

Université de Montréal

Beyond hairballs: Depicting complexity of a kinase-phosphatase network in the budding yeast

par Diala Abd-Rabbo

Département de biochimie et médecine moléculaire
Faculté de médecine

Thèse présentée en vue de l'obtention du grade de
Philosophiæ Doctor (Ph.D.)
en Bio-informatique

Janvier, 2017

© Diala Abd-Rabbo, 2017

Résumé

Les kinases et les phosphatases (KP) représentent la plus grande famille des enzymes dans la cellule. Elles régulent les unes les autres ainsi que 60 % du protéome, formant des réseaux complexes kinase-phosphatase (KP-Net) jouant un rôle essentiel dans la signalisation cellulaire. Ces réseaux caractérisés d'une organisation de type commandes-exécutions possèdent généralement une structure hiérarchique. Malgré les nombreuses études effectuées sur le réseau KP-Net chez la levure, la structure hiérarchique ainsi que les principes fonctionnels sont toujours peu connus pour ce réseau. Dans ce contexte, le but de cette thèse consistait à effectuer une analyse d'intégration des données provenant de différentes sources avec la structure hiérarchique d'un réseau KP-Net de haute qualité chez la levure, *S. cerevisiae*, afin de générer des hypothèses concernant les principes fonctionnels de chaque couche de la hiérarchie du réseau KP-Net.

En se basant sur une curation de données d'interactions effectuée dans la présente et dans d'autres études, le plus grand et authentique réseau KP-Net reconnu jusqu'à ce jour chez la levure a été assemblé dans cette étude. En évaluant le niveau hiérarchique du KP-Net en utilisant la métrique de la centralisation globale et en élucidant sa structure hiérarchique en utilisant l'algorithme vertex-sort (VS), nous avons trouvé que le réseau KP-Net possède une structure hiérarchique ayant la forme d'un sablier, formée de trois niveaux disjoints (supérieur, central et inférieur). En effet, le niveau supérieur du réseau, contenant un nombre élevé de KPs, était enrichi par des KPs associées à la régulation des signaux cellulaires; le niveau central, formé d'un nombre limité de KPs fortement connectées les unes aux autres, était enrichi en KPs impliquées dans la régulation du cycle cellulaire; et le niveau inférieur, composé d'un nombre important de KPs, était enrichi en KPs impliquées dans des processus cellulaires diversifiés.

En superposant une grande multitude de propriétés biologiques des KPs sur le réseau KP-Net, le niveau supérieur était enrichi en phosphatases alors que le niveau inférieur en était appauvri, suggérant que les phosphatases seraient moins régulées par phosphorylation et déphosphorylation que les kinases. De plus, le niveau central était enrichi en KPs représentant des « bottlenecks », participant à plus d'une voie de signalisation, codées par des gènes

essentiels et en KPs qui étaient les plus strictement régulées dans l'espace et dans le temps. Ceci implique que les KPs qui jouent un rôle essentiel dans le réseau KP-Net devraient être étroitement contrôlées. En outre, cette étude a montré que les protéines des KPs classées au niveau supérieur du réseau sont exprimées à des niveaux d'abondance plus élevés et à un niveau de bruit moins élevé que celles classées au niveau inférieur du réseau, suggérant que l'expression des enzymes à des abondances élevées invariables au niveau supérieur du réseau KP-Net pourrait être importante pour assurer un système robuste de signalisation.

L'étude de l'algorithme VS a montré que le degré des nœuds affecte leur classement dans les différents niveaux d'un réseau hiérarchique sans biaiser les résultats biologiques du réseau étudié. En outre, une analyse de robustesse du réseau KP-Net a montré que les niveaux du réseau KP-Net sont modérément stable dans des réseaux bruités générés par ajout d'arrêtes au réseau KP-Net. Cependant, les niveaux de ces réseaux bruités et de ceux du réseau KP-Net se superposent significativement. De plus, les propriétés topologiques et biologiques du réseau KP-Net étaient retenues dans les réseaux bruités à différents niveaux. Ces résultats indiquent que bien qu'une robustesse partielle de nos résultats ait été observée, ces derniers représentent l'état actuel de nos connaissances des réseaux KP-Nets.

Finalement, l'amélioration des techniques dédiées à l'identification des substrats des KPs aideront davantage à comprendre comment les réseaux KP-Nets fonctionnent. À titre d'exemple, je décris, dans cette thèse, une stratégie que nous avons conçu et qui permet à déterminer les interactions KP-substrats et les sous-unités régulatrices sur lesquelles ces interactions dépendent. Cette stratégie est basée sur la complémentation des fragments de protéines basée sur la cytosine désaminase chez la levure (OyCD PCA). L'OyCD PCA représente un essai *in vivo* à haut débit qui promet une description plus précise des réseaux KP-Nets complexes. En l'appliquant pour déterminer les substrats de la kinase cycline-dépendante de type 1 (Cdk1, appelée aussi Cdc28) chez la levure et l'implication des cyclines dans la phosphorylation de ces substrats par Cdk1, l'essai OyCD PCA a montré un comportement compensatoire collectif des cyclines pour la majorité des substrats. De plus, cet essai a montré que la tubuline- γ est phosphorylée spécifiquement par Clb3-Cdk1, établissant ainsi le moment pendant lequel cet événement contrôle l'assemblage du fuseau mitotique.

Mots-clés : Réseau des kinases-phosphatases, *Saccharomyces cerevisiae*, Complexité des réseaux, Structure hiérarchique des réseaux, Algorithmes de décomposition de réseaux, propriétés topologiques, Intégration de données multi-omiques, Essai de complémentation des fragments de protéine basé sur la cytosine désaminase chez la levure, Kinase cycline-dépendante de type 1 et la tubuline- γ .

Abstract

Kinases and phosphatases (KP) form the largest family of enzymes in living cells. They regulate each other and 60 % of the proteome forming complex kinase-phosphatase networks (KP-Net) essential for cell signaling. Such networks having the command-execution aspect tend to have a hierarchical structure. Despite the extensive study of the KP-Net in the budding yeast, the hierarchical structure as well as the functional principles of this network are still not known. In this context, this thesis aims to perform an integrative analysis of multi-omics data with the hierarchical structure of a bona fide KP-Net in the budding yeast *Saccharomyces cerevisiae*, in order to generate hypotheses about the functional principles of each layer in the KP-Net hierarchy.

Based on a literature curation effort accomplished in this and in other studies, the largest bona fide KP-Net of the *S. cerevisiae* known to date was assembled in this thesis. By assessing the hierarchical level of the KP-Net using the global reaching centrality and by elucidating its hierarchical structure using the vertex-sort (VS) algorithm, we found that the KP-Net has a moderate hierarchical structure made of three disjoint layers (top, core and bottom) resembling a bow tie shape. The top layer having a large size was found enriched for signaling regulation; the core layer made of few strongly connected KPs was found enriched mostly for cell cycle regulation; and the bottom layer having a large size was found enriched for diverse biological processes.

On overlaying a wide range of KP biological properties on top of the KP-Net hierarchical structure, the top layer was found enriched for and the bottom layer was found depleted for phosphatases, suggesting that phosphatases are less regulated by phosphorylation and dephosphorylation interactions (PDI) than kinases. Moreover, the core layer was found enriched for KPs representing bottlenecks, pathway-shared components, essential genes and for the most tightly regulated KPs in time and space, implying that KPs playing an essential role in the KP-Net should be firmly controlled. Interestingly, KP proteins in the top layer were found more abundant and less noisy than those of the bottom layer, suggesting that availability of enzymes at invariable protein expression level at the top of the network might be important to ensure a robust signaling.

Analysis of the VS algorithm showed that node degrees affect their classification in the different layers of a network hierarchical structure without biasing biological results of the sorted network. Robustness analysis of the KP-Net showed that KP-Net layers are moderately stable in noisy networks generated by adding edges to the KP-Net. However, layers of these noisy overlap significantly with those of the KP-Net. Moreover, topological and biological properties of the KP-Net were retained in the noisy networks to different levels. These findings indicate that despite the observed partial robustness of our results, they mostly represent our current knowledge about KP-Nets.

Finally, enhancement of techniques dedicated to identify KPs substrates will enhance our understanding about how KP-Nets function. As an example, I describe here a strategy that we devised to help in determining KP-substrate interactions and the regulatory subunits on which these interactions depend. The strategy is based on a protein-fragment complementation assay based on the optimized yeast cytosine deaminase (OyCD PCA). The OyCD PCA represents a large scale *in vivo* screen that promises a substantial improvement in delineating the complex KP-Nets. We applied the strategy to determine substrates of the cyclin-dependent kinase 1 (Cdk1; also called Cdc28) and cyclins implicated in phosphorylation of these substrates by Cdk1 in *S. cerevisiae*. The OyCD PCA showed a wide compensatory behavior of cyclins for most of the substrates and the phosphorylation of γ -tubulin specifically by Clb3-Cdk1, thus establishing the timing of the latter event in controlling assembly of the mitotic spindle.

Keywords: Kinase-phosphatase network, *Saccharomyces cerevisiae*, Network complexity, Network hierarchical structure, Network decomposition algorithms, Topological properties, Integration of multi-omics data, protein-fragment complementation assay based on the optimized yeast cytosine deaminase, Cyclin-dependent kinase 1 and γ -tubulin.

Table of Contents

RÉSUMÉ	I
ABSTRACT	IV
TABLE OF CONTENTS	VI
LIST OF TABLES	XII
LIST OF FIGURES	XIII
LIST OF ABBREVIATIONS	XV
ACKNOWLEDGMENTS	XVIII
CHAPTER 1 INTRODUCTION	20
1.1 Complexity of living cells	20
1.2 Complexity of the kinase-phosphatase networks (KP-Nets)	21
1.3 Kinase-phosphatase networks	23
1.3.1 Phosphorylation and dephosphorylation	23
1.3.2 Kinases and phosphatases	24
1.3.2.1 Protein kinases	25
1.3.2.2 Protein phosphatases	28
1.3.3 Mapping phosphorylation networks	29
1.3.3.1 <i>In vitro</i> experimental methods	29
1.3.3.1.1 In vitro kinase assays using protein microarrays	30
1.3.3.1.2 In vitro kinase assays using peptide libraries	30
1.3.3.1.3 In vitro kinase assays using analogue-sensitive kinase alleles	31
1.3.3.2 <i>In vivo</i> experimental methods	32

1.3.3.2.1	In vivo phosphorylation technologies based on liquid-chromatography and mass-spectrometry.....	32
1.3.3.2.2	In vivo Protein-fragment Complementation Assay (PCA) based on the optimized yeast cytosine deaminase.....	33
1.3.4	Protein phosphorylation and protein interaction databases-----	34
1.4	Analyzing biological networks-----	34
1.4.1	Studying biological processes using control theory-----	35
1.4.2	Assessing network topological properties -----	37
1.4.2.1	Local topological properties of networks-----	38
1.4.2.2	Global topological properties of networks-----	39
1.4.2.3	Network logic motifs-----	40
1.4.3	Network models-----	41
1.4.3.1	Erdős-Rényi model - Random networks-----	41
1.4.3.2	Watts-Strogatz model - Small-world networks-----	43
1.4.3.3	Barabási-Albert model - Scale-free networks-----	43
1.4.3.4	Ravaz-Barabási model - Hierarchical networks-----	44
1.4.4	Studying hierarchical structures in networks-----	45
1.4.4.1	Types of network hierarchical structures-----	45
1.4.4.2	Measuring network hierarchical level -----	47
1.4.4.3	Elucidating network hierarchy-----	48
1.5	Correlation between topological properties and biological functions -----	53
1.5.1	Do node topological properties predict node biological functions?-----	53
1.5.2	Does logic motif topology predict logic motif biological functions?-----	54
1.5.3	Does network topology predict network biological functions?-----	55
1.6	Data integration-----	56
1.6.1	Integrating other biological networks with phosphorylation networks -----	57
1.6.2	Integrating biological data with KP-Nets -----	58
1.7	Thesis objectives and organization -----	59
1.7.1	Thesis aim and objectives -----	59
1.7.2	Thesis organization -----	59

CHAPTER 2 DELINEATING FUNCTIONAL PRINCIPLES OF THE BOW TIE STRUCTURE OF A KINASE-PHOSPHATASE NETWORK IN THE BUDDING YEAST-----	61
2.1 Authors' contributions -----	62
2.2 Abstract-----	63
2.3 Background-----	64
2.4 Results -----	66
2.4.1 The kinase-phosphatase network (KP-Net) -----	66
2.4.2 The KP-Net possesses a “corporate” hierarchical structure in the form of a bow tie with a strongly connected core layer -----	67
2.4.3 The three layers of the KP-Net have dissimilar biological roles and subcellular localization -----	69
2.4.4 Phosphatases are less regulated by phosphorylation than kinases-----	71
2.4.5 KP-Net upper levels are the least regulated and KP-Net lower levels are the least to regulate other KPs -----	71
2.4.6 The KP-Net core layer is enriched for essential genes, bottlenecks, and pathway-shared components -----	73
2.4.7 Molecular switches are enriched in KPs in core and bottom layers -----	73
2.4.8 Core layer KPs employ scaffolding to prevent unwanted pathway crosstalk-----	75
2.4.9 Core layer KPs undergo more spatial organization changes than top and bottom layer KPs-----	75
2.4.10 Top layer KP proteins are more abundant and less noisy than bottom layer KPs of the KP-Net -----	76
2.4.11 The VS algorithm depends on node degree to classify network nodes in three layers -----	77
2.4.12 Biological properties of KPs are independent of their in- and out-degrees-----	79
2.4.13 Robustness of results and incompleteness of data -----	80
2.4.14 Predicting Kinases acting on substrates on high osmolarity stress -----	81
2.4.15 Discussions-----	82
2.4.16 Conclusions -----	84
2.5 Methods-----	84
2.5.1 Over-representation of various logic motifs in the KP-Net -----	84
2.5.2 Network randomization -----	85
2.5.2.1 Degree preserving randomization (DPR)-----	85

2.5.2.2	Similar degree preserving randomization (SDPR)-----	85
2.5.2.3	Out-degree preserving randomization (ODPR)-----	85
2.5.2.4	Degree non-preserving randomization (DNP) -----	85
2.5.3	The matching algorithm -----	85
2.5.4	Testing whether the KP-Net GRC is bigger than Erdős–Rényi network GRCs-----	86
2.5.5	Comparing means of node properties in two layers using RT-----	86
2.5.6	Generating subsampled/noisy networks and assessing their layers stability and their overlap with KP-Net layers -----	86
2.5.7	Predicting kinases -----	87
2.6	Acknowledgements -----	87
CHAPTER 3 DISSECTION OF CDK1-CYCLIN COMPLEXES <i>IN VIVO</i> -----		88
3.1	Authors' contributions -----	89
3.1.1	Quantifying colony growth -----	89
3.1.2	Preys in OyCD PCA are implicated in various biological processes-----	90
3.1.3	Comparison of OyCD PCA data with other datasets-----	91
3.1.3.1	Comparison of first screen of OyCD PCA data with other datasets -----	92
3.1.3.2	Comparison of second screen of OyCD PCA data with other datasets -----	94
3.1.4	Other contributions-----	94
3.2	Abstract-----	95
3.3	Significance-----	95
3.4	Introduction-----	96
3.5	Results -----	98
3.5.1	Identifying Cdk1 complexes <i>in vivo</i> -----	98
3.5.2	Cyclin dependency of Cdk1 complexes -----	100
3.5.3	γ -Tubulin is an <i>in vivo</i> target of Cdk1–Clb3 -----	104
3.5.4	Clb3–Cdk1 preferentially targets γ -tubulin <i>in vitro</i> -----	106
3.5.5	Discussion -----	106
3.6	Materials and methods -----	108

3.6.1 Yeast strains -----	108
3.6.2 Plasmid construction -----	108
3.6.3 Gateway cloning-----	109
3.6.4 OyCD PCA to detect protein–protein interactions with Cdk1 -----	109
3.6.5 Detecting protein–protein interactions in the different cyclin deletion strains -----	110
3.6.6 Analysis of the cyclin deletion strains -----	110
3.6.7 Purification of γ-TUSC substrate and particle analysis -----	111
3.6.8 Cdk1 kinase assays-----	111
3.7 Note added in proof-----	112
3.8 Acknowledgements -----	112
CHAPTER 4 CONCLUSIONS AND FUTURE WORK-----	113
4.1 Conclusions-----	113
4.2 Future directions-----	124
REFERENCES -----	129
APPENDIX 1.METHODS, TABLES AND SUPPLEMENTARY FIGURES OF CHAPTER 2 -- I	
Supplementary materials-----	i
Supplementary methods -----	iii
Supplementary tables -----	vii
Supplementary figures -----	x
APPENDIX 2.SUPPLEMENTARY METHODS, TABLES AND FIGURES OF CHAPTER 3	
-----	XIX
Supplementary methods -----	xix
Supplementary tables -----	xxi
Supplementary figures -----	xxiii
APPENDIX 3.SCIENTIFIC CONTRIBUTIONS -----	XXV

Publications -----	xxv
Conferences and presentations-----	xxvi

List of tables

Table 1.I.	Families of protein kinases and their corresponding members. -----	26
Table 1.II.	Protein phosphatases families and their corresponding phosphatases members. -----	28
Table 3.I.	Top enriched GO terms in the preys that were tested to interact with Cdc28 -----	91
Table 3.II.	Significance of the overlap between Cdc28 interactors identified in Ear et al. and in other datasets.-----	94
Table A1.I.	Annotation of dephosphorylation interactions.-----	vii
Table A1.II.	K-Nets and KP-Nets studies. -----	viii
Table A1.III.	Phosphorylation and dephosphorylation interactions used to assemble the KP-Net (Supplementary file on CD-ROM) -----	ix
Table A1.IV.	Core layer KPs implicated in decision-making.-----	ix
Table A1.V.	The kinase-substrate interactions that were predicted to be implicated in osmotic shock in this study (Supplementary file on CD-ROM)-----	ix
Table A2.I.	List of prey proteins that interact with Cdk1 identified by OyCD PCA -----	xxi
Table A2.II.	False-negative rates (FNR) of the Cdk1 screen-----	xxii
Table A2.III.	Sensitivity and specificity of the Cdk1–protein interaction screen in cyclin deletion strain backgrounds -----	xxii
Table A2.IV.	Cdk1 interaction death index for all prey genes (Supplementary file on CD-ROM). -----	xxii
Table A2.V.	Cyclin contingency of 21 Cdk1 interacting proteins (Supplementary file on CD-ROM). --	xxii

List of figures

Figure 1.1. Mechanisms of kinases specificity.....	27
Figure 1.2. A closed-loop control system.....	35
Figure 1.3. Different types of network logic motifs.....	41
Figure 1.4. Network models.....	42
Figure 1.5. Different types of network hierarchical structures.....	46
Figure 1.6. Application of the different sorting algorithms on a toy network and a comparison of their performance.....	52
Figure 2.1. The pipeline used to assemble and to sort the KP-Net, and the KP-Net bow tie structure.....	67
Figure 2.2. Depleted and enriched biological processes and cellular components in each of the KP-Net layers.....	70
Figure 2.3. Topological and biological properties of KPs in the different layers of the KP-Net.....	72
Figure 2.4. Biochemical and spatiotemporal modulators of KPs in the different layers of the KP-Net.....	74
Figure 2.5. mRNA and protein turnover related properties of KPs in the different layers of the KP-Net.....	77
Figure 2.6. The VS algorithm depends on node degrees to sort network nodes in three layers.....	78
Figure 2.7. Stability of KP-Net layers and their overlap with subsampled/noisy network layers.....	81
Figure 3.1. Dissecting Cdk1 complexes using the OyCD PCA.....	97
Figure 3.2. Identification of interaction partners of Cdk1.....	99
Figure 3.3. Cyclin dependence of Cdk1–prey protein interactions.....	101
Figure 3.4. γ -Tubulin is a Clb3–Cdk1 substrate <i>in vitro</i>	105
Figure A1.1. Distributions of the cumulative degree and of the clustering coefficient of KP-Net nodes follow a power law distribution.....	x
Figure A1.2. Detailed hierarchical structure of the KP-Net.....	xi
Figure A1.3. Percentage of phosphatases in each of the three layers of the KP-Net and distribution of phosphosites among KPs and among the different layers of the KP-Net.....	xii
Figure A1.4. Clustering of the different sets of randomized networks.....	xiii
Figure A1.5. Correlation between KP in-degrees and KP numerical properties.....	xiv
Figure A1.6. In-degree distribution of KPs characterized and not characterized by each of the studied categorical properties.....	xv
Figure A1.7. Correlation between KP out-degrees and KP numerical properties.....	xvi

Figure A1.8. Out-degree distribution of KPs characterized and not characterized by each of the studied categorical properties.....	xvii
Figure A1.9. Distribution of the different properties of KPs in each layer of 100 noisy networks.	xviii
Figure A2.1. Optimized yeast <i>Saccharomyces cerevisiae</i> prodrug-converting enzyme cytosine deaminase (OyCD) protein fragment complementation assay cyclin-dependent kinase (Cdk)-prey interaction assays in nine cyclin deletion strains.	xxiii

List of abbreviations

5-FC	5-fluorocytosin
ATP	Adenosine triphosphate
BFS	Breadth-first search
Cdk	Cyclin-dependent kinase
DNPR	Degree non-preserving randomization
DPR	Degree preserving randomization
ELM	Eukaryotic linear motif database
FBL	Feedback loop
FCY1	yCD gene
FFL	Feedforward loop
FN	False negative
FNR	False-negative rate
GO	Gene ontology
GRC	Global reaching centrality
GST	Gluthathione-S-transferase
GTP	Guanosine triphosphate
HIDEN	Hierarchical decomposition of regulatory network
HiNO	Hierarchical regulatory networks organization
HS	Hierarchical score
HSM	Hierarchical score maximization
HT	Hypergeometric test
HTP	High-throughput
IDPR	In-degree preserving randomization
KP	Kinase and phosphatase
K-Net	Kinase phosphorylation network
KID	Kinase interaction database
KP-Net	Kinase-phosphatase network
LBM	Linear binding motif
LC	Liquid chromatography
LC-MS/MS	Liquid-chromatography and mass-spectrometry

LTP	Low-throughput
MAPK	Mitogen-activated protein kinase
MS	Mass-spectrometry
ODPR	Out-degree preserving randomization
ORF	Open reading frame
OyCD	Optimized yeast cytosine deaminase
PCA	Protein-fragment complementation assay
PDI	Phosphorylation and dephosphorylation interaction
PPI	Protein-protein interaction
PKA	Protein kinase A
PPM	protein phosphatases manganese or magnesium dependent
PPP	Phosphoprotein phosphatase
PTM	Post-translational modification
PTP	Protein Tyr-specific phosphatase
RT	Randomization test
SCC	Strongly connected component
SDPR	Similar degree preserving randomization
Ser	Serine
SIM	Single-input module
SPB	Spindle pole body
STE	Sterile
TAP	Tandem affinity purification
TF	Transcription factor
TF-Net	Transcription network
Thr	Threonine
TP	True positive
Tyr	Tyrosine
VS	Vertex-sort
WT	Wild type
WWW	World wide web

To my Creator,

*to the One Who created the immense universe
with a lot of secrets, and invites us to unravel
them in order to know Him better and
to appreciate His gifts.*

To God, the One and only Creator.

Acknowledgments

Words cannot express my deepest and sincere gratitude for all the blessings that God has given me. It is because of Him first that I was able to realize this research.

I am forever indebted to my advisor, Prof. Stephen Michnick, for his confidence in me providing me the opportunity to pursue a Ph.D. degree in his laboratory. I would like to express my sincere gratitude for his enthusiasm, invaluable advices, guidance, and for being constantly available even during his vacations. I greatly appreciate his great human qualities including but not limited to listening to me, his patience, his exceptional understanding and his constant support and encouragement during difficult moments along this thesis.

I would like also to express my gratitude to the members of the jury of my thesis, Dr. Martine Raymond, Dr. Sébastien Lemieux, Dr. Mathieu Blanchette and Dr. Janos Filep, for having accepted to read and evaluate my work despite their busy agenda, and for their valuable and instructive comments.

My acknowledgements go equally to all my colleagues in the Michnick laboratory, in particular to Jacqueline Kowarzyk, Lara Matta, Emmanuelle Tchekanda, Luz Carrillo, Alessandra Nurrisso, and Durgajini Sivanesan for their support, encouragement, and for the unforgettable moments that we spent together. I would like also to thank Emmanuel Levy and Abdellali Kelil for their pertinent advices and the instructive discussions that I had with them. I would like to say a big thank you for the administrative responsible for the Bioinformatics program, Mrs Elaine Meunier. She provided me with invaluable support and she was always available with her charming smile and affectionate attention.

No words can express my sincere and deepest gratitude to my exceptional mom, Hana, who encouraged me continuously with a lot of affection and who had invested considerable time, energy and effort to help me overcome a major vision deficiency to finally succeed my scholar studies. She tirelessly and with a lot of determination instilled in me great confidence. I owe my grandparents and my uncle so much, Teita Fathiyyeh, Jeddo Shafik and uncle Saed for their infinite love, and touching attention. I will never know how to express how much I love them, how much I miss them, and how much I think of them, may their souls rest in peace. An exceptional thank you goes to my special dad, Ahmad, an exceptional example of

an avid reader and a well instructed person who taught me the importance of knowledge; to my step father, Ghassan, a great dentist who taught me how to adopt a systematic and efficient approach in my studies. A huge thank you for the love of my life, Salaheddine, who believed in me and since the first day of our marriage, more specifically four years before the accomplishment of this thesis, he already awarded me the title of Doctor. I would like to thank my sisters, Reem and Noor and my brothers Abboud and Samer, my aunts, in particular, aunt Sana and aunt Sahar for their continuous encouragement, their distinguished attention and their prayers. I would like to infinitely thank my sister and my dearest and closest friend Rahima Ziane, for her friendship, for her precious advices, for her support and for the tons of positive waves she sent me throughout this thesis. Finally, I would like to thank my lovely friends Fati and Dima for being available to listen to me and to always encourage me to continue till the end.

Chapter 1 Introduction

1.1 Complexity of living cells

In nature, living cells are constantly exposed to unstable environmental conditions leading to stress situations such as change in acidity, osmolarity, temperature, availability of nutrient supplies, exposure to radiations, and many other external factors. Interestingly, cells have a remarkable capacity to adapt to new environmental conditions by coordinating their intracellular activities and to correct incidental internal errors (Gasch and Werner-Washburne, 2002). For instance, when DNA damage occurs during DNA replication, the cell cycle stalls and a DNA repair process is launched (Putnam et al., 2009). In case of a successful repair of the DNA damage, both DNA replication and the cell cycle are resumed. Otherwise, the cell is destroyed by launching a programmed cell death called apoptosis. Another example of internal errors occurs when proteins are incorrectly folded. To remedy such situations, the cell attempts to refold these proteins. In case of failure, the cell targets misfolded proteins for degradation (Kriegenburg et al., 2012). These examples and many others show that biological events are well organized in time and space and importantly, are driven by an apparent capacity of the living cells for decision-making.

Notably, the biological events that are triggered, at least in the two examples mentioned above, involve various biochemical mechanisms such as: (1), signaling to sense unusual conditions (DNA damage or protein misfolding in these cases) and to transmit information to other molecules after signal detection; (2), protein-protein interactions (PPI) to form biochemical machineries required for the repair process; (3), post-translational modifications (PTM) including phosphorylation and dephosphorylation interactions (PDI) to regulate individual proteins or the assembled complexes; and (4), metabolic reactions to transform the damage products perceived as cellular toxic to non-toxic products (Chan and Dedon, 2010). Accordingly, living cells can be described as a complex system in which different networks [e.g. signaling network, PPI network, metabolic network, transcription network (TF-Net), ubiquitination network (Venancio et al., 2009) and acetylation network (Choudhary et al., 2009)] interact with each other and affect each other. Complexity in living cells emerges from the interplay between different biological networks on the one hand and on the other hand from the complexity

encountered within each of these networks known usually as complex networks. In fact, there is no single and formal definition for complex networks in the literature. For instance, Barabási and Albert described complex networks as large networks (having a large number of nodes) that are characterized by topological properties different from those of random networks (Paragraph 1.4.3.1); in particular they contain few nodes that are extremely connected to other nodes (hubs), they expand continuously by adding new nodes, and their new nodes connect preferentially to well-connected nodes (Paragraph 1.4.3.3) (Barabasi and Albert, 1999). Whereas, Steen described a network as complex, if it is so large in size and so interconnected, that it is difficult to understand the behavior of the whole network by examining the behavior of each of its nodes separately (Steen, 2010). This thesis adopts Steen's definition for complex networks with a minor addition highlighting the importance of the organizational structure of a network. Consequently, in the context of this thesis, a complex network represents a large network formed of extremely interconnected nodes; this interconnectedness hinders the identification of the organizational structure of this network and hinders understanding the behavior of the whole network by observing the individual behavior of its nodes.

1.2 Complexity of the kinase-phosphatase networks (KP-Nets)

Among the various biological networks, the KP-Net is the one that exhibits the most prevalent, immediate and profound effect on the cellular biochemical machinery implicated in a vast range of biological processes (Paragraph 1.3.1). KP-Nets play an essential role in signaling pathways (Graves and Krebs, 1999; Pawson and Scott, 2005). Textbooks usually describe signaling pathways as simple isolated linear cascades of enzymes, particularly of kinases and phosphatases (KP), which are activated in response to a stimulus ultimately to generate a cellular response such as metabolic alterations, gene activation and repression, and PDIs. This cellular response aims at changing the cell behavior to maintain cellular homeostasis. It is now well established that signaling does not occur through linear pathways, but through highly interconnected pathways forming complex signaling networks.

Recent evidence suggests that KP-Nets are extremely complex for different reasons. First, KPs in these networks not only regulate proteins having no enzymatic activity, but also regulate each other, making KPs regulators of regulators or super-regulators (Bodenmiller et al., 2010; Breitkreutz et al., 2010; Fiedler et al., 2009). Second, various types of network building blocks

known as logic motifs (subgraphs) are over-represented in KP-Nets (Alon, 2007). These over-represented logic motifs have a more complex topology than the simple linear chains corresponding to linear cascades of KPs in KP-Nets (e.g. single-input modules (SIM), feedforward loops (FFL), diamonds and bi-fans, Paragraph 1.4.2.3). Third, results from a recent study showed a high level of complexity associated to a KP-Net in the budding yeast, for which the perturbation of most KPs affect large parts of the KP-Net (Bodenmiller et al., 2010). Finally, another level of complexity in KP-Nets results from the compensatory behaviors of certain KPs. On the knockdown of a particular KP, another KP closely related to the deleted one could partially or completely compensate for the action of the deleted enzyme. In brief, these findings suggest that the KP-Net is an extremely complex network similar to a hairball that cannot be organized in order to delineate its structure and in order to ultimately understand how it functions.

In the following sections, first, I will introduce KP-Nets, by describing: KPs families, the mechanisms guiding the substrates specificity of KPs, the large-scale experiments used to map KP-Nets, strengths and limitations of these experiments, as well as the available databases designed to annotate interactions implicating KPs. Second, I will introduce the tools used to analyze biological networks, by presenting: a general overview about the application of the control theory to study the structure and dynamics of small networks, the network topological properties used to describe networks, the different network models conceived to build networks having predefined properties, and the most used bioinformatics methods applied to determine and elucidate network architectures. Third, I will also touch on studies that aimed at inferring the biological functions of nodes, of logic motifs and of the networks themselves from topological properties of each of these entities. I will also outline the important efforts performed to integrate data within KP-Nets. Finally, given the extensive use of the budding yeast *Saccharomyces cerevisiae* as a model organism for studying KP-Nets, a special attention to the literature related to *S. Cerevisiae* will be considered in this introduction and the rest of the thesis.

1.3 Kinase-phosphatase networks

1.3.1 Phosphorylation and dephosphorylation

Phosphorylation is a biochemical interaction by which a kinase catalyzes the transfer of a phosphate group from adenosine triphosphate (ATP) or guanosine triphosphate (GTP) principally to serine (Ser), threonine (Thr) or tyrosine (Tyr) residues and less commonly to other residues such as histidine, arginine and lysine of a protein known as the kinase substrate (Ciesla et al., 2011; Ubersax and Ferrell, 2007). Dephosphorylation represents the inverse reaction by which a phosphatase catalyzes the transfer of a phosphate group from an amino acid residue of a phosphoprotein to a water molecule. Hence, the phosphorylation level of a protein residue is affected by the simultaneous activity of KPs on this site. Phosphorylation and dephosphorylation could also target small-molecules such as lipids and carbohydrates (Fiedler et al., 2009), but work in this thesis will focus on protein PDIs.

Phosphorylation is the most prevalent PTM of proteins that affects their function by altering their structural conformation and consequently their enzymatic activity in case they have one, their subcellular localization, their abundance by targeting them for ubiquitination in a phospho-dependent manner and their molecular interactions by shaping their interaction surfaces to create new or disrupt existing PPIs (Hunter, 2000; Seet et al., 2006). This PTM regulates a staggering number of cellular processes ranging from stress response, cell cycle, cell proliferation and death to cellular metabolism. Usually, regulation of biological processes by phosphorylation interaction might necessitate a combined sequence of many PDIs (Graves and Krebs, 1999; Pawson and Scott, 2005).

Generally, a phosphosite is said to be functional if it has a known biological function or if its mutation results in an observable phenotype or change in fitness of an organism (Lasalde et al., 2014; Yoshimi et al., 2012). To date, the function of a limited number of phosphosites has been characterized. Recent studies suggest that most phosphosites are non-functional and that they represent noise due to non-specific phosphorylation of, particularly, abundant proteins encountered by kinases in the crowded environment of the cell (Ba and Moses, 2010; Landry et al., 2009; Levy et al., 2012; Malik et al., 2008). These studies suggest that functional phosphosites could be systematically identified if they meet the following criteria: (1), to occur

in lower abundance proteins; (2), to be on average more conserved across species; and (3), to have high stoichiometry in comparison with non-functional sites (phosphosites having no known function). Although these studies provided elegant approaches to distinguish between functional and non-functional phosphosites, these criteria are not necessarily predictive of phosphosite functionality (Tan et al., 2010; Wu et al., 2011). Moreover, numerous phosphosites appear to have no functional consequences on their mutation, because their mutation would not result in any cellular phenotypic change, or because of the existence of other phosphosites having a redundant role (Lasalde et al., 2014; Yoshimi et al., 2012). Therefore, a safe approach to confidently determine non-functional phosphosites would be to investigate not only the non-functional consequence of each site separately, but also for all site combinations per protein (Lienhard, 2008). Currently, no screens are considered to assess the functionality of individual and combined phosphosites on a large scale for two reasons. First, the mutation of this staggering number of phosphosites, apart from being tremendously time-consuming, is not feasible given the currently available experimental approaches. Second, it is extremely difficult to find a general experimental assay to determine the functionality of a phosphosite. And yet, advanced site-directed mutagenesis technologies combined with phenotypic high-throughput (HTP) screens could open new avenues for unraveling a considerable number of functional phosphosites in the future.

1.3.2 Kinases and phosphatases

Kinases and phosphatases belong to the class of phosphotransferases, the largest class of enzymes in eukaryotic cells. They are responsible for phosphorylating and dephosphorylating their substrates respectively. Kinases and phosphatases can also act on themselves and on each other (Bodenmiller et al., 2010; Whinston et al., 2013; Wu et al., 2009). They both represent ~2 % of eukaryotic cells proteome. The number of kinases is estimated to 133 in *Saccharomyces cerevisiae*, 454 in *Caenorhabditis elegans*, 251 in *Drosophila melanogaster* and 518 in *Homo sapiens*, whereas the number of phosphatases is estimated to 41 in *Saccharomyces cerevisiae*, 185 in *Caenorhabditis elegans*, 86 in *Drosophila melanogaster* and 126 in *Homo sapiens* (Manning et al., 2002b; Morrison et al., 2000). Recently, KPs were shown to regulate at least 60 % of budding yeast proteins (Yachie et al., 2011). Hence, it is not surprising that aberrant regulation of KPs leads to serious perturbations of the regulatory and signaling networks, a result

that has been clearly observed on the knockdown of most of these enzymes one at a time (Bodenmiller et al., 2010).

1.3.2.1 Protein kinases

Kinases can be subdivided into three groups according to the specificity they exhibit towards the phosphorylated residue: Ser/Thr-specific, Tyr-specific and dual-specificity kinases. Most of the kinases in the budding yeast are Ser/Thr-specific. Only one has been shown to be Tyr-specific (Malathi et al., 1999), a number that could be underestimated (Yachie et al., 2011; Zhu et al., 2000). And seven kinases showed a dual-specificity; they can phosphorylate Ser, Thr or Tyr residues (Hunter and Plowman, 1997).

Kinases can also be classified according to the primary sequence similarity of their catalytic domains into seven families: (1), the family that groups protein kinases A, G and C; (2), the calcium/calmodulin-dependent kinase family; (3), the family that groups Cdks, mitogen-activated protein kinases (MAPK), glycogen synthase kinase 3 β and Cdk-like kinases; (4), the casein kinase 1 family; (5), the “STERile” family (STE) or kinases of MAPK grouping Ste7, Ste11 and Ste20 subfamilies; (6), the other family grouping kinases that do not share strong similarity with the previously mentioned kinase families; and (7), the atypical family, grouping proteins that lack sequence similarity to the conventional eukaryotic protein kinase catalytic domains, yet they were experimentally shown to have a kinase activity (Hunter and Plowman, 1997) (Table 1.I).

Kinase catalytic domains, made of ~250 amino acids and having the form of a cleft, binds the ATP and the kinase substrate and catalyzes the transfer of a phosphate from the ATP to a Ser (72 %), Thr (~23 %) and rarely to a Tyr (~5 %) residue within the substrate protein sequence in the budding yeast (Yachie et al., 2011). The sequence recognized by a kinase catalytic domain is called the consensus sequence and varies in length between four and eight residues (Ubersax and Ferrell, 2007). Identification of these sequences became possible with the emergence of peptide library screens and other similar technologies (Mok et al., 2010; Olsen et al., 2006; Pinna and Ruzzene, 1996). Although phosphorylation consensus sequences enhance the recognition of a kinase to its substrates, these elements seem to be insufficient to determine protein kinase specificity towards their substrates for at least three reasons. First, different Ser/Thr-specific kinases were found to have similar minimal phosphorylation consensus sequences. For instance,

Cdk and protein kinase A (PKA) both recognize the [S/T]P sequence (Ubersax and Ferrell, 2007). Second, consensus sequences are characterized to be of short length and to have a degenerate nature. Finally, certain authentic phosphorylation sequences do not match identified consensus phosphorylation sequences (Kreegipuu et al., 1998; Ubersax et al., 2003).

Table 1.I. Families of protein kinases and their corresponding members.

Kinases belonging to the different kinase families were determined from (Hunter and Plowman, 1997; Manning et al., 2002a; Thomas et al., 2003).

Protein kinases A, G and C Family (AGC: 17)									
CBK1	DBF2	DBF20	FPK1	KIN82	PKC1	PKH1	PKH2	PKH3	RIM15
SCH9	TPK1	TPK2	TPK3	YPK1	YPK2	YPK3			
Calcium/calmodulin-dependent Kinases Family (CaMK: 22)									
CHK1	CMK1	CMK2	DUN1	ELM1	FRK1	GIN4	HSL1	KCC4	KIN1
KIN2	KIN4	MEK1	PRR1	RCK1	RCK2	SAK1	SNF1	SPK1	TOS3
YMR291W YPL150W									
Cyclin dependent, Mitogen-activated, Glycogen synthase and Cyclin dependent-like kinases Family (CMGC: 21)									
CAK1	CDC28	CTK1	FUS3	HOG1	IME2	KDX1	KIN28	KNS1	KSS1
MCK1	MRK1	PHO85	RIM11	SGV1	SKY1	SLT2	SMK1	SSN3	YAK1
YGK3									
Casein Kinases 1 Family (CK1: 4)									
HRR25	YCK1	YCK2	YCK3						
"STERile" kinases Family (STE: 14)									
BCK1	CDC15	CLA4	KIC1	MKK1	MKK2	PBS2	SKM1	SPS1	SSK2
SSK22	STE11	STE20	STE7						
Other kinases Family (40)									
AKL1	ALK1	ALK2	ARK1	ATG1	BUB1	CDC5	CDC7	CKA1	CKA2
ENV7	FMP48	GCN2	HAL5	HRK1	IKS1	IPL1	IRE1	ISR1	KIN3
KKQ8	KSP1	MPS1	NNK1	NPR1	PRK1	PRR2	PTK1	PTK2	RTK1
PSK1	PSK2	RAD53	SAT4	SCY1	SKS1	SWE1	VHS1	VPS15	YPD1
Atypical Kinases Family (15)									
BUD32	COQ8	MCP2	MEC1	PKP1	PKP2	RIO1	RIO2	SLN1	TEL1
TOR1	TOR2	TRA1	TWF1	YPL109C					

Therefore, living cells possess additional mechanisms to assure kinase specificity towards their substrates (Figure 1.1) (Ubersax and Ferrell, 2007), including: (1), docking site in kinases, more specifically on the substrates binding interaction domains in kinases, increases the local concentration of the substrate around the kinase catalytic domain (Biondi and Nebreda, 2003); (2), docking site in kinase regulatory subunit facilitates targeting the kinase to particular substrates [e.g. some cyclins representing the regulatory subunits of cyclin-dependent kinase 1

(Cdk1), also called Cdc28, contain a hydrophobic patch that recognizes an [R/K]XL motif in Cdc28 substrates (Koivomagi et al., 2011; Loog and Morgan, 2005)]; (3), conditional docking sites on substrates target the kinase to specific substrates that has been already phosphorylated [e.g. Plk1 and Gsk3 target substrates that have been already phosphorylated (Elia et al., 2003; Frame et al., 2001)]; (4), localization of kinases or their regulatory partners in different cellular compartments increases the phosphorylation rate towards substrates in the same subcellular localization as the kinase in certain conditions and limits the number of substrates accessible to the kinase in other conditions (Miller and Cross, 2000); and finally (5), adaptors or scaffolds organize the interaction between the kinase and their substrates by recruiting those substrates to the same complex (Bhattacharyya et al., 2006; Pawson and Scott, 1997). Interestingly, conditional docking motifs function as an AND logic gate representing a means by which living cells temporally organize their biochemical events with respect to each other. Similarly, subcellular localization and scaffolding mediate the spatial organization of these biochemical events. In summary, these mechanisms not only enhance kinases specificity towards their substrates, but also constitute the primary platform based on which living cells organize biological events in time and space.

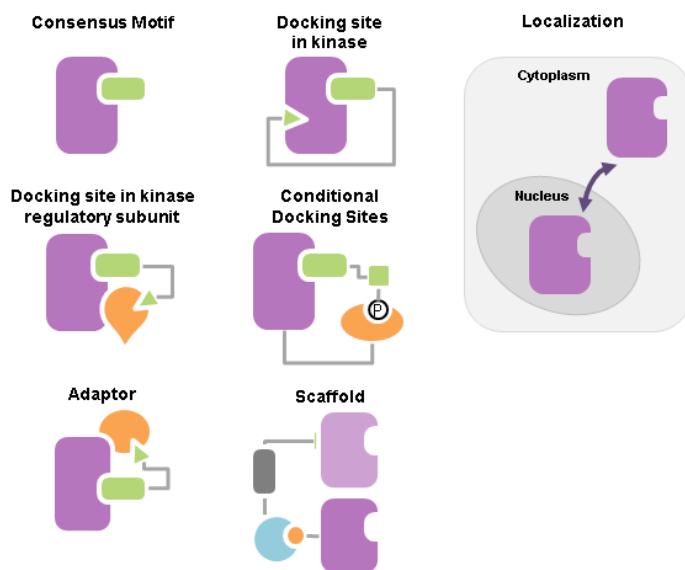


Figure 1.1. Mechanisms of kinases specificity.

The mechanisms used by kinases to enhance their specificity towards their substrates. Adapted from (Bhattacharyya et al., 2006).

1.3.2.2 Protein phosphatases

Similar to kinases, phosphatases can also be subdivided, according to the residues they dephosphorylate, mainly into two groups. The first group corresponds to the Ser/Thr-specific protein phosphatases and consists of two structurally different families: the phosphoprotein phosphatases (PPP) family and the protein phosphatases manganese or magnesium dependent (PPM) family. The second group corresponds to the protein Tyr phosphatases and contains both the protein Tyr-specific phosphatases (PTP) family and the dual-specificity phosphatases family which can dephosphorylate all three phospho-amino acids (Ser, Thr and Tyr) (Barford et al., 1998) (Table 1.II).

Table 1.II. Protein phosphatases families and their corresponding phosphatases members.

Phosphatases belonging to the different phosphatase families were determined from (Breitkreutz et al., 2010; Cherry et al., 2012; Stark, 1996).

PhosphoProtein Phosphatase Family (PPP: 17)									
CMP2	CNA1	FCP1	GLC7	PPG1	PPH21	PPH22	PPH3	PPQ1	PPT1
PPZ1	PPZ2	RTR1	RTR2	SIT4	SSU72	YNL217W			
Protein Phosphatase Metal-dependent Family (PPM: 7)									
PTC1	PTC2	PTC3	PTC4	PTC5	PTC6	PTC7			
Protein Tyrosine Phosphatase Family (PTP: 8)									
LTP1	MIH1	OCA1	PTP1	PTP2	PTP3	SIW14	YCH1		
Dual-Specificity Phosphatases Family (DSP: 5)									
CDC14	MSG5	PPS1	SDP1	YVH1					
Other Family (4)									
NEM1	PSR1	PSR2	YER134C						

Previously, phosphatases were thought of as housekeeping enzymes and phosphate erasers with no major impact on the regulation of cell signaling. Consequently, they were subject of somewhat dismissive attitude from the scientific community in comparison with the extensive attention that was expressed towards kinases. In contrast, it is now well accepted that phosphatases represent true counter-actors of kinases playing a complementary role as critical as that of kinases in controlling signal transduction (Tonks, 2013). Studies based on mathematical modeling and quantitative experimentation of signaling pathways suggest that while kinases

seem to control the amplitude of a signaling response, phosphatases tend to regulate both the amplitude and duration of these responses (Heinrich et al., 2002; Hornberg et al., 2005).

Notably despite the key role phosphatases play in the regulation of cell signaling, a quick comparison of the number of phosphatases (24 Ser/Thr-specific, 8 Tyr-specific and 5 dual-specificity phosphatases, Table 1.II) with that of kinases (125 Ser/Thr-specific, 1 Tyr-specific and 7 dual-specificity phosphatases, Table 1.I) reveals that the number of phosphatases is in general smaller than that of the kinases in eukaryotic cells and in particular, the number of Ser/Thr-specific phosphatases is much smaller than that of Ser/Thr-specific kinases in budding yeast. In addition, phosphatases showed low specificity towards substrates in *in vitro* assays (Virshup and Shenolikar, 2009). These observations led to suspect that phosphatases represent promiscuous enzymes. However, emerging evidence shows that specificity of phosphatases is enhanced independently of their catalytic domains by a wide range of regulatory subunits associated to catalytic domains of PPPs on one hand and by supplementary domains or conserved sequence motifs in PPMs and PTPs on the other hand (Shi, 2009; Tonks, 2006; Ubersax and Ferrell, 2007; Virshup and Shenolikar, 2009). Similar to kinases, phosphatases has also been observed to exhibit spatial and temporal organization in the budding yeast (Jin et al., 2008; Rossio and Yoshida, 2011).

1.3.3 Mapping phosphorylation networks

Over the past decade, significant efforts have been made to map phosphosites and to identify responsible kinases for those phosphorylation events using *in vitro* and *in vivo* HTP experiments. As mentioned previously, phosphatases have much less attracted the attention of the scientific community in comparison with kinases and consequently have rarely made the subject of such HTP experiments. Therefore, the following sections will introduce HTP methods used to identify kinases substrates.

1.3.3.1 *In vitro* experimental methods

In vitro biochemical approaches consist of incubating the kinase of interest with the proteome exposed differently in the presence of $\gamma^{32}\text{P}$ -ATP. The HTP *in vitro* approaches that significantly contributed in identifying potential kinase substrates in yeast include *in vitro* kinase assays using: (1), protein microarrays (Mok et al., 2009; Ptacek et al., 2005); (2), peptide libraries (Mah et al.,

2005; Mok et al., 2010); and (3), analogue-sensitive kinase alleles (Dephoure et al., 2005; Koivomagi et al., 2011; Loog and Morgan, 2005; Ubersax et al., 2003). These *in vitro* experimental methods made seminal contributions to identify potential substrates of the studied kinases. Each of these methods has its proper strengths and limitations.

1.3.3.1.1 In vitro kinase assays using protein microarrays

Similar to the DNA microarray concept, a large number of purified proteins are spotted on a chip, which is then incubated with a purified kinase in the presence of radiolabeled $\gamma^{32}\text{P}$ -ATP (Ptacek et al., 2005). The chips are then washed, dried and analyzed by autoradiography. Proteins that are phosphorylated *in vitro* by the kinase of interest could be identified depending on their spatial position on the chip. Kinase assays using protein chips are very practical, since one assay could be used to screen the whole proteome against a specific kinase. But, the main disadvantage of these assays is that proteins are not equally represented on the proteome chip for various reasons: missing clones, misfolded proteins and variable clones expression (Mok et al., 2011).

1.3.3.1.2 In vitro kinase assays using peptide libraries

A huge number of peptides having a fixed length and having a phosphorylation acceptor residue (Ser or Thr) in a predefined central position are generated randomly (Mok et al., 2010). Peptides are then incubated with a purified kinase and radiolabeled ATP in different wells. Subsequently, the aliquots are simultaneously spotted using a pin tool on a membrane. The membrane is then washed, dried and exposed to a phosphor screen, allowing the identification of the phosphorylated peptides. Phosphopeptides could be identified depending on their spatial position on the membrane. Peptides library techniques permitted the mapping of potential consensus phosphorylation motifs targeted by a kinase of interest. However, similar to kinase assays, peptides library techniques also possess drawbacks. Their major drawback is that they considerably depend on the bioinformatics algorithms used to characterize the phosphorylation motifs targeted by the studied kinase.

1.3.3.1.3 In vitro kinase assays using analogue-sensitive kinase alleles

The ATP-binding pocket of a kinase of interest is mutated to accept bulky radiolabeled ATP analogues that are not used by wild type (WT) kinases, producing the so-called analogue-sensitive kinases (Shah et al., 1997; Ubersax et al., 2003). In order to determine whether a protein is phosphorylated by an analogue-sensitive kinase of interest, a library of yeast strains is prepared, in which each strain expresses an open reading frame tagged with a molecular marker such as the glutathione-S-transferase (GST) tag for protein purification. The addition of radiolabeled $\gamma^{32}\text{P}$ -ATP analogue to the cell lysates of strains of this library containing an analogue-sensitive kinase of interest permits phosphorylation of the analogue-sensitive kinase substrates by radiolabeled ATP. The tagged proteins are then purified and their phosphosites are identified. Kinase assays involving analogue-sensitive kinase alleles succeeded to identify direct potential kinase substrates in an environment having physiological conditions close to those found in living cells; however, their application is limited, because not all kinases could be genetically mutated into the analogue-sensitive version and some analogue-sensitive kinase alleles lose their enzymatic capacity or their specificity towards substrates (Knight et al., 2013). Moreover, it has been shown that some WT kinases could use the analogue bulky form of ATP (Carlson et al., 2011).

Nevertheless, all *in vitro* experiments share a well-known technical limitation resulting in the generation of high false positive rates. The artificial environment in which these assays are performed lacks the cellular compartmentalization, the temporal expression profiles and the PTM of proteins met in living cells. Hence, a large number of proteins that could never represent genuine substrates of the assayed kinases in natural conditions will be phosphorylated in these experiments lacking the spatial, temporal and PTMs coordination and will be erroneously identified as kinase substrates, a misleading result that increases false positive rates of these approaches. Obviously, absence of natural context is not a problem for *in vivo* HTP experiments.

1.3.3.2 *In vivo* experimental methods

1.3.3.2.1 In vivo phosphorylation technologies based on liquid-chromatography and mass-spectrometry

Most of the *In vivo* large-scale technologies are based on mass-spectrometry (MS) to map phosphosites in cell lysates. Cell extracts are treated with phosphatase inhibitors to protect phosphoproteins from being dephosphorylated by phosphatases in cell lysates and with protease to generate a mixture of small peptides to permit their analysis by MS. Phosphopeptides are then isolated from non-phosphorylated peptides, enriched using different enrichment strategies, introduced into the spectrometer using liquid chromatography (LC), ionized and identified by comparing their fragmentation spectra measured by MS with those in primary sequence databases. *In vivo* phosphorylation technologies based on Liquid-chromatography and mass-spectrometry (LC-MS/MS) is a large-scale method renowned to detect small phosphoproteins abundance and accurately identify phosphorylated residues as well as the proteins to which they belong.

These technologies were used to quantify phosphosites in cell lysates (Chi et al., 2007; Ficarro et al., 2002) or phosphosites exhibiting changes in their phosphorylation level between WT and KP inhibited or deficient cells, one KP at a time (Bodenmiller et al., 2010; Holt et al., 2009) or between cells under normal and stress conditions. For instance, the latter approach has been used to analyze phosphorylation interactionss under different conditions, such as in cells exposed to alpha factor (Gruhler et al., 2005; Li et al., 2007), to DNA damage agents (Chen et al., 2010; Smolka et al., 2007) or to high osmolarity stress (Kanshin et al., 2015).

Obviously, *in vivo* screens guarantee that the quantified phosphorylation interactions occurred in their intrinsic environment. Yet, a major drawback of this technique is that the identified proteins showing a decrease in their phosphorylation level on the knockdown or inhibition of a given kinase could be direct substrates of this kinase, substrates of another kinase activated by the kinase of interest, or substrates of a phosphatase inhibited by the kinase of interest.

1.3.3.2.2 In vivo Protein-fragment Complementation Assay (PCA) based on the optimized yeast cytosine deaminase

Recently, a promising *in vivo* PCA based on the optimized mutant form of the reporter yeast cytosine deaminase (OyCD) has been devised to identify substrates of the cyclin-dependent kinase, Cdc28, and the cyclin regulatory subunit(s) on which the interactions between Cdc28 and its substrates depend in the budding yeast (Ear et al., 2013). The Cdc28 is a cyclin dependent kinase that with its nine cyclins, regulatory subunits, orchestrate regulation of mitotic and meiotic cell cycle events in the budding yeast. Cyclins are classified as G1 phase cyclins (Cln1, Cln2 and Cln3), S phase cyclins (Clb5 and Clb6) and M phase cyclins (Clb1, Clb2, Clb3 and Clb4) depending on the phase in which their expression peaks. They are believed to enhance Cdc28 specificity towards its substrates.

The OyCD PCA is called also the death selection assay, because the enzyme yCD reporter catalyzes the deamination of a yCD-specific prodrug called 5-fluorocytosine (5-FC) to 5-fluorouracil (5-FU). The 5-FU will then be converted by other enzymes in the pyrimidine salvage pathway to 5-FUTP, a toxic compound that causes cell death. This PCA consists of two screens. The first screen is to test the interaction between Cdc28 and a protein of interest using the OyCD PCA and the second screen retests the same interaction but in a strain lacking one of the nine cyclins. The OyCD PCA consists of splitting the yCD into two fragments F[1] and F[2] and fusing one of them to Cdc28 and the other one to the protein of interest. If Cdc28 interacts with this protein, then F[1] and F[2] will be brought into proximity allowing the OyCD reporter enzyme to fold and restore its activity causing cell death in the presence of 5-FC in the first screen. Moreover, if this interaction depends on a given cyclin, so on retesting the same interaction in cells lacking this cyclin, the interaction will not take place, consequently the reporter enzyme will not refold and cells will grow in the second screen.

In contrast to LC-MS/MS, although the OyCD PCA identifies potential direct substrates of Cdc28, it was not conceived to identify amino acids phosphorylated by Cdc28 in its corresponding substrates. However, the OyCD PCA is the first experimental method designed to identify the regulatory subunit or subunits on which the kinase depends to phosphorylate its substrates. Importantly, this PCA could be applied on any KP that depends on regulatory subunit(s) to modify their substrates. Moreover, the OyCD PCA can be performed on the level of

the proteome, because its fragments are compatible with the Gateway cloning system. Finally, this assay is affordable, because it does not require expensive reagents.

1.3.4 Protein phosphorylation and protein interaction databases

Various databases were designed to store phosphorylation data and physical interactions between KPs and other proteins (e.g. regulatory subunits, adaptors and scaffolds) including their substrates. Among those that offered a seminal contribution annotating these events, particularly in the budding yeast, we mention: the PhosphoELM database which provides access to phosphosites detected in *in vivo* and *in vitro* experiments (Dinkel et al., 2011); the PhosphoGRID database which contains information about phosphosites validated by *in vivo* low-throughput (LTP) and HTP experiments and about KPs targeting these phosphopeptides (Stark et al., 2010); the BioGRID database which stores genetic and physical PPI data inferred from HTP and LTP studies (Stark et al., 2006); and the kinase interaction database (KID) which represents the most kinase specialized repository that annotates genetic, physical and phosphorylation interactions of budding yeast kinase with other proteins, validated by LTP and HTP studies and providing details about the experimental techniques used to identify a kinase-protein interaction (Sharifpoor et al., 2011). Also, the KID database annotates kinase-protein interaction based on co-localization, chemical co-fitness and phenotypic correlation studies. Unfortunately, phosphorylation interactions are much more annotated in comparison with dephosphorylation interactions in these databases.

1.4 Analyzing biological networks

Mapping biological networks and designing databases to store their basic components provided a catalogue of KPs, their substrates as well as the PDIs associating these three entities together. Nevertheless, this catalogue of individual components is not sufficient to gain information about how KP-Nets function. We need to understand how these entities are assembled to function together in order to shape the cellular behavior and how changes in one entity of this system may affect other entities.

1.4.1 Studying biological processes using control theory

Studying the structure and the dynamic behavior of networks underlying biological processes is possible by benefiting from the principles of control theory, an interdisciplinary branch originating from engineering and mathematics. Control theory aims to study the characteristics and behavior of closed-loop control systems (Nise, 2015). A closed-loop system is a system in the form of a feedback loop (FBL) in which system inputs are determined partially by the system outputs. In its simplest forms, the closed-loop system is composed of a controller, a plant and a sensor (Figure 1.2). The controller drives a plant; the plant is the part of a system to control; and the sensor detects the output of a plant. This system receives an input, generates an output, and can be exposed to disturbance. The difference between the input and the measured output is called the error (Figure 1.2). This error drives the plant to make a correction so that the output will become equal to the input. If this error is zero, then the system does not drive the plant, because the plant is producing the desired output. The system detects the disturbance and corrects for it via the closed-loop, and thus yields the correct output. Various examples in everyday life and in human and other organism physiology represent closed-loop control system (e.g. heating systems, elevators, radar antenna, bacterial chemotaxis, pancreatic secretion of insulin and other hormones, and FBLs in transcription regulatory networks and signaling networks). For instance, in a heating system, the input is a position on a thermostat, the output is the heat, the sensor is a thermostat, the plant is the heater, and the controller is composed of fuel valves and an electrical system to control these valves. As long as a room temperature does not correspond to the desired temperature, the heater will be activated and when the position on the thermostat corresponds to the measured temperature in the room, the heater will stop.

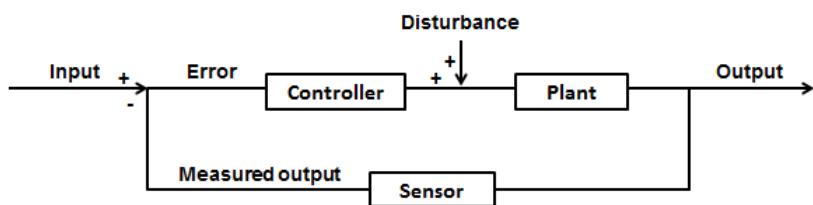


Figure 1.2. A closed-loop control system.

The arrangement of the main components (the controller, the plant, and the sensor) of a control system in a simple closed-loop control system.

Given the conceptual similarities between the engineering and cellular regulatory mechanisms, the tools used to model engineering closed-loops are used to analyze biological networks. FBLs in signaling pathways represent an example of a closed-loop control system. For instance, G-protein coupled receptor signaling in eukaryotic cells represents a FBL responding to a wide range of stimuli (e.g. pheromone α -factor and presence or absence of nutrients in the budding yeast, sense of smell in mice, and hormones and neurotransmitters in human) (Alberts et al., 2002). The G-protein is composed of three subunits α , β , and γ . In the inactive form of the G-protein, its α -subunit binds GDP. The first step in the activation of the G-protein consists of the binding of a ligand to a transmembrane receptor, causing the α -subunit to bind to the ligand-receptor complex. This binding causes a conformational change in the α -subunit, leading it to release GDP, bind GTP, and to dissociate from the $\beta\gamma$ subunits. The GTP-bound α -subunit activates a downstream response, then converts the GTP to GDP, and at the end, the G-protein three subunits are re-associated, closing thus the FBL (Ingalls, 2013).

Each of the described interactions in the G-protein signaling mechanism or in a similar regulatory mechanism can be represented in the form of mathematical formulas such as differential equations (Yi et al., 2003). The terms of these formulas describe the rate of change of the abundance of each component of the model. This approach can be applied to various categories of biological processes such as metabolic, transcription regulatory and signaling pathways (Hynne et al., 2001; Perkins et al., 2006; Yi et al., 2003). Modelling biological processes serves to study different types of the dynamic responses of these processes (graded response, ultrasensitivity (switch-like response), bistability, and multistability) (Szomolay and Shahrezaei, 2012). Moreover, control theory provides a theoretical basis to conceive reverse engineering approaches aiming at determining the structure of regulatory networks (Khammash, 2008; Perkins et al., 2006). Details about the modelling approaches and the nature of the studied responses are outside the scope of this thesis. For a review, refer to (Fischer, 2008; Iglesias, 2013). Although, mathematical modelling yields important details about the structure and dynamic behavior of networks underlying biological processes, it can be computationally expensive. Mathematical modelling, however, can be fruitfully applied to networks composed of a limited number of nodes (in the order of tens). Other approaches are needed to describe the structure and dynamics of larger networks and to help shed light on how these networks function, though limited in detail compared to mathematical modelling.

The subsequent sections introduce how studying organizational structure of biological networks could enhance our understanding about how biological networks work. This involves: (1), assessing networks topological properties (local properties, global properties and logic motifs); (2), classifying networks into categories according to the assessed properties they share with known network models; (3), measuring network hierarchical level; and (4), elucidating their hierarchical structure in case their hierarchical level is perceptible. In the following sections, I will introduce these approaches as well as the four network models that provided valuable insights for understanding and describing diverse real-world networks.

1.4.2 Assessing network topological properties

The charted cellular networks that have been annotated in various databases require a mathematical framework for representation and analysis purposes. The graph representation and graph theory principles constitute a universal and efficient computational tool to study different biological networks. A graph is a set of nodes that are connected by edges. Nodes correspond to molecules such as genes, transcripts, proteins, metabolites or small molecules and edges designate the interactions between these entities. Edges could take the form of arrows or simple links depending on whether node interactions are directed or not, defining directed and undirected networks, respectively. For instance, PPI networks represent undirected networks, because if node u interacts with node v , this reciprocally means that node v interacts with node u . In contrast, KP-Nets represent directed networks, because if node u (de)phosphorylates node v this does not imply that the inverse is also true.

Representing biological networks as a graph composed of a set of nodes and edges permitted the study of biological networks from a topological perspective. Watts and Strogatz were among the first to suggest that networks topology affects networks function, supporting their claims by finding that the infectious diseases spread and information spread are affected by the social network structure and that robustness of power transmission is affected by the power grid topology (Strogatz, 2001; Watts and Strogatz, 1998). From then, topological properties of networks were extensively studied to characterize biological networks in an attempt to infer functional principles of these networks. This discipline encompasses assessing a myriad of metrics that have been proposed to describe biological networks on both local and global levels. As their names indicate, the former category characterizes local network structures, whereas the

latter one describes global network features. This approach also involves identifying the under- and over-representation of logic motifs which represent subgraphs of the network.

1.4.2.1 Local topological properties of networks

In what follows, we define a network G by a set of nodes V and a set of edges E . The number of nodes in G is $n=|V|$ and of edges is $m=|E|$. Among the local metrics that are commonly studied, we focused on: the node degree, the node clustering coefficient, the shortest path between two nodes and the node betweenness centrality (Kantarci and Labatut, 2014).

The ***node degree*** is the number of edges connected to this node. In directed networks, the node degree consists of the sum of its in- and out-degrees (k_{in} and k_{out}), which correspond to the node in- and out-going edges, respectively.

The ***clustering coefficient***, C , of a given node, called also the node transitivity, measures the probability that neighbors of this node are also neighbors. Neighbors of a node are nodes that are directly connected to the node in question. Formally, the clustering coefficient of a node u is equal to the ratio of the number of triangles actually connected to u and the number of all possible triangles that could be centered on u (Wasserman and Faust, 1994).

$$C(u) = \frac{2e_u}{k_u(k_u - 1)}$$

Where e_u is the number of edges actually linking the neighbors of the node u and k_u is the degree of the node u .

The ***shortest path*** between two nodes is the minimal set of consecutive edges that should be traversed to go from one node to another. The length of a shortest path represents the number of edges in the shortest path.

The ***betweenness centrality*** of a given node quantifies the number of shortest paths going through this node and is calculated using the following formula:

$$\text{Betweenness}(v) = \sum_{u \neq v, u \neq w, v \neq w} \left(\frac{\sigma_{uw}(v)}{\sigma_{uw}} \right)$$

Where $\sigma_{uw}(v)$ is the number of all shortest paths going through the node v and σ_{uw} is the number of all shortest paths in the network. This metric assesses the load of traffic going through a given node assuming that the information flow is primarily mediated via shortest available paths.

1.4.2.2 Global topological properties of networks

The most commonly used properties to provide a global overview of complex networks topology include: the network degree distribution, the network average degree, the network average path length, the network diameter and the network clustering coefficient.

The ***network degree distribution***, $P(k)$, represents the probability that a node of the network has a degree k . We described below some of the degree distributions of various network models and how they contribute in shaping the network architecture (Paragraph 1.4.3). Various biological networks have been shown to be scale-free having a power-law degree distribution (Albert, 2005; Barabasi and Oltvai, 2004; Jeong et al., 2000), an observation that has been debated and rejected for certain networks (Khanin and Wit, 2006).

The ***network average degree*** is simply the average of all nodes degree in the network and it assesses the density of connections or the so-called connectivity of the network. Although most of the biological networks are thought to be dense, their average degree is much smaller than what is expected (Kantarci and Labatut, 2014). This might return to the considerable amount of missing data in these networks.

The ***network average path length***, $\langle L \rangle$, quantifies the average of the shortest paths length between all pairs of nodes in the network. It represents the number of edges that the information needs on average to traverse from a source node to reach a target node.

The ***network clustering coefficient***, $\langle C \rangle$, is simply the average of the clustering coefficient of all nodes in the network (Paragraph 1.4.2.1). This metric reflects the cliquishness level of the network, in other words the probability of the presence of clusters made of three nodes in the network. Moreover, a high network clustering coefficient is an indicator of the small-world property introduced in (Paragraph 1.4.3.2).

The ***network diameter***, D , is the maximum of the shortest path lengths between any pair of nodes in the network. This property measures the maximal number of edges needed by the information to traverse from a source node to attain its target.

1.4.2.3 Network logic motifs

Network logic motifs represent recurring patterns of interconnections that occur much more frequently in real-world networks in comparison with random networks and this independently of the network size (Milo et al., 2002). They are considered to serve as basic building blocks of a network; they have different shapes (Figure 1.3) and their under- and over-representation are thought to differ depending on the network functional role. Identifying the set of over-represented logic motifs in a given network could unravel valuable information about structural design principles of the studied network. To determine whether a given logic motif is over-represented in a network, the number of occurrences of this logic motif in the network of interest is compared with that in a random network having the same node in- and out-degrees as the studied network.

It is generally accepted that networks having similar roles share the same set of enriched logic motifs and that those having distinct functions differ in their set of enriched logic motifs. For instance, it has been shown that the over-represented motifs in the world wide web (WWW), designed to provide short paths between related pages, are different from those in the food web networks, permitting a flow of energy from bottom to top food chains, which are both different from those in the TF-Net and neural connectivity network specialized in information processing (Milo et al., 2002). To further support this idea, the authors of the same study showed that although eukaryote cells are considerably different from prokaryote cells, their TF-Nets are enriched for the same set of logic motifs (FFL and bi-fans).

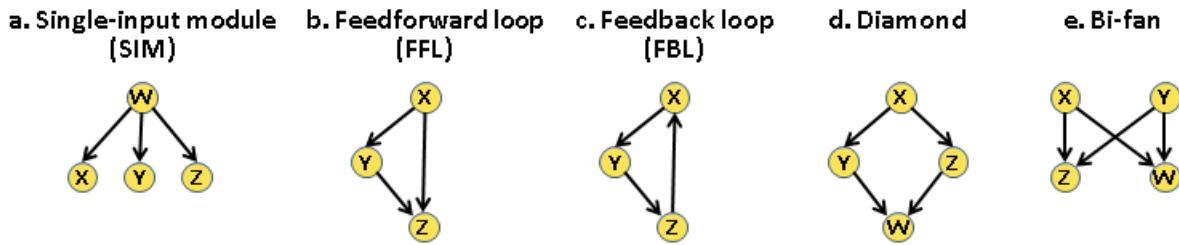


Figure 1.3. Different types of network logic motifs.

(a) Single-input module (SIM) represents a node controlling more than one node. (b) Feedforward loop (FFL) corresponds to a node x that regulates a node y and together they co-regulate a node z . (c) Feedback loop (FBL) designates a complete ring in which each node regulates the other consecutively, in this figure a FBL made of three nodes is shown, but FBLs could be made of only one node or more than one node. (d) Diamond stands for a node x that regulates two nodes y and z and the two latter nodes co-regulate another node w . Diamond is a multilayer motif that could be made of three layers, as illustrated here, or of more than three layers in which nodes y and z regulates respectively two other nodes which co-regulate the w node. (e) Bi-fan represents two upstream nodes x and y that co-regulate two downstream nodes z and w .

1.4.3 Network models

Other than a mathematical framework for studying cellular networks, research in this field needed models to explain the emergence and behavior of these complex systems. Many network models had been designed; four of which had a direct impact on our understanding of biological networks: the Erdős-Rényi model, the Watts-Strogatz model, the Barabási-Albert model and the Ravaz-Barabási model (Figure 1.4). These network models do not necessarily share all, but they share some of the previously defined topological properties with many real-world networks, as it will be shown in the subsequent sections.

1.4.3.1 Erdős-Rényi model - Random networks

Among the earliest proposed probabilistic graph models, Erdős-Rényi random graph was introduced as a network model to generate random networks. This model assumes that for a given number of nodes, n , in a network, each pair of nodes has a constant and independent probability of being connected. This generates a homogeneous network made mostly of nodes having approximately the same number of interactions, k , called the node degree (Erdős and Rényi, 1960) (Figure 1.4a, upper panel). Nodes degree of Erdős-Rényi networks follows a Poisson distribution peaking strongly at the mean of k , $\langle k \rangle$; $P(k) \sim e^{-k}$, for $k \gg \langle k \rangle$ (Figure 1.4a,

middle panel). Erdős-Rényi networks are characterized by a small average path proportional to $\log(n)$ and a clustering coefficient independent of node degrees (Figure 1.4a, lower panel).

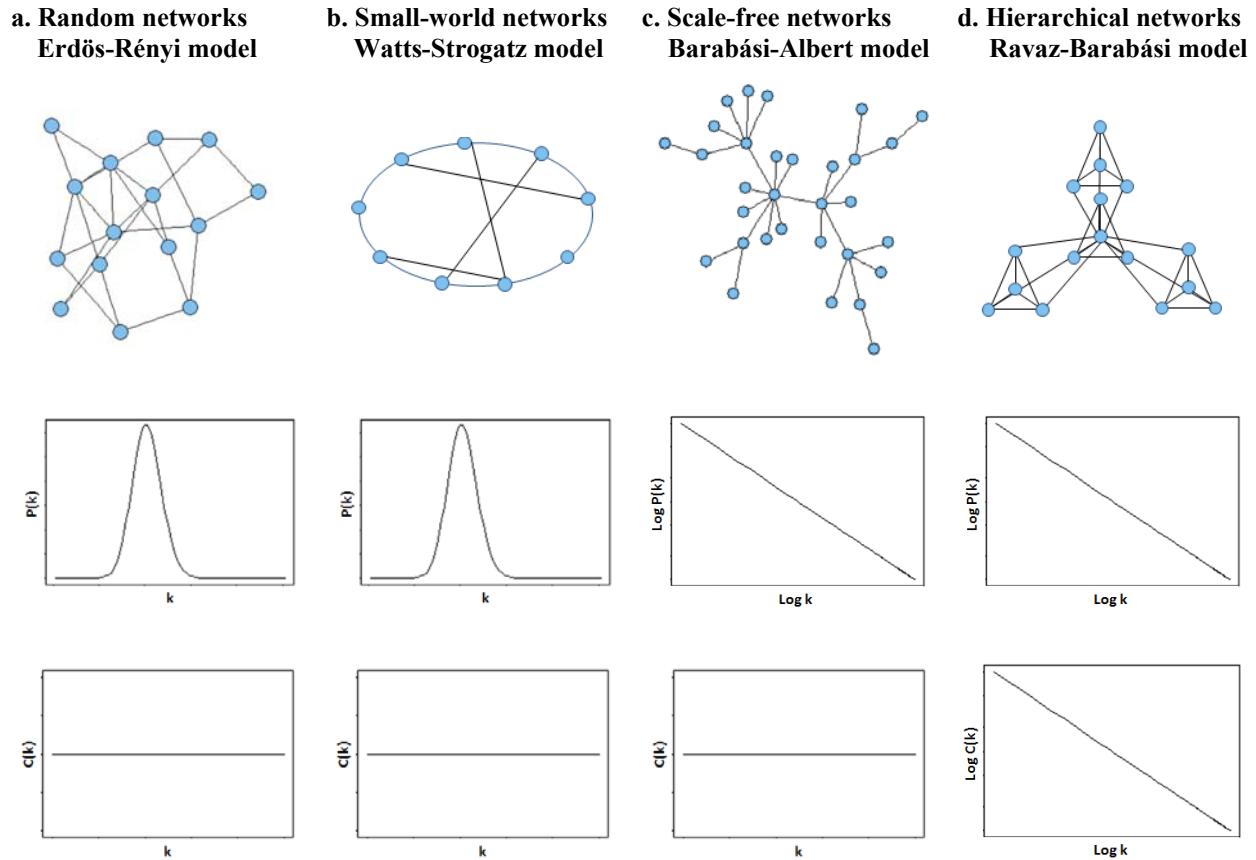


Figure 1.4. Network models.

Illustration of the four network models: **(a)** random networks, **(b)** small-world networks, **(c)** scale-free networks and **(d)** hierarchical networks (upper panels), their nodes degree distribution (middle panels) and the scaling of their nodes clustering coefficient in function of their nodes degree (lower panels). Adapted from (Barabasi and Oltvai, 2004).

With the growing interest in complex systems and the emergence of the Erdős-Rényi model, a relevant question was to investigate whether real-world networks match this model and whether they are fundamentally random. Empirical studies of different networks, such as the WWW, Internet, social networks and various biological networks, showed that the Erdős-Rényi model failed to efficiently describe real-world networks (Barabasi and Albert, 1999; Watts and Strogatz, 1998). These findings necessitated the suggestion of another model that better describes these networks.

1.4.3.2 Watts-Strogatz model - Small-world networks

The small-world model was designed by Watts and Strogatz to describe a wide range of networks having a small average path length and a high clustering coefficient (Watts and Strogatz, 1998) (Figure 1.4b). These networks lie between two extremes of connection topology found in complete regular and complete random networks. The construction algorithm begins with a complete ring having n nodes each linked to k neighbors. Then shortcuts between the ring nodes are created by rewiring some edges to a node chosen at random with a probability p . When p approaches 0, the model matches a regular ring and when p approaches 1 the model matches a complete random network. In a specific region between 0 and 1, this algorithm could create networks retaining a short average path length as in the Erdős-Rényi model, with accounting for a high clustering coefficient that is missing in Erdős-Rényi networks.

The Watts-Strogatz model succeeded in partially explaining the small-world phenomenon (Frigyes, 1929; Milgram, 1967), also known as the six degrees of separation. This phenomenon suggests that any pair of persons in the world could be linked to each other through at most six acquaintances (Guare, 1990). Watts and Strogatz showed the validity of the small-world phenomenon to various real-world networks such as the actors' collaboration in films, the neural network in the nematode worm *C. elegans* and the electrical power grid of the western United States (Watts and Strogatz, 1998). Nevertheless, networks observed in the nature still differ from those generated by the Watts-Strogatz model mainly by having hubs.

1.4.3.3 Barabási-Albert model - Scale-free networks

Most of real-world networks are heterogeneous; their nodes are far from having similar degrees. Few nodes are highly connected forming what is known as hubs, whereas the majority of the remaining nodes are less connected (Figure 1.4c, upper panel). To account to this property, Barabási and Albert proposed a model characterized by a power-law degree distribution: $P(k) \sim k^{-\gamma}$ where γ is the degree exponent and k represents the nodes degree (Barabasi and Albert, 1999) (Figure 1.4c, middle panel). To form a network having a power-law degree distribution, the algorithm begins with a small number of nodes m_0 and adds one node at a time with m ($m \leq m_0$) links that connect to the already existing nodes with a probability proportional to their node degrees. It is this probability expressed in function of existing node degrees that leads to the appearance of hubs in scale-free networks, coining what is known as the preferential attachment

phenomenon. Another essential property of networks generated by the Barabási-Albert model is their open and growing nature encountered in real-world networks that are constantly expanding over time (e.g. WWW, the author publications, the actors' collaboration in movie networks). Scale-free networks are also characterized by a short average path length proportional to $\log(\log(n))$ (Cohen and Havlin, 2003) and a node clustering coefficient independent of node degree (Figure 1.4c, lower panel). Finally, a well-known consequence of a scale-free architecture in networks is their robustness against random errors represented by random removal of network nodes, coupled with their fragility against hubs targeted attacks (Albert et al., 2000).

1.4.3.4 Ravaz-Barabási model - Hierarchical networks

Although the Barabási-Albert model succeeded in describing real-world networks, this model was still unable to account for the modularity property distinguishing these networks. A network is said modular, if it could be partitioned into different functional modules (Hartwell et al., 1999; Lauffenburger, 2000). Practically, network modules represent a set of nodes that are highly interconnected and more interconnected in comparison with other nodes. For instance, modules in social networks correspond to coworkers or groups of friends (Granovetter, 1973), modules in the WWW correspond to communities sharing interests (Flake et al., 2000) and modules in PPI networks correspond to protein complexes (Dittrich et al., 2008). To account for the simultaneous existence of hubs and modules in real-world networks, Ravaz and Barabási proposed a new model to which the investigators refer as the hierarchical network (Ravasz and Barabasi, 2003; Ravasz et al., 2002) (Figure 1.4d). This network begins with four nodes extensively connected forming a cluster or a module. The latter module is replicated three times and peripheral nodes of each of the replicates are connected to the central node of the first existing module, obtaining a new module made of sixteen nodes. The latter step could be repeated indefinitely. The architecture of the created hierarchical network is characterized by integrating modular structure with the scale-free topology. This hierarchy, although obvious on visual inspection when the network is hierarchically represented as in (Figure 1.4d, upper panel), it could be further understood by noticing the integrated way by which the replicates are assembled to the original module. Ravaz and Barabási suggested that a unique signature of hierarchical networks is the inverse relationship between node degrees and node clustering coefficients in these networks, an unshared feature with the network models described above

(Figure 1.4d, lower panel). But, this signature was later contested by Hao et al. who showed that this inverse relationship between nodes degree and clustering coefficient is due to the modularity and not to the hierarchical structure of the network (Hao et al., 2012).

Since the hierarchical structure could be encountered in many real-world networks especially directed ones having the command-execution organization and since the signature proposed by the Ravaz-Barabási model has been refuted (Hao et al., 2012), other methods aiming to measure and elucidate networks hierarchical structure were needed. Before presenting those methods in the subsequent sections, the types of network hierarchical structures will be introduced first.

1.4.4 Studying hierarchical structures in networks

1.4.4.1 Types of network hierarchical structures

A network is considered hierarchical if its nodes could be ordered on the basis of a given criterion depending on the context of interest. A hierarchy can be classified into three different types depending on how they are defined. First, the order hierarchy is a structure in which elements are ordered based on the values of a variable defined on a set of elements such as size, complexity or other properties (Figure 1.5a). Second, the inclusion or nested hierarchy represents a structure in which elements are ordered into nested clusters of other elements such as “Chinese boxes” (Simon, 1973) (Figure 1.5b). Finally, the flow or level hierarchy is a hierarchy in which elements of a network are classified into different levels so that elements in upper levels control those in lower levels (Lane, 2006; Mones, 2013) (Figure 1.5c). To understand how KP-Nets, representing the subject of this thesis, process information and how information flows in these networks, we are definitely interested in the latter category of hierarchies.

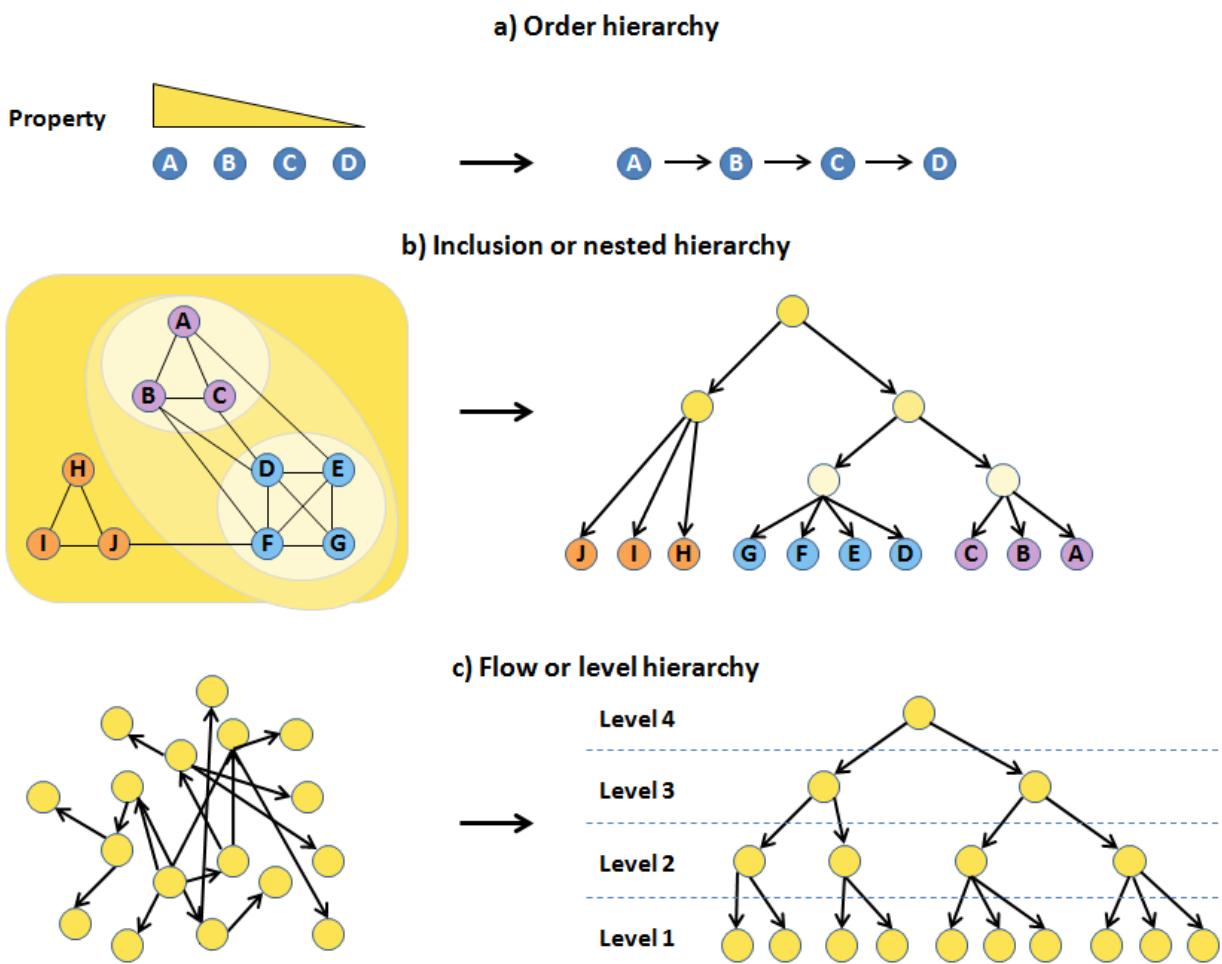


Figure 1.5. Different types of network hierarchical structures.

(a) the order hierarchy in which elements are ordered according to the values of a property characterizing these elements, (b) the inclusion or nested hierarchy in which elements are arranged into nested clusters in which a new node is created to represent each cluster and (c) the flow or level hierarchy in which nodes are classified into different levels. Adapted from (Mones, 2013).

A flow or level hierarchy occurs when the network contains nodes having different influential levels on each other and consequently on the whole network. Assessing the nodes influence was quantified based on various concepts, but the bigger the difference is between the most and least influential nodes, the more hierarchical the network is. This influence level ranges from strong to weak. Some examples of nodes with high influence include: managers in a company (Rowe et al., 2007), group leaders in a terrorist group (Memon et al., 2008) or group leaders in birds travelling together (Nagy et al., 2010), proteins having maximal influence over

others in different regulatory networks (Bhardwaj et al., 2010b; Ma et al., 2004; Yu and Gerstein, 2006) and opinion leaders in blogospheres (Song et al., 2007).

1.4.4.2 Measuring network hierarchical level

Various metrics have been proposed to quantify the hierarchical level or the amount of hierarchy in a flow network. A non-exhaustive inventory of these metrics is provided with a briefly description below.

Carmel et al. conceived an algorithm to graphically illustrate directional networks (Carmel et al., 2004). This algorithm is based on the Sugiyama method which associates x and y coordinates to each node of the network (Sugiyama et al., 1981). While x coordinates are determined to produce nice graphical representations, y coordinates reflect the network hierarchy and are calculated by minimizing a simple energy function. The hierarchy amount in this model corresponds to the difference between the maximum and minimum y coordinates normalized by the network diameter.

other efforts considered a network to have a maximal hierarchy such as in complete trees and quantified its hierarchical level by only penalizing the proportion of edges forming loops or multiple edges in the network (Krackhardt, 1994).

Another study based on the concept of hierarchical path assessed the hierarchical level of a network by measuring the fraction of shortest paths that are hierarchical (Trusina et al., 2004). A path between two nodes is considered hierarchical, if there is an up- or down-path between these two nodes. An up- and down-path is a path in which you can go from a node u to a node v if $k_u \leq k_v$ and $k_u \geq k_v$, respectively; where k_u represents the degree of the node u .

Recently, Mones et al. defined a new metric called the global reaching centrality (GRC) to quantify the hierarchy amount of a network (Mones, 2013; Mones et al., 2012). GRC is a global property measuring the normalized sum of the differences between the highest local reaching centrality of all nodes and the local reaching centrality of each node in the network. The local reaching centrality of a node u is equal to the proportion of nodes in the network accessible from the node u through its out-going edges.

Finally, the latest defined metrics to quantify the degree of hierarchy of a network is the hierarchical score (HS) which represents the ratio of the number of downward arrows to that of

upward arrows balanced by the number of horizontal arrows (Cheng et al., 2015). Downward, upward and horizontal arrows represent arrows pointing from node in upper levels to those in lower ones, arrows pointing from nodes in lower levels to those in upper ones, and arrows linking nodes in the same level, respectively. Networks having a large number of levels tend to have a high HS, hence HS of networks differing in their number of levels could not be directly compared. To overcome this limitation, a corrected HS was defined to take into consideration not only the observed, but also the expected number of downward, upward and horizontal arrows.

In exception of the corrected HS, the more the previously defined metrics quantifying network hierarchy amount are closer to 1, the more the network is considered hierarchical. As for the corrected HS, its value changes between 1 for complete random networks and ∞ for networks having highest hierarchy scores such as complete trees. The variety of existing concepts used to measure the networks hierarchical level might return to the multitude of definitions and methods used to approach hierarchical networks. But, we think that GRC is the measure that reflects better the flow of information in hierarchies, because it reflects the contribution of nodes to control the network through their accessibility to other nodes. Moreover, Mones et al. noted that some metrics measuring network hierarchical levels depend on certain parameters that could not be defined in all types of networks such as those proposed in (Carmel et al., 2004; Rowe et al., 2007) and other metrics are applied on exclusively undirected or directed networks but not on mixed ones (Carmel et al., 2004; Krackhardt, 1994; Trusina et al., 2004). These disadvantages do not apply for the GRC metric.

1.4.4.3 Elucidating network hierarchy

The hierarchical structure of a network could be elucidated by employing two approaches: process approaches and structural approaches (Lane, 2006). Process approaches unravel how information is processed within the network hierarchy using bottom-up methods, whereas structural approaches order nodes in the network hierarchy using up-bottom methods used mainly for representation purposes. Many methods, called decomposition algorithms, were devised to elucidate networks hierarchical structure. These algorithms consist of classifying nodes in multiple levels or layers. Seven algorithms from the process approaches are worth outlining:

1) Leaf-removal represents an iterative algorithm that consists of removing leaves of the network as well as their connected edges, in each iteration, till the network is completely decomposed (Ma et al., 2004). Leaves of a network are nodes having an out-degree of zero. The first removed leaves are placed in the level 1. The removed leaves in each subsequent step are classified in the next higher levels (the previous level incremented by 1). Since the leaf-removal begins sorting nodes from bottom to top, it is known as a bottom-up algorithm.

2) Breadth-first search (BFS) is also a bottom-up algorithm which first classifies nodes that do not regulate other regulators or those that regulate only themselves into the bottom level (level 1) (Bhardwaj et al., 2010b; Yu and Gerstein, 2006). Subsequently, the BFS algorithm assigns to each unclassified node a level equal to the length of the shortest path between this node and any node in the bottom level. This algorithm could result in inconsistent ordering of nodes within the network hierarchy, in particular for those making part of FFLs.

3) Hierarchical regulatory networks organization (HiNO), an improved version of the BFS algorithm, implements a recursive correction approach to resolve the conflicting assignment of nodes level in the network hierarchy (Hartsperger et al., 2010). This problematic is associated only to nodes making part of FFLs. The correction is made of two steps: First, a downgrade step that consists of classifying each node in the least level to which any of its regulators belong. Second, an upgrade step consists of assigning each node, not having any regulator and having all of its targets placed in the same level, to the next higher level.

4) Intuitive hierarchy is a simple method based on nodes in-degrees, in which nodes are classified into three layers as follows: (1), the top layer contains nodes having a zero in-degree, in other words regulators that are not regulated by other regulators; (2), the core or middle layer is composed of nodes having a non-zero in-degree and regulating other regulators; and (3), the bottom layer contains nodes also having a non-zero in-degree, but not regulating any regulator (Bhardwaj et al., 2010a).

5) Hierarchical decomposition of regulatory networks (HIDEN) transforms the problem of discovering the underlying hierarchical structure of a network to a mixed integer programming problem for networks having less than 100 nodes (Gulsoy et al., 2012). The HIDEN algorithm uses the divide and conquer strategy for bigger networks in order to generate networks having smaller sizes, and then resolves the mixed integer programming problem for

each of these generated partitions apart. mixed integer programming is an optimization method aiming at maximizing or minimizing a linear function of a mixture of discrete and continuous variables that respect linear constraints (Garfinkel and Nemhauser, 1972). The mixed integer programming in the HIDEN algorithm consists on minimizing a penalty function that penalizes edges having an upward or a horizontal direction in the hierarchy.

6) The Vertex-Sort (VS) algorithm classifies network nodes into various levels by: (1), transforming the network into an acyclic graph; (2), applying the leaf-removing algorithm on the network and on its transpose; (3), merging the results of the two leaf-removing applications into a global result, in which the level of each node is determined by the possibility of a node to span many levels and (4), grouping nodes into three non-overlapping layers: the top, core and bottom layers (Jothi et al., 2009). The core layer is made of the nodes composing the biggest strongly connected component (SCC) of the network and the top and bottom layers contain nodes that regulate and are regulated by the core layer, respectively. Transforming a network from a cyclic to an acyclic graph is achieved by collapsing each SCC into a super-node. Transposing a graph consists on inverting its edges direction. An SCC is a graph in which between each pair of nodes u and v there is a path from u to v and from v to u . The VS algorithm represents a global algorithm, defined as a combination of bottom-up and up-bottom algorithms, since it applies the leaf-removing algorithm on the graph and on its transpose.

7) The hierarchical score maximization (HSM) algorithm is a simulated annealing algorithm that aims at maximizing the HS introduced previously (see Paragraph 1.4.4.2) and conceived by (Cheng et al., 2015). The HSM procedure consists of: first, randomly assigning the different nodes of a network to a predefined number of levels and calculating the network HS. Second, randomly selecting a node, assigning it to another level and recalculating the network HS. If the new HS is larger than the previous one, the new hierarchical structure is accepted with a probability depending on the previous HS, the current HS and a global parameter T called temperature. If not, the new hierarchical structure is rejected. Third, the second step is repeated a p number of times. Finally, all the previous steps are repeated 1000 times in order to select the network hierarchical structure having the highest HS.

A comparison between the application of the different sorting algorithms on a toy network showed that the VS algorithm outperforms other algorithms (Jothi et al., 2009). First,

the leaf-removing algorithm cannot be applied on a cyclic graph, because in a cycle there are no leaves and consequently the algorithm will not be able to decompose the network (Figure 1.6a). Second, the BFS and the HIDEN algorithms could generate wrong classifications of certain nodes causing many arrows to point upwards (colored in red) instead of downwards (Figure 1.6b,d). Third, all algorithms except the leaf-removing and the VS algorithms are not scalable, because on the addition of a new node to the sorted network, the relative order between old nodes could be altered (Figure 1.6b-e). For instance, by adding a node I to Figure 1.6b having an in-coming arrow from the node D and an out-going arrow to the node H, a new shortest path of length two will be created between nodes D and H. Consequently, the node D will be placed in the same level as the other nodes A, B and E which were previously in a lower level. Finally, none of the algorithms, except the VS algorithm, produces a global solution, a unique result that stems from sorting the network from bottom to top and *vice versa* (Figure 1.6f). This comparison did not cover the HSM algorithm, due to the unavailability of its code for the public use.

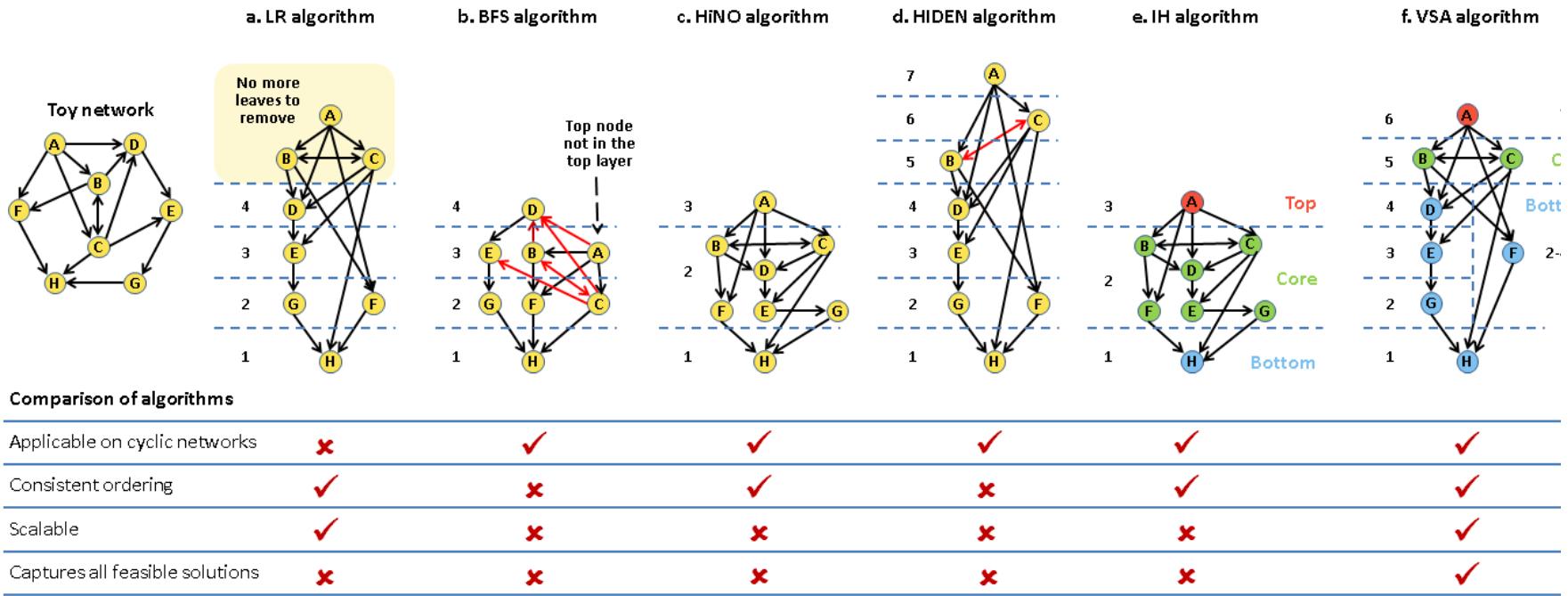


Figure 1.6. Application of the different sorting algorithms on a toy network and a comparison of their performance.

Application of **(a)** the leaf removal (LR) algorithm, **(b)** the breadth-first search (BFS) algorithm, **(c)** the hierarchical regulatory networks organization (HiNO) algorithm, **(d)** the hierarchical decomposition of regulatory network (HIDEN) algorithm, **(e)** the intuitive hierarchy algorithm and **(f)** the vertex-sort (VSA) algorithm on a toy network. Adapted from (Jothi et al., 2009).

1.5 Correlation between topological properties and biological functions

The approaches introduced previously, that consist of assessing networks topological properties and architectural structure, were extensively used to get insights into biological roles of these networks and of their components. But, to assess the efficiency of these approaches in reaching their objective, we need to know to what extent structure determines biological functions and how well topology alone predicts biological functions. Actually, this question has been previously addressed on the level of network nodes, network building blocks (logic motifs) and even on the level of the networks themselves.

1.5.1 Do node topological properties predict node biological functions?

Based on the idea that nodes having similar topological signatures tend to possess similar biological characteristics and hence perform similar biological functions, many efforts have been made to unravel biological roles of these network entities. For instance, cancer proteins have been shown to exhibit higher connectivity and centrality in comparison with non-cancer proteins (Jonsson and Bates, 2006). Nevertheless, proteins causing cancer do not always exhibit such topological properties, as they do not always represent hubs (Goh et al., 2007). Moreover, proteins having similar topological properties and/or neighborhoods in a PPI network are assumed to have similar biological traits or function and to belong to the same protein complex (Guerrero et al., 2008; Milenkovic and Przulj, 2008; Winterbach et al., 2013). Furthermore, clusters enriched for proteins involved in cancer were successfully identified by comparing node neighborhoods topological properties in a PPI network (Milenkovic et al., 2010).

What stands out most from these efforts is that topology does encode certain amount of information about node biological functions; however, the reliability solely on topological properties in inferring biological significance of protein functions in networks is questionable and could be misleading (Hakes et al., 2008). Actually, biological networks remain unrepresentative of the underlying studied systems, due mainly to missing and noisy data and to the dynamic nature of these systems changing in time and space, two criteria that are poorly represented in the mapped biological networks at the present time. Hence, more sophisticated approaches aiming at integrating biological data proved to better tackle these problems. For

instance, overlaying biological data, such as DNA neighborhood, gene expression pattern, sequence similarity, gene ontology (GO) and knowledge of functional or disease modules of proteins, on top of PPI networks, representing the topological facet of the problem, had substantially increased success rate of valid disease genes prediction (Aragues et al., 2008; Lage et al., 2007; Oti and Brunner, 2007; Oti et al., 2006; Wu et al., 2008).

1.5.2 Does logic motif topology predict logic motif biological functions?

Advances in synthetic biology permitted many studies to investigate biological functions of various logic motifs. For example, SIM were associated to temporal expression coordination of a set of genes sharing similar functions (Ronen et al., 2002); FFLs were linked to persistence detection, pulse generation, response acceleration and sign-sensitive delays (delay in response based on the activation state of the node x in FFL, Figure 1.3b) (Basu et al., 2004; Mangan et al., 2003); FBL were related to switch between two steady states, promotion of signal persistence, information storage devices and limitation of signal propagation (Bhalla and Iyengar, 1999; Sneppen et al., 2010); diamonds were associated to degradation of performance upon loss of components and generalization of information from partial signals (Alon, 2007; Hertz et al., 1991); and bi-fans were linked to temporal regulation of signal propagation, signal sorting, filtering and synchronization (Alon, 2007; Lipshtat et al., 2008).

Although the topology of a logic motif determines to a certain extent the set of biological tasks this motif could perform, the observed diversity of biological functions associated to each type of the logic motifs results from a combination of three features: (1), The biochemical parameters characterizing each interaction making the logic motif. (2), The sign (positive or negative) of each edge of the logic motif represents the effect the regulator exerts on the regulated entities in the logic motif. Positive/negative effect corresponds to expression/repression of the regulated gene in the TF-Net and activation/inhibition of the regulated protein in the KP-Net respectively. (3), The logic functions (e.g. *AND* gate and *OR* gate) determining the output of each logic motif. Motifs having an *AND* gate respond when all conditions are in favor of a response, whereas those having an *OR* gate respond when at least one of the conditions is in favor of a response. For instance, an *AND* gate in a FFL in a KP-Net necessitates that both KPs (x and y) in the FFL act on the third KP z , whereas an *OR* gate requires that either KPs (x or y) in the FFL acts on the third KP z , to produce a response (Figure 1.3b).

In consequence, topology of logic motifs could be associated to a set of functions, but when logic motifs topology is coupled with biological parameters (represented in the three features mentioned above), they both specify the biological function of a logic motif.

1.5.3 Does network topology predict network biological functions?

Addressing this question returns to investigate whether networks that share global topological properties have the same biological function. Certain biological networks have been associated with well-defined biological functions. For instance, it has been shown that signaling networks and transcriptional networks are dedicated to information processing. Signaling networks do not represent infrastructures that only receive signals from outside of the cell via its surface receptors and transduce them to the inside but also, they process and integrate these signals and take appropriate decisions according to the cellular environment to preserve homeostasis (e.g. TOR signaling pathway, osmotic stress signaling pathway) (Bhalla and Iyengar, 1999; Bray, 1990, 1995; Helikar et al., 2008). As for transcriptional networks, they are responsible for controlling gene transcription and in part protein production based on the cell environment (Milo et al., 2002; Santini et al., 2012; Shen-Orr et al., 2002). Other biological networks such as PPI networks have not been associated to a well-defined biological role yet, other than its basic function - forming protein complexes. This might be because the PPI network is implicated in a staggering number of biological processes (signal transduction, protein folding, transport, transcription, translation and PTMs) (Royer, 1999). Disregarding the fact that the biological role of certain networks is still vague, the minimal requirement that should be met to know whether biological networks sharing topological properties also share biological functions is that the assessed global topological properties of these networks should be authentic. This condition should be respected, otherwise inaccurate models lead to inaccurate predictions.

The network degree distribution represents one of the most assessed network topological properties and was considered responsible of network behaviors. Many studies characterized most of the biological networks to be scale-free having a power-law degree distribution (Bollobàs and Riordan, 2003), a property that has been refuted for the PPI network and the metabolic network (Jeong et al., 2001; Przulj et al., 2004; Stumpf et al., 2005a). On the other hand, it has been shown that sub-networks generated from scale-free networks are not scale-free and that the observed scale-free characteristic of sub-networks cannot be extrapolated to the

complete networks with high confidence (Han et al., 2005; Stumpf et al., 2005b). Therefore, since the currently mapped networks are far from being complete and represent incomplete snapshots of the true underlying studied systems, the current state of data suggesting that the mapped networks are scale-free cannot guarantee that whole real-networks are scale-free. This means that some global topological properties of real biological networks are still hard to be well assessed. Hence, indeed topology could lead to certain important conclusions, but could not be used by itself to infer biological significance. However, integration of data with the network structure [e.g. the work performed in Chapter 2 in (Abd-Rabbo and Michnick, 2017), and in (Jothi et al., 2009)] could enhance our knowledge about the underlying biological mechanisms and strategies used by the biological networks to respond to different stimuli.

In summary, topological properties of network entities on all scales (nodes, logic motifs and networks themselves) partially determine the biological functions of these entities. However, integration of biological data with topological properties of these entities could exponentially enhance our understanding about the functional principles used by living cells to shape out their behavior.

1.6 Data integration

Despite the great contribution of the previously introduced network analyses in describing KP-Nets, most of the efforts studied KP-Nets as static maps that are disconnected from the dynamic nature of the cellular behavior. Recently, an increasingly used approach shows a great success in describing KP-Nets as dynamic maps and consists of integrating various experimental resources with KP-Nets. So far, data integration has been used to: (1), enhance the quality of the studied datasets as in (Breitkreutz et al., 2010); (2), identify hetero-regulatory motifs originating from different types of biological networks as in (Csikasz-Nagy et al., 2009; Ptacek et al., 2005; Wang et al., 2012a); (3), predict protein biological functions as in (Breitkreutz et al., 2010; Wang et al., 2012a); and (4), characterize the hierarchical structure of a KP-Net by the biological properties of KPs belonging to the different layers of the network hierarchy as in (Bhardwaj et al., 2010b; Cheng et al., 2015; Jothi et al., 2009). The important efforts that have been dedicated in this direction are introduced in the following sections.

1.6.1 Integrating other biological networks with phosphorylation networks

By integrating a PPI network and a TF-Net with a kinase phosphorylation network (K-Net) in the budding yeast, Ptacek et al. showed that transcription factors (TF) represent the largest class of kinase substrates. Moreover, the authors assessed the over-representation of hetero-regulatory motifs involving at least one phosphorylation interaction and/or a TF. This revealed six potential regulatory mechanisms governing kinases, kinase substrates, kinase regulatory subunits and TFs (Ptacek et al., 2005).

Another study integrated a TF-Net with a KP-Net in *S. cerevisiae*, and identified hetero-regulatory FFLs made of Cdk1, a TF and a protein (Csikasz-Nagy et al., 2009). In the identified hetero-regulatory FFLs, Cdk1 acts on a given protein directly by phosphorylating it and indirectly by phosphorylating TFs that regulate the gene expression of this same protein. Using mathematical modelling, the authors showed that these FFLs play a major role in controlling the temporal progression of cell cycle in the budding yeast (Csikasz-Nagy et al., 2009).

In another work, Wang et al. integrated a genetic network, a PPI network, a K-Net with a TF-Net in *S. cerevisiae* (Wang et al., 2012a). This work showed that the studied TF-Net is enriched for bi-fan motifs, whereas the studied K-Net is enriched for linear chains, suggesting that TF-Nets and K-Nets use different regulatory mechanisms to accomplish their biological processes. Wang et al. also predicted novel hetero-regulatory modules implicating a KP, a TF and a protein in a combined network formed from the genetic KP network and the TF-Net. These novel hetero-regulatory modules confirmed the structure and cross-talk of MAPK pathways, predicted a new function of the Sok2 TF in the HOG pathway and suggested an explanation for the reduced mating efficiency occurring on Fus3 deletion in MAPK pathways.

Finally, Breitkreutz et al. explored the interplay between the PPI network and the KP-Net by identifying the KP physical interaction network in the budding yeast using the MS technology (Breitkreutz et al., 2010). The authors built a KP global network, that represents an enhanced KP network, by combining the KP physical interaction network obtained by MS in this study with other KP physical interactions of high confidence obtained from LTP and HTP experiments. Interestingly, Breitkreutz et al. observed an over-representation of kinase-kinase interactions in the KP global network, suggesting that these interactions constitute an essential infrastructure for the coordination of cellular responses. Furthermore, the authors uncovered the roles of Cdc14 in

the MAPK signaling pathway, DNA damage response and metabolism; they also revealed the implication of new kinases in nitrogen and carbon metabolism.

1.6.2 Integrating biological data with KP-Nets

A common applied strategy of data integration within networks consists of overlaying static biological data, such as GO terms, gene essentiality, mRNA/protein abundance, mRNA/protein half-life and mRNA/protein noise, on top of the studied network. Yet, this approach has been rarely applied on K-Nets, but not on KP-Nets in the budding yeast. Moreover, to our knowledge, integration of dynamic biological data, such as temporal gene/protein expression and temporal localization, with both K-Nets and KP-Nets is still not performed in the budding yeast.

In 2010, Bhardwaj et al. elucidated the hierarchical structure of a K-Net in *S.cerevisiae* and assessed the enriched biological processes of each layer (top, core and bottom) of the K-Net hierarchy (Bhardwaj et al., 2010b). Bhardwaj et al found that top layer kinases were enriched for signaling response, the core layer for signal transduction and cell cycle regulation and the bottom layer for metabolic processes. Another effort integrating PPI network and multi-omics data (protein abundance, protein disorder, and literature-derived signaling reactomes) with phosphorylation data showed that phosphoproteins have much more interacting partners than non-phosphoproteins do and this independently of protein abundance and protein disorder level (Yachie et al., 2011). This finding highlights the large effect that the phosphorylation interactions have on both PPIs and on protein complexes formation. An equally important result from the latter study is the characterization of the co-phosphorylation of interacting proteins by the same kinase as a common regulatory mechanism in living cells. Finally, in a parallel effort to that in this thesis, Cheng el al. identified the hierarchical structure of a K-Net of the budding yeast assembled from *in vitro* phosphorylation interactions and integrated many biological properties of kinases in the K-Net (Cheng et al., 2015). Biological processes of kinases in the K-Net in top and core layers overlap those observed in (Bhardwaj et al., 2010b). The most interesting result found by Cheng et al. is that the top layer was found depleted for whereas core and bottom layers were enriched for kinases located in the bud and the bud neck. Cheng et al. suggested that top layer kinases might remain in the mother cells, while middle and bottom layer kinases might play a direct role in bud and bud neck subcellular compartments during yeast budding.

1.7 Thesis objectives and organization

1.7.1 Thesis aim and objectives

The aim of this thesis is to perform an integrative analysis of multi-omics data with the hierarchical structure of a bona fide KP-Net in the budding yeast, in order to generate hypotheses about the functional principles of each layer in the KP-Net hierarchy. This could ultimately permit the understanding of how the whole KP-Net functions and shapes cellular behavior.

I set out in this thesis to achieve four objectives: (1), to assemble a KP-Net from high confidence PDIs in the budding yeast; (2), to assess the hierarchical level of the assembled KP-Net and to elucidate its hierarchical structure (classify KPs among the top, core and bottom layers), (3), to overlap multi-omics data on top of the KP-Net hierarchical structure and suggest biological roles of each layer according to the topological and biological properties of KPs in each layer; and (4), to present an *in vivo* experimental method devised in the Michnick laboratory giving an example of enhanced techniques used to identify bona fide KP substrates and the regulatory subunits implicated in the modification of these substrates.

1.7.2 Thesis organization

Chapter 2, Delineating functional principles of the bow tie structure of a KP-Net in the budding yeast, presents a study of a KP-Net assembled from bona fide PDIs in the budding yeast *Saccharomyces cerevisiae*. The KP-Net was shown to have a hierarchical structure having the form of a bow tie (hour-glass) with a strongly connected core layer. Using the VS algorithm, this hierarchy was elucidated to have three disjoint layers: top, core and bottom. The overlay of a wide range of biological properties of KPs on top of the KP-Net hierarchical structure showed that phosphatases are less regulated by phosphorylation than kinases and that KPs that occupy a central position in the network, that receive a large number of inputs and that are implicated in more than one pathway, in cell-cycle processes and decision making, are among the most tightly regulated KPs in space and time. Moreover, the VS algorithm was shown to depend on both node in- and out-degrees to classify nodes in the different layers without biasing the biological results of the sorted network. Analysis of the robustness of our results showed that the three KP-Net layers were unstable on adding or removing edges, indicating that our findings reflect the current

state of knowledge in the literature. Nevertheless, the KP-Net layers significantly overlap with those of the resampled networks, showing a certain level of robustness of our results. Finally, an R package was developed to permit the application of the VS algorithm on other networks. This manuscript was submitted to the BMC Systems Biology journal

Chapter 3, Dissection of Cdk1-cyclin complexes *in vivo*, represents a collaboration work in the Michnick laboratory in which Ear et al. devised an innovative *in vivo* experimental method using the PCA based on the OyCD to identify substrates of the kinase Cdc28 and also cyclins that were involved in interactions between Cdc28 and their substrates in *Saccharomyces cerevisiae*. Many new candidate substrates and known substrates of Cdc28 were identified in this study. Also, the cyclin(s) implicated in the interactions between Cdc28 and its substrates were unprecedentedly identified *in vivo*, determining thus the phase of the cell cycle in which these interactions occurred. The mitotic spindle protein γ -tubulin (Tub4) was shown to be phosphorylated by the Clb3-Cdc28 complex, which suggests that this event occurs on early spindle assembly during the S-phase. Finally, most of the interactions between Cdc28 and its substrates were found to involve many cyclins, an observation that confirms complexity of KP-Nets in which many regulatory subunits could have redundant functions and consequently have compensatory roles. This manuscript was published in the PNAS journal.

Chapter 4, Conclusions, represents a discussion of the originality, the limitations, the impact of this research project on the current knowledge in the field of cellular regulation and signaling and finally, the expected future directions in this field.

Chapter 2 Delineating functional principles of the bow tie structure of a kinase-phosphatase network in the budding yeast

Diala Abd-Rabbo^{a,b} and Stephen Michnick^{a,b,1}

Article published in **BMC Systems Biology**, 11(1):38-52 (Abd-Rabbo and Michnick, 2017)

Included with the permission of BMC Systems Biology

^aDépartement de Biochimie et Médecine Moléculaire, Université de Montréal, C.P. 6128, Succ. Centre-ville, Montréal, Québec, H3C 3J7, Canada. ^bCentre Robert-Cedergren, Bio-Informatique et Génomique, Université de Montréal, C.P. 6128, Succ. Centre-ville, Montréal, Québec, H3C 3J7, Canada. ¹To whom correspondence should be addressed: E-mail: stephen.michnick@umontreal.ca and Tel : (514) 343-5849

2.1 Authors' contributions

Diala Abd-Rabbo and Stephen W. Michnick conceived the study. **Diala Abd-Rabbo** performed the research and the bioinformatics analyses, and wrote the manuscript. Stephen W. Michnick corrected the manuscript.

2.2 Abstract

Background: KPs form complex self-regulating networks essential for cellular signal processing. In spite of having a wealth of data about interactions among KPs and their substrates, we have very limited models of the structures of the directed networks they form and consequently our ability to formulate hypothesis about how their structure determines the flow of information in these networks is restricted.

Results: We assembled and studied the largest bona fide KP-Net known to date for the yeast *Saccharomyces cerevisiae*. Application of the VS algorithm on the KP-Net permitted to elucidate its hierarchical structure in which nodes are sorted into top, core and bottom layers, forming a bow tie structure with a strongly connected core layer. Surprisingly, phosphatases tend to sort into the top layer, implying they are less regulated by phosphorylation than kinases. Superposition of the widest range of KP biological properties over the KP-Net hierarchy shows that core layer KPs: (i), receive the largest number of inputs; (ii), form bottlenecks implicated in multiple pathways and in decision-making; (iii), and are among the most regulated KPs both temporally and spatially. Moreover, top layer KPs are more abundant and less noisy than those in the bottom layer. Finally, we showed that the VS algorithm depends on node degrees without biasing the biological results of the sorted network. Our findings provide a model and allow us to propose functional principles of a KP-Net based on knowledge current state. The VS algorithm is available as an R package (Supplementary methods file on CD-ROM).

Conclusions: the KP-Net model we propose possesses a bow tie hierarchical structure in which the top layer appears to ensure highest fidelity owing to high abundance of its members and limited variation in abundance across populations of individual cells. The core layer, by virtue of its enrichment for bottlenecks and KPs involved in decision-making appears to mediate signal integration and cell state-dependent signal interpretation. Our model of the yeast KP-Net provides both functional insight into its organization as we understand today and a framework for future investigation of information processing in yeast and eukaryotes in general.

Keywords: Kinase-phosphatase signaling network, Network hierarchical structure, Topological properties, Biological properties, Vertex-sort algorithm, Functional principles of cell behavior, *Saccharomyces cerevisiae*

2.3 Background

To maintain normal homeostasis, living cells continuously accommodate changes to their internal and external environment *via* signaling pathways. Protein KPs play an essential regulatory role in signaling pathways through PDIs that cause profound effects on substrates, affecting their turnover, localization and interactions with other proteins (Novak et al., 2010).

Numerous efforts have been made to reconstruct the budding yeast KP-Net from various types of interactions (Bhardwaj et al., 2010b; Bodenmiller et al., 2010; Breitkreutz et al., 2010; Cheng et al., 2015; Fiedler et al., 2009; Ptacek et al., 2005). Despite these efforts, KP-Nets assembled so far do not accurately represent genuine networks in which a KP acts directly on its substrate for the following reasons. First, dephosphorylation interactions are underrepresented in KP-Nets, because on one hand, dephosphorylation interactions are poorly annotated in public databases (Table A1.I) and on the other hand, phosphatases have been modestly studied in comparison to kinases. Second, K-Nets that were assembled from *in vitro* phosphorylation interactions do not include phosphatases and contain a considerable number of false positives due to the artificial environment in which phosphorylation interactions occur (Bhardwaj et al., 2010b; Cheng et al., 2015; Ptacek et al., 2005). Finally, KP-Nets that were assembled from PPIs and from genetic interactions and KP-Nets that were elucidated by knocking out a KP lack two crucial properties that usually characterize the command-execution aspect of regulatory networks: causality and directionality (Bodenmiller et al., 2010; Breitkreutz et al., 2010; Fiedler et al., 2009). Causality determines which KP directly acts on which substrate, whereas directionality indicates the direction of the interaction between the two interactors, which is especially required when substrates are themselves KPs. Interestingly, KP-Nets assembled from high quality PDIs are not characterized by the previously mentioned drawbacks and hence describe better genuine KP-Nets. Despite the large number of KP-Net studies, to our knowledge, no investigations were based on bona fide KP-Nets. More specifically, studies of KP-Nets in the budding yeast did not include interactions characterized by both causality and directionality (Bodenmiller et al., 2010; Breitkreutz et al., 2010; Fiedler et al., 2009). KP-Net studies that did include interactions characterized by both causality and directionality were not performed *in vivo* and did not include phosphatases (Bhardwaj et al., 2010b; Cheng et al., 2015; Ptacek et al., 2005)

(Table A1.II). Hence, constructing a bona fide KP-Net is still a missing fundamental constituent for analysis of signaling networks.

There have been a number of efforts to determine rules governing the organization and function of biological regulatory networks. For instance, a number of studies invoke command-execution organization characterizing directed networks to elucidate their hierarchical structure and this by applying network decomposition methods on various regulatory networks (Bhardwaj et al., 2010b; Cheng et al., 2015; Gerstein et al., 2012; Gulsoy et al., 2012; Jothi et al., 2009; Kim et al., 2012; Ma et al., 2004; Yu and Gerstein, 2006). Decomposition methods classify network nodes into different layers to elucidate information flow in network hierarchies. The majority of these efforts were aimed at TF-Nets, but rarely at other regulatory networks, including KP networks. In addition, network layers in these studies were characterized by topological and rarely by biological properties of their nodes; that is, KP-Nets are rarely characterized according to the features of the gene products that represent nodes such as stability, abundance and noise in mRNA and protein gene products (Table A1.II). However, biological properties are the ones that profoundly affect the regulatory state of any biological network.

Despite the wealth of available evidence, deciphering the complexity of KP-Nets to gain insights into their functional principles is still challenging. Here, we overcame two basic gaps in knowledge in previous studies: first, we constructed the largest bona fide KP-Net for the yeast *Saccharomyces cerevisiae*. Second, we elucidated the KP-Net hierarchical structure using the VS algorithm and unprecedentedly, we integrated the widest range of KP biological properties within this hierarchy. We found that the KP-Net has a bow tie hierarchy formed of three layers (top, core and bottom) and that the different biological properties of KPs are unevenly distributed among KP-Net layers. This uneven distribution reveals general biological properties of KPs in each layer from which we could postulate the behaviours and information processing functions of each layer in the KP-Net hierarchy. We suggest that high protein abundances and low protein noise in KP-Net top layer could result in signal fidelity, whereas enrichment for decision-making and bottleneck proteins in the core layer may underlie signal integration. Finally, we showed that node degrees affect the way the VS algorithm sorts nodes within a network but we also showed that our results and conclusions are not biased by node degrees. We developed an R package called the VertexSort to facilitate VS algorithm application to other networks (Table A1.III).

2.4 Results

2.4.1 The kinase-phosphatase network (KP-Net)

The KID database provides the most detailed and specialized annotation of kinase-protein interactions; however, phosphatase-protein interactions are not included in and many kinase-protein interactions are missing or partially annotated in this database. Hence, we collected these interactions from different sources, then, curated, annotated and scored the collected interactions according to the KID database pipeline with minor adjustments to annotate phosphatase-protein interactions (Figure 2.1a and Appendix 1: Supplementary methods) (Ba and Moses, 2010; Fiedler et al., 2009; Magrane and Consortium, 2011; Muller et al., 2004; Stark et al., 2006; Stark et al., 2010) (<http://www.ebi.ac.uk/citexplore>). The KID pipeline associates a confidence score to each interaction based on the extent to which the different experimental methods that validate an interaction contribute to identifying a true positive Kinase-protein interaction. To ensure that the interactions assembled in the KP-Net represent PDIs rather than simply Kinase-protein or phosphatase-protein interactions, we selected interactions having a confidence score ≥ 4.52 (corresponding to a $P \leq 5 \times 10^{-2}$) and those validated by at least a biochemical experiment showing the occurrence of a PDI (*in vitro* kinase assay, *in vivo* or *in vitro* phosphosite mapping, mobility shift of phosphoproteins on gel or substrate trapping by a dead phosphatase catalytic domain). The assembled KP-Net contains 1087 directed interactions (918 and 169 PDIs, respectively) implicating 616 proteins [101 kinases and 31 phosphatases, covering ~77 % of these enzymes and 484 proteins, most of which are KP substrates that are not KPs, (Figure 2.1a and Table A1.III)]. Similar to other biological networks, the KP-Net possesses a scale-free structure ($P(K) \sim K^{-2.58}$ with a goodness-of-fit test $P = 1.3 \times 10^{-2}$) in which most KPs regulate few proteins and few KP hubs regulate a large number of proteins (Appendix 1: Supplementary methods and Figure A1.1).

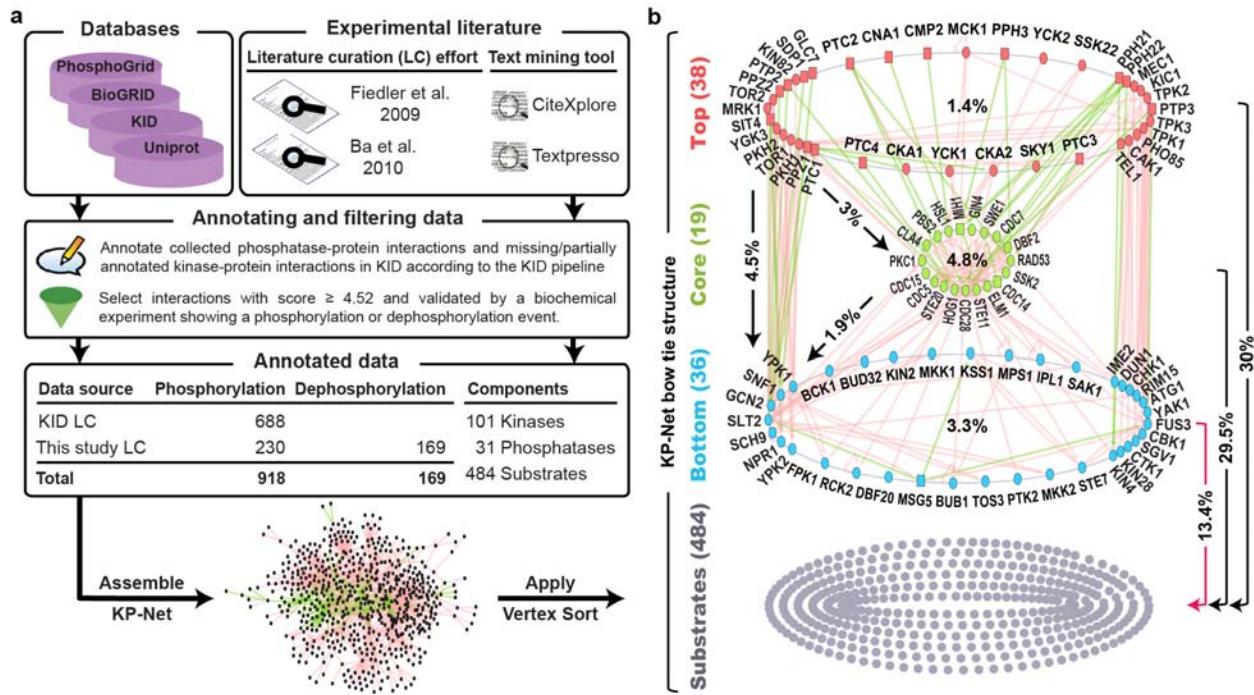


Figure 2.1. The pipeline used to assemble and to sort the KP-Net, and the KP-Net bow tie structure.

(a) The steps followed to elucidate the KP-Net hierarchical structure going from the different sources used to collect kinase-protein and phosphatase-protein interactions, passing through the data annotation procedure and filtering criteria applied to select high quality PDIs, to the assembly and sorting of the KP-Net by the VS algorithm. **(b)** The bow tie structure of the KP-Net showing how KPs are classified in top, core and bottom layers. Top layer KPs control core layer KPs; top and core layer KPs control bottom layer KPs and KPs in the three layers control proteins in the substrates layer formed of proteins that are not KPs and of KPs having no substrates. Numbers between parentheses represent number of nodes in each layer. Arrows represent directed interactions (red: phosphorylation, green: dephosphorylation and black: both). Percentages designate percentage of interactions within and between layers.

2.4.2 The KP-Net possesses a “corporate” hierarchical structure in the form of a bow tie with a strongly connected core layer

We assessed the amount of the hierarchical structure of the KP-Net by calculating its GRC, which represents a normalized average of the proportions of nodes accessible from each node in the network (Mones et al., 2012). The closer the GRC is to 1, the more hierarchical the network is. The KP-Net has a moderate GRC of 0.61, suggesting that the KP-Net represents a hierarchical structure that could be placed between two extremes: (i), an autocratic structure comparable to a complete tree; and (ii), a democratic structure in which collaborative regulation dominates and no hierarchy exists (Bhardwaj et al., 2010b). Bhardwaj et al. observed a similar moderate

hierarchy in a co-phosphorylation network and described it as a corporate hierarchy (Bhardwaj et al., 2010b). Obviously, the KP-Net does not represent a complete tree, as it is enriched for many logic motifs that do not occur in trees: FFLs (a structure in which a node regulates another node and together they regulate a third one), two node FBLs (two nodes that regulate each other), and bi-fans (a structure in which two nodes regulate two other nodes) ($P < 10^{-3}$, Methods). Moreover, the KP-Net does not represent democracies and encapsulates a hierarchical structure, as its GRC is significantly higher than that of Erdős–Rényi random networks (non-hierarchical networks) having the same number of nodes and edges as the KP-Net ($P < 10^{-4}$, Methods). Interestingly, the GRC of the KP-Net is significantly smaller than that of random networks generated by degree preserving randomization (DPR, Methods). This result is not surprising, as the degree distribution of a network is essential to determine its organizational structure, meaning networks having the same degree distributions will have similar organizational structures. Thus the GRC of the KP-Net was expected to be comparable to that of DPR networks, but was found significantly smaller than the GRC of DPR networks; probably an indication of the heavy loop system (enrichment for FBLs) that exists in KP-Nets.

Subsequently, we applied the VS algorithm on the KP-Net to elucidate the network hierarchical structure and the signal flow within the elucidated hierarchy. The VS algorithm is one of the best network decomposition algorithms applied by Jothi et al. to a transcription regulatory network of the budding yeast to elucidate its network hierarchical structure (Bhardwaj et al., 2010b; Cheng et al., 2015; Gulsoy et al., 2012; Hartsperger et al., 2010; Jothi et al., 2009; Ma et al., 2004; Yu and Gerstein, 2006). The VS algorithm sorts nodes into different levels so that nodes in upper levels control those in lower levels (Jothi et al., 2009). It first transforms a cyclic graph to an acyclic one by collapsing each SCC (a sub-graph where each node pair is related by two paths of opposite directions) into a super node and then it applies the leaf removal algorithm on the resulting graph and on its transpose. This generates global solutions in which a node could span a range of levels, taking into account the huge amount of missing data in and the dynamic nature of biological networks.

Application of the VS algorithm on the KP-Net elucidated a hierarchical structure in which KPs are sorted into 9 levels that we subsequently grouped into three non-overlapping layers: top, core and bottom (Figure A1.2a). As in Jothi et al., we first identified KPs of the largest SCC and classified them in the core layer (19 KPs). We then classified KPs that regulate

core layer KPs into the top layer (38 KPs) and those that are regulated by core layer KPs into the bottom layer (36 KPs) (Figure 2.1b) (Jothi et al., 2009). Thirty-eight nodes (33 KPs and five proteins that are not KPs) were excluded from further analysis (Figure A1.2b). The thirty-three KPs cannot be classified in the KP-Net hierarchy because they are not connected to any KP in the KP-Net (Figure A1.2b). In addition, the five proteins are substrates of the excluded KPs, therefore they also cannot be included in the KP-Net hierarchy (Figure A1.2b). The three layers of the KP-Net generated a bow tie structure in which the core layer has relatively fewer nodes than top and bottom layers (Figure 2.1b). It is important to note that the bow tie shape of the KP-Net is not the result of the application of the VS algorithm on the KP-Net. More specifically, the bow tie structure is not the result of choosing the core layer as the SCC of the KP-Net for two reasons. First, by applying the VS algorithm in the same way, the hierarchical structure of the regulatory network elucidated by Jothi et al. do not have a bow tie shape (top, core and bottom layers contain 25, 64 and 59 nodes, respectively) (Jothi et al., 2009). Second, the probability of generating a bow tie shape from random networks generated by DPR is 0.5 (Appendix 1: Supplementary methods).

Interestingly, KP-Net top, core and bottom layers regulate 235, 276 and 148 proteins, respectively, corresponding to 38 %, 45 % and 24 % of the KP-Net nodes, respectively. Although the core layer is ~2 times smaller in size than top and bottom layers, it regulates a number of substrates that is 1.2 and 1.9 times larger than that regulated by top and bottom layers, respectively, implying an essential role of the core layer in the KP-Net.

2.4.3 The three layers of the KP-Net have dissimilar biological roles and subcellular localization

To unravel biological roles of the KP-Net layers, we performed a GO enrichment/depletion analysis for KPs in each of these layers (Appendix 1: Supplementary methods). We found that the KP-Net top layer is enriched mostly for signal regulation and transduction; interestingly, the core layer is enriched for signaling also, for metabolic processes, but mostly for cell cycle, organization processes related to cell cycle and decision-making (Table A1.IV), confirming the essential role of the core layer in the KP-Net; and the bottom layer is enriched for few GO terms, suggesting that it has a less specialized and more diverse biological role (Figure 2.2a). These results are in line with the findings of Bhardwaj et al. (Bhardwaj et al., 2010b).

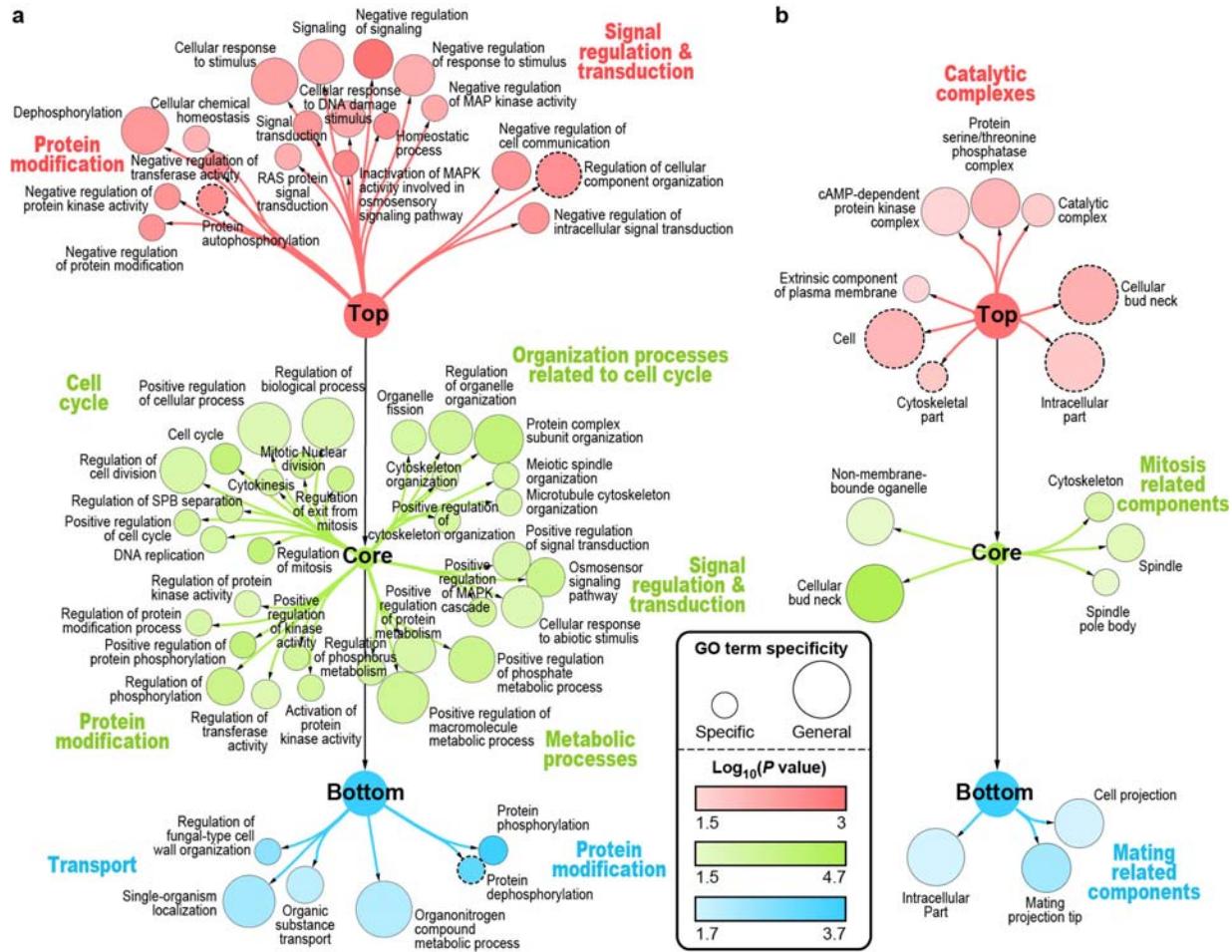


Figure 2.2. Depleted and enriched biological processes and cellular components in each of the KP-Net layers.

Depleted and enriched **(a)** biological processes and **(b)** cellular compartments associated to KPs in each of the KP-Net layers (top: red; core: green; and bottom: blue). Nodes represent the different enriched and depleted GO terms. Color gradients represent \log_{10} of P -values ($\log_{10}(P)$) of enriched and depleted GO terms. Size of nodes indicates the specificity of each GO term (small: specific and large: general). Enriched GO terms are encircled with solid border while depleted ones are encircled with a dashed border.

On another level, the top layer is depleted for, whereas the core layer is enriched for KPs located in the bud neck (Figure 2.2b), a result that has been already observed by Cheng et al. (Cheng et al., 2015). We further found that the bottom layer is enriched for KPs located in the mating projection tip (Figure 2.2b). The latter observations suggest that top layer KPs might remain in the mother cell to regulate signaling, while core layer KPs might enrich in the daughter cell as a reflection of their roles in mitosis, while bottom layer KPs might reside in the cell projection to contribute in mating. These findings are interesting and merit further investigation.

Strikingly, dephosphorylation is enriched in the top layer and depleted in the bottom layer of the KP-Net (Figure 2.2a), proposing that phosphatases are over-represented in signaling pathway upstream parts and depleted in signaling pathway downstream parts. The latter results are consistent with dynamic phosphoproteomic studies showing that at least 50 % of early responses to cell perturbations are depletion of phosphosites (Kanshin et al., 2015).

2.4.4 Phosphatases are less regulated by phosphorylation than kinases

Our findings confirmed our propositions; indeed, the top layer is enriched for phosphatases whereas the bottom layer is depleted for phosphatases (Figure A1.3a, $P = 2.2 \times 10^{-5}$ and $P = 4.1 \times 10^{-4}$ respectively; hypergeometric test (HT)). Besides, we observed that 81 % of the top layer phosphatases have a zero in-degree. Using high quality phosphoproteomic data annotated in the PhosphoGRID database, we also found that the number of phosphosites identified in phosphatase protein sequences is smaller than that identified in kinases (Figure A1.3b, $P = 2.3 \times 10^{-3}$; randomization test (RT), Methods). These results strongly suggest that phosphatases are less regulated by phosphorylation than kinases are. Our suggestion is also supported by the great variety of regulatory subunits controlling phosphatases (Shi, 2009) and by the large number of cellular mechanisms, other than phosphorylation, reported to regulate phosphatases, including phosphorylation of the regulatory subunits of phosphatases (Ahn et al., 2007; Janssens et al., 2008; Maeda et al., 1993; Mitchell and Sprague, 2001; Trinkle-Mulcahy et al., 2006; Trockenbacher et al., 2001).

2.4.5 KP-Net upper levels are the least regulated and KP-Net lower levels are the least to regulate other KPs

Top layer KPs in-degrees are on average smaller than KP in-degrees in core and bottom layers (Figure 2.3a, $P < 10^{-4}$; RT, Methods). This observation is a direct result of the VS algorithm application ($P = 10^{-3}$; degree non-preserving randomization (DNPR), Methods) on a network, but it agrees with organizational principles found in hierarchical systems in which members of upper levels are the least regulated (e.g. pyramid networks). In contrast, the out-degree of the bottom layer is significantly smaller than that of top and core layers (Figure 2.3b, $P = 3 \times 10^{-3}$; RT, Methods). This finding is independent of the VS algorithm application ($P = 0.7$; DNPR, Methods) on a network and has been previously observed in the hierarchical structure of a yeast

transcriptional regulatory network elucidated by a decomposition algorithm (BFS) different than the VS algorithm (Yu and Gerstein, 2006). Finally, the observed features related to node in- and out-degrees were implemented in two network decomposition algorithms, other than the VS algorithm, to classify nodes in top and bottom layers, respectively (Bhardwaj et al., 2010b; Gerstein et al., 2012).

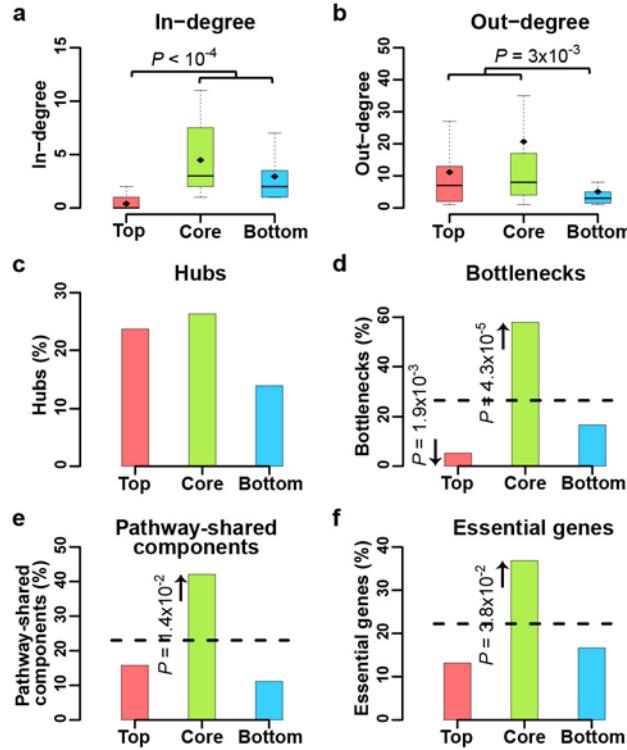


Figure 2.3. Topological and biological properties of KPs in the different layers of the KP-Net.

Distribution of (a) in-degree and (b) out-degree of KPs in each layer of the KP-Net and percentage of KPs representing (c) hubs, (d) bottlenecks, (e) shared components between pathways (KPs involved in at least two pathways), and (f) essential genes in each layer of the KP-Net. The broken line in bar plots represents the expected mean of the corresponding percentage in each layer. Black diamonds in box plots designate the average of the corresponding property of KPs. Outliers were omitted from box plots to simplify data representation. P -values were calculated by comparing property means of two layers and the enrichment/depletion of a property within a layer using the RT (Methods) and HT, respectively. For description of the used datasets see Supplementary materials in Appendix 1.

In the following sections, we examined the distribution of an extensive range of KP biological properties among the three layers of the KP-Net hierarchy in an attempt to delineate the functional role of each of these layers within the hierarchy.

2.4.6 The KP-Net core layer is enriched for essential genes, bottlenecks, and pathway-shared components

To better grasp our knowledge of signal flow in the KP-Net, we analysed the distribution of hubs, bottlenecks, pathway-shared components (KPs involved in at least two pathways) and essential genes in the three layers of the KP-Net. Hubs and bottlenecks are defined as the 20 % of KPs in the KP-Net that have, respectively, the highest degree and the highest betweenness (fraction of shortest paths between all pairs of nodes that pass through a single node; this measure captures how much signaling passes through a node). The hubs are equally distributed among the three layers, reflecting the prevalence of parallel regulation as a principle emerging from the three layers of the KP-Net (Figure 2.3c). Interestingly, the core layer is enriched for bottlenecks, pathway-shared components and essential genes (Figure 2.3d-f, $P = 4.3 \times 10^{-5}$, $P = 1.4 \times 10^{-2}$ and $P = 3.8 \times 10^{-2}$, respectively; HT), suggesting that most of the signal integration and crosstalk between pathways occur in the core layer.

2.4.7 Molecular switches are enriched in KPs in core and bottom layers

Molecular switches represent phosphosites within or adjacent to linear binding motifs (LBM) which mediate “on demand” controls switching proteins between different functional states (on-off, specificity, cumulative and sequential switches) (Van Roey et al., 2012). Given their fundamental role in controlling signaling networks, we investigated the distribution of KP molecular switches in the KP-Net hierarchy. We predicted protein disordered regions in KP protein sequences and LBMs within these predicted disordered regions using the IUPred and ANCHOR algorithms, respectively (Appendix 1: Supplementary methods) (Dosztanyi et al., 2005; Meszaros et al., 2009). We then overlaid bona fide *in vivo* phosphosites from the PhosphoGRID database on top of KP protein sequences (Appendix 1: Supplementary materials). We found that percentage of predicted disordered regions in KP proteins in core and bottom layers are on average higher compared to the top layer (Figure 2.4a, $P < 2.3 \times 10^{-2}$; RT, Methods). The same trend is observed for: (i), the percentage of sequences predicted to contain LBMs (Figure 2.4b, $P < 2.1 \times 10^{-2}$; RT, Methods); (ii), the number of phosphosites in KP sequences generally (Figure A1.3c, $P < 6.1 \times 10^{-4}$; RT, Methods) and (iii), in the predicted LBMs particularly (Figure 2.4c, $P < 2.2 \times 10^{-2}$; RT, Methods); and (iv), the number of potential molecular switches in each KP (Figure 2.4d, $P < 3.1 \times 10^{-3}$; RT, Methods). Interestingly, our

findings suggest that phosphorylation of KPs in lower layers could form molecular switches important for KP temporal regulation. Two out of many examples confirming our suggestions are: (1), the specificity switch in Hsl1 (core layer kinase and morphogenesis checkpoint regulator) leading to a G2 arrest essential for cell survival upon osmotic shock; and (2), the on-switch in Swe1 (core layer kinase) maintaining Cdc28 in an inhibited form essential for entry of cells into mitosis (Clotet et al., 2006; Harvey et al., 2005).

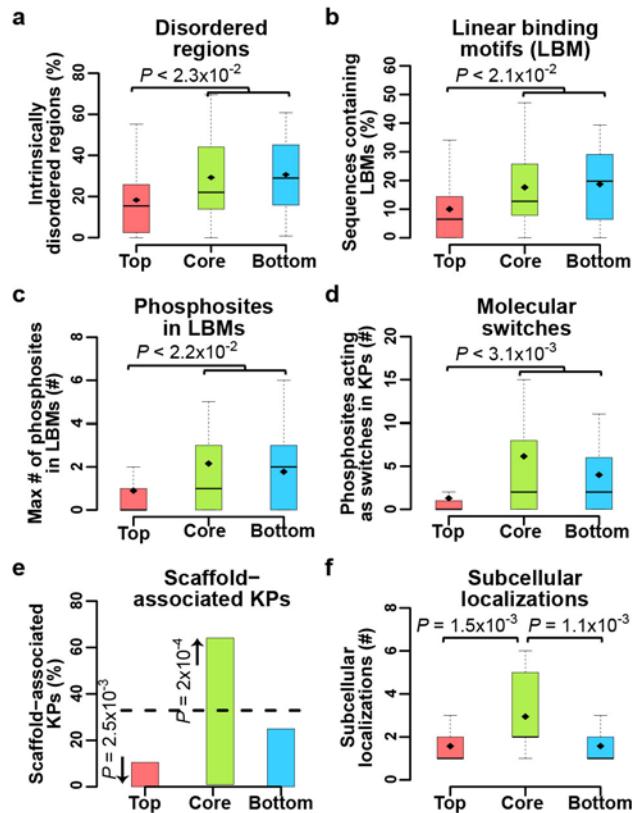


Figure 2.4. Biochemical and spatiotemporal modulators of KPs in the different layers of the KP-Net.

Distribution of (a) the percentage of disordered regions, (b) the percentage of predicted linear binding motifs (LBM), (c) the maximum number of phosphosites within a predicted LBM, (d) the number of phosphosites acting as molecular switches in each KP in KP-Net layers and (e) the percentage of scaffold-associated KPs and (f) the distribution of the number of subcellular localizations in which a KP was detected for KPs in each of the three KP-Net layers. For description of box plots and bar plots, see Figure 2.3 and for description of used datasets see Supplementary materials in Appendix 1.

2.4.8 Core layer KPs employ scaffolding to prevent unwanted pathway crosstalk

It is now well established that redirecting information flow within signaling networks is accomplished through scaffolding and is required for the insulation of interconnected pathways (Pawson and Scott, 1997). Interestingly, the KP-Net core layer is enriched for pathway-shared components (Figure 2.4e) and for LBMs (Figure 2.4b), suggesting that core layer KPs that are shared between pathways associate with scaffolds through LBMs. Indeed, although core and bottom layers are enriched for potential LBMs, only the core layer is enriched for scaffold-associated KPs (Figure 2.4e, $P = 2 \times 10^{-4}$; HT). This indicates that scaffolding is extensively employed at the core layer where most pathway crosstalk occurs (Figure 2.4d-e), in order to prevent inappropriate cellular responses resulting from the activation of undesired pathways. For instance, the mitogen extracellular signal-regulated kinase kinase Ste11, a core layer kinase, is involved in three pathways: high osmolarity, filamentous growth and pheromone pathway. Association of Pbs2 (a MAPK kinase and a scaffold protein implicated in the HOG signaling pathway) and Ste5 (a pheromone-responsive MAPK scaffold protein) with Ste11 reorients signal flow by activating the HOG signaling pathway and the mating pathway, respectively; whereas, unavailability of both Pbs2 and Ste5 favour filamentous growth (Schwartz and Madhani, 2004).

2.4.9 Core layer KPs undergo more spatial organization changes than top and bottom layer KPs

Controlling spatial distribution of KPs plays an essential role in tuning KP activity and specificity towards their substrates (Mattison and Ota, 2000; Ubersax and Ferrell, 2007). By superposing microscopy subcellular localization data of proteins in single cells under different stress conditions (Chong et al., 2015) on top of the KP-Net hierarchy, we observed that KPs in the core layer dynamically redistribute among more subcellular compartments than KPs in top and bottom layers (Figure 2.4f, $P < 1.6 \times 10^{-3}$; RT, Methods). This indicates that core layer KPs might be subject to a more stringent control than top and bottom layer KPs to tightly restrict their localization. Hog1 is a relevant example of a core layer kinase that is translocated from the cytoplasm to the nucleus to trigger a wide transcriptional response on exposure to a high osmolarity stimulus (Muzzey et al., 2009). Another typical example of tight localization control

is Cdc14, a core layer phosphatase essential for mitotic exit, which after its sequestration in the nucleolus, is released to the nucleus and the cytoplasm where it associates with the spindle pole body (SPB) during early anaphase (Bloom et al., 2011).

2.4.10 Top layer KP proteins are more abundant and less noisy than bottom layer KPs of the KP-Net

Since KPs turnover determines their availability and thus their activity, we overlaid various information of KP turnover (Appendix 1: Supplementary materials) on top of the KP-Net hierarchy (Arava et al., 2003; Basehoar et al., 2004; Belle et al., 2006; Eser et al., 2014; Miura et al., 2008; Newman et al., 2006; Wang et al., 2012b). While transcripts coding for core layer KPs are synthesized at a higher rate than top and bottom layers (Figure 2.5a, $P < 3.9 \times 10^{-3}$; RT, Methods), mRNA of top layer KPs have longer half-lives than core and bottom layers (Figure 2.5b, $P < 4.6 \times 10^{-3}$; RT, Methods). However, mRNA abundance has a similar trend to mRNA half-life, implying that mRNA degradation (the process that determines half-lives) is more important than synthesis rate in determining mRNA abundance (Figure 2.5c, $P < 1.8 \times 10^{-2}$; RT, Methods). Similarly, mRNA of top layer KPs are translated at higher rates than core and bottom layers (Figure 2.5d, $P < 4.8 \times 10^{-2}$; RT, Methods). However, half-lives of KP proteins are statistically comparable among the three layers of the KP-Net (Figure 2.5e; RT, Methods), suggesting that proteins abundance should have the same trend as the translation rate of mRNA molecules. This is partially true, since top layer KP proteins are more abundant than the bottom layer (Figure 2.5f, $P = 3.3 \times 10^{-2}$; RT, Methods), but not more abundant than the core layer. This discrepancy might be due to the fact that KP proteins in the core layer tend to have longer half-lives (mean values are reported; 95 minutes, Fig. 5e) than the top layer (69 minutes, Figure 2.5e). On another level, percentages of noisy KP genes at the mRNA level are comparable among the three KP-Net layers (Figure 2.5g; HT). Moreover, top layer KP proteins are less noisy than core and bottom layers in starving *S. cerevisiae* cells (Figure 2.5h, $P < 2.2 \times 10^{-2}$; RT, Methods). Interestingly although, we observed significant relative differences in each of protein abundance and noise between KP-Net layers, notably proteins were abundant (Figure 2.5f, top 5336 molecules/cell, core 3041 molecules/cell and bottom 2436 molecules/cell) and not noisy (Figure 2.5h, top -0.94 a.u., core -0.05 a.u. and bottom 0.12 a.u.) in the three layers of the KP-Net. Taken together, these results suggest that higher protein abundance coupled with lower

protein noise in the three layers and in particular in the top layer, might confer high signaling fidelity to the KP-Net.

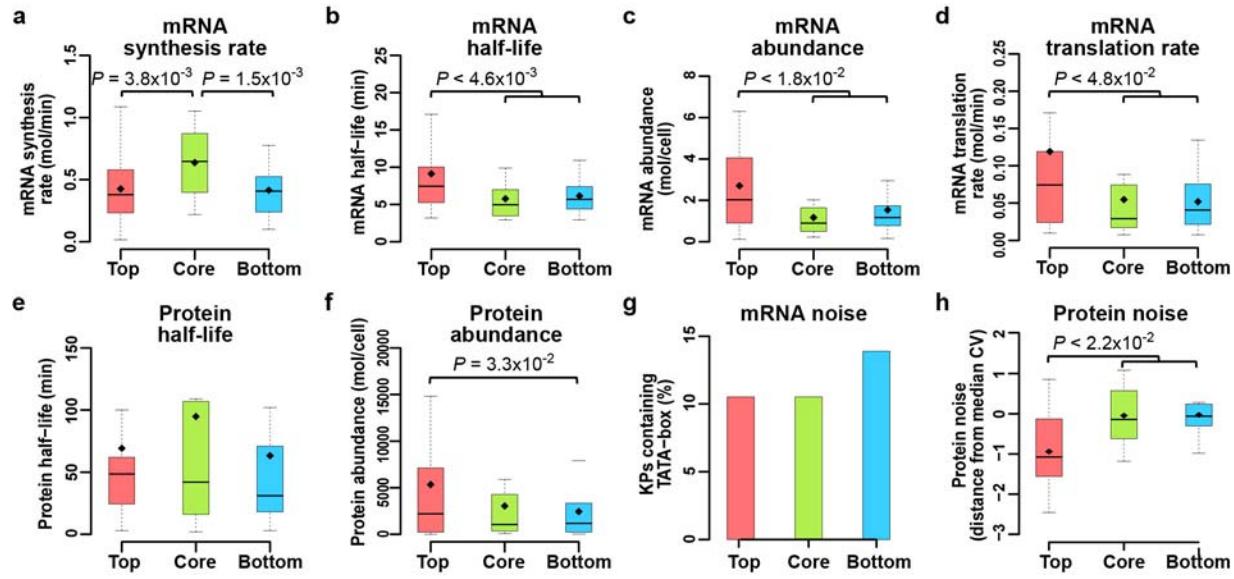


Figure 2.5. mRNA and protein turnover related properties of KPs in the different layers of the KP-Net.

Distribution of (a) mRNA synthesis rate, (b) mRNA half-life, (c) mRNA abundance, (d) mRNA translation rate, (e) protein half-life, (f) protein abundance, (g) percentage of noisy mRNA KPs and (h) distribution of noise in KP protein abundance of KPs in the different layers of the KP-Net. A KP is considered to be noisy at the transcriptomic level, if the promoter region of its gene was predicted to contain a TATA-box consensus sequence. Protein noise was defined as the distance of coefficient of variation (CV) of protein abundance from a running median of protein abundance CV. For description of box plots and bar plots, see Figure 2.3 and for description of used datasets see Supplementary materials in Appendix 1.

2.4.11 The VS algorithm depends on node degree to classify network nodes in three layers

As the findings of this study mainly result from the application of the VS algorithm, we asked whether the VS algorithm depends on a specific node property to sort nodes into three layers and whether these findings reflect the biology underlying the KP-Net. To address these questions, we generated five sets of 1000 random networks produced using five randomization methods: degree preserving randomization (DPR), similar degree preserving randomization (SDPR), in-degree preserving randomization (IDPR), out-degree preserving randomization (ODPR), and degree non-preserving randomization (DNPR) (Methods). We then applied the VS algorithm on these random networks and plotted means of KP properties in each layer of the KP-Net (black

diamonds, Figure 2.6), means of KP properties in each layer of random networks (points joined by colored lines, Figure 2.6) and the 95 % confidence interval of random network means (colored vertical segments, Figure 2.6).

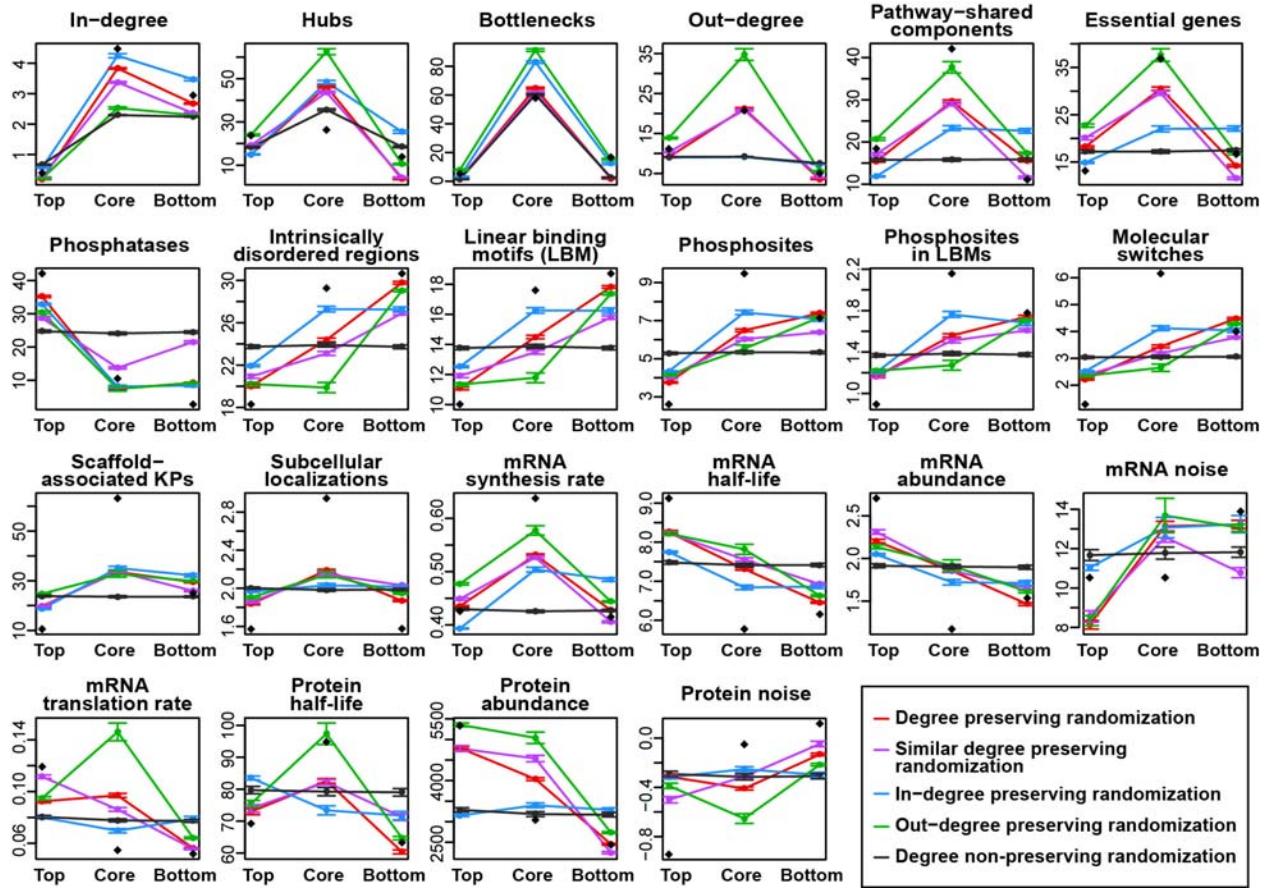


Figure 2.6. The VS algorithm depends on node degrees to sort network nodes in three layers.

The mean and its 95 % confidence interval of the studied properties of KPs in the three layers of each of the five sets of the 1000 random networks generated by: degree preserving randomization (DPR, red line), similar degree preserving randomization (SDPR, pink line), in-degrees preserving randomization (IDRP, blue line), out-degrees preserving randomization (ODRP, green line) and degree non-preserving randomization (DNPR, black line). The black diamonds represent the mean of studied properties of KPs in the three layers of the KP-Net.

Strikingly, we observed that the distribution of all properties, except in-degrees, hubs and bottlenecks, of the three layers form a straight horizontal line for DNPR networks (Figure 2.6, black line), showing that the VS algorithm produces a particular global signature (they peak at the core layer) in completely random networks for only these three properties that are all related to node degrees. Interestingly, the distribution of all properties in the DPR and SDPR networks

(red and pink lines, Figure 2.6) are the closest to each other when node degrees are similar to each other (DPR and SDPR cluster together in Figure A1.4). Taken together, our observations suggest that the VS algorithm depends on node degree to sort network nodes in the different layers. Moreover, on clustering the five sets of randomized networks using the Euclidean distance between the different properties of their KPs, we found that ODPR networks are closer to DPR networks than IDPR networks (Figure A1.4), suggesting that the VS algorithm depends on node out-degrees more than node in-degrees. However, the VS algorithm obviously depends also on node in-degrees, as any node with a zero in-degree will be automatically placed in the top layer. Therefore, the VS algorithm depends on both nodes in- and out-degrees. Nevertheless, although the VS algorithm depends on node degrees to classify network nodes into different layers, two observations suggest that KP biological properties are not associated with KP degrees and that they are not the result of a bias in the VS algorithm: (i), all biological properties showed a straight line distribution in completely random networks (Figure 2.6, black line); and (ii), most of the means of KP biological properties in KP-Net layers (black diamonds, Figure 2.6) are outside of the 95 % confidence interval of the means of the corresponding properties in random network layers.

2.4.12 Biological properties of KPs are independent of their in- and out-degrees

To confirm that the observed biological properties of the KP-Net are not the result of a bias in the VS algorithm, we investigated the association between the different studied properties of KPs and their in- and out-degrees, respectively. Unsurprisingly, KP topological properties (hubs and bottlenecks) are associated with their in- and out-degrees, but most of KP biological properties are not or are weakly associated with their in- and out-degrees (Figures I1.5-8). Only, the number of KP phosphosites is moderately correlated with KP in-degrees (Figure A1.5, $R = 0.45$; Spearman correlation), and the association of KP with scaffold and the type of enzyme (KP) are associated with KP in-degrees (Figure A1.6, $P = 2.6 \times 10^{-3}$ and $P = 3.4 \times 10^{-2}$, respectively; one-way ANOVA test). To note that, the first result is a consequence of the KP-Net definition: the more KP sequences contain phosphosites, the more KPs are phosphorylated or dephosphorylated, thus the higher their in-degree is and vice-versa. The third result was expected, as phosphatases are over-represented in the top layer and they contain less

phosphosites than kinases (Figure A1.3a-b), suggesting that phosphatase in-degrees are smaller than kinase in-degrees. Moreover, KP protein abundance and noise are also mildly correlated with their out-degrees (Figure A1.7, $R = 0.35$; Spearman correlation) and KP gene essentiality is associated to KP out-degrees (Figure A1.8, $P = 1.4 \times 10^{-3}$; one-way ANOVA test). In conclusion, 6 out of 18 KP biological properties are associated with either their in- or out-degrees and most of the KP biological properties (12 out of 18) are neither associated with their in- nor with their out-degrees. Thus, most KP biological properties describing each layer of the KP-Net are intrinsic to the network and are classified independently of KP degrees.

2.4.13 Robustness of results and incompleteness of data

It did not escape our attention that the KP-Net that was assembled in this study represents a small snapshot of the whole phosphorylation network of the budding yeast. Therefore, we assessed the robustness of our results to missing interactions by generating noisy networks (adding edges to the KP-Net) and the robustness of our results to false positives by generating subsampled networks (deleting edges from the KP-Net) (Methods). We then assessed the stability of KP-Net layers using the Jaccard coefficient as a measure of similarity between KP-Net layers and noisy/subsampled network layers (Methods) (Henning, 2007). Also, we assessed the significance of the overlap between KP-Net layers and noisy/subsampled network layers using the HT (Methods) (Henning, 2007). We observed that the KP-Net is more robust to removing than to adding edges (Figure 2.7a and 7c). Moreover, the more edges are added to and removed from the KP-Net, the more the three layers become unstable (Figure 2.7a and 7c). However, in spite of this instability, all layers in noisy/subsampled networks significantly overlap with the KP-Net layers (Figure 2.7b and 7d), showing that our findings are sufficiently robust to describe the KP-Net with our current knowledge. Finally, properties characterizing the KP-Net were retained to different levels in the noisy networks (Appendix 1: Supplementary methods and Figure A1.9), confirming that the characteristics of the KP-Net elucidated in this study represent the best of our knowledge to date about KP-Nets.

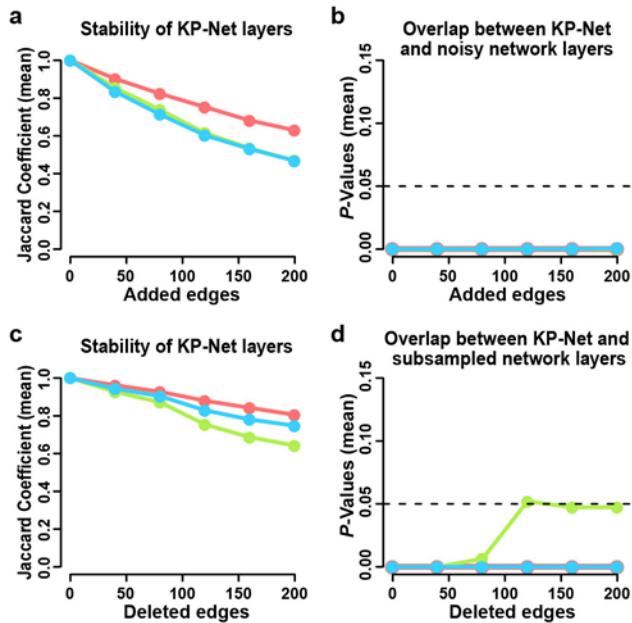


Figure 2.7. Stability of KP-Net layers and their overlap with subsampled/noisy network layers.

(a) Stability of KP-Net layers on adding edges to the KP-Net. **(b)** Significance of the overlap between KPs in each layer of the KP-Net and noisy network layers on adding edges. **(c)** Stability of KP-Net layers on deleting edges from the KP-Net. **(d)** Significance of the overlap between KPs in each layer of the KP-Net and subsampled network layers on deleting edges. Stability was quantified using the Jaccard coefficient as a similarity measure between KPs belonging to KP-Net layers and those belonging to noisy/subsampled network layers. P -values in (b) and (d) were calculated using the hypergeometric test. Colors designate the different layers of the noisy/subsampled networks (top: red, core: green and bottom: blue).

2.4.14 Predicting Kinases acting on substrates on high osmolarity stress

Presently, one of the most active areas of research consists of linking each KP to its substrates. As an example, we attempted to predict the kinases that could possibly phosphorylate substrates characterized by a change in their level of phosphorylation in cells exposed to osmotic shock. We used the KP-Net as a gold standard; we overlaid on top of it phosphorylation consensus motifs curated from the literature and proteins that undergo time-dependent phosphorylation or dephosphorylation following osmotic shock from Kanshin & Bergeron-Sandoval et al. (Kanshin et al., 2015). We identified 57 interactions linking 19 kinases to 25 potential substrates (Methods and Table A1.V). The overlap between the predicted kinases in our study and the kinases that underwent changes in phosphorylation in Kanshin & Bergeron-Sandoval et al. was significant ($P = 3.8 \times 10^{-2}$; HT). This result suggests, first, that a significant number of the 19 kinases that

we predicted to act on 25 potential substrates do undergo time-dependent changes in phosphorylation that may reflect their activation or deactivation in response to osmotic shock; second, that the interactions forming the KP-Net that was assembled in this study are of high confidence; and finally, that this same KP-Net could be used as a benchmark with other phosphoproteomic data to identify kinases and may be phosphatases that might act on a set of substrates.

2.4.15 Discussions

In this study, we assembled the largest bona fide KP-Net known to date for the yeast *Saccharomyces cerevisiae*. We found, first, that the KP-Net has a moderate hierarchical structure made of three layers (top, core and bottom) in the form of a bow tie structure having a strongly connected core layer. Second, phosphatases are for the first time shown to be less regulated by kinases than are kinases by each other. Third, the observed high abundance and low noise of KP proteins in the three layers of the KP-Net, but notably in the top layer, may reflect an adaptation by which maximal sensitivity to signals at the earliest steps of signaling is assured. Finally, the tight temporal and spatial regulation that we observed for the core layer of the KP-Net can be justified by both the high load of signals received by this layer and its enrichment for KPs implicated in critical processes as cell cycle and decision-making.

Recently, Cheng et al. overlaid many of the biological properties studied here on top of a K-Net assembled from *in vitro* phosphorylation interactions in the budding yeast (Table A1.II) (Cheng et al., 2015). In contrast to our findings, most of the examined biological properties by Cheng et al. were statistically comparable among the three layers (gene essentiality, abundance, half-life and noise on mRNA and protein levels). It is important to note that properties of each layer depend on the identity and the properties of the proteins belonging to each layer. Difference between our findings and those of Cheng et al. could be due to: (i), the lack of phosphatases in the network analyzed by Cheng et al. (ii), the high number of false positives that normally exists in large scale data generated *in vitro*, which were used by Cheng et al. These false positives could affect sorting of nodes into the different layers and thus directly affect layer properties. (iii), Cheng et al. applied a decomposition method different from the VS algorithm. (iv), any combination of these reasons. Interestingly though, protein noise results of Cheng et al. concord

partially with our findings as proteins in the top layer were less noisy than those in the bottom layer.

A limitation of the KP-Net generated in this study is that it cannot be used to predict novel PDIs or pathways, to note that this was not among the objectives of this study. However, the KP-Net can serve as a gold standard in future investigations of signaling networks to suggest a set of KP candidates that might act on substrates under a given condition, as we showed in predicting kinases that act on substrates following osmotic stress. Another limitation is that although the choice of the largest SCC to represent the core layer was subjective and inspired by previous application of the VS algorithm to a transcription regulatory network, we can justify the validity of our choice by the concordance of our observations with those in the literature (Jothi et al., 2009). In the literature, a core layer of a bow tie structure is usually associated with critical decisions determining the system outputs (Tieri et al., 2010). This perfectly concords with our findings showing that 79 % of the core layer KPs are implicated in cell cycle and decision-making processes, to note that the VS algorithm does not necessarily generate a bow tie structure as in reference (Jothi et al., 2009) (Figure 2.2a; Table A1.IV). Finally, the assembled KP-Net represents a small snapshot of the real-world KP-Net affecting 60 % of the proteome. Advances in HTP technologies should eventually complete the KP-Net by unravelling missing PDIs. As with any network reconstruction exercise, there is the risk that a different sorting of KPs within the KP-Net hierarchical structure could lead to different interpretations of the KP-Net. However, when we randomly added edges to the KP-Net in order to create “noisy networks”, we observed that the layers of the noisy KP-Nets become less stable by adding more edges; but at the same time, they overlap significantly with KP-Net layers (Figure 2.7a and 7b). These results show that the properties of the KP-Net layers are robust to describe how the KP-Net functions with the best of our current knowledge, which represents the principal objective of this study.

Despite the limitations mentioned above, the functional principles of the KP-Net that are proposed in this study are consistent with other observations. Interestingly, bow tie structures are frequently associated with robustness against removal of some of their components and to external perturbations (Kitano, 2004; Ma and Zeng, 2003; Tieri et al., 2010; Whitacre, 2012). Robustness of the KP-Net bow tie structure could be ensured by the following factors. First, the degeneracy (overlapping functions) of many KPs in the top layer [e.g. PKAs, Tel1-Mec1 and calcineurins, (Figure 2.1b)] guarantees that failure of a KP to activate a given pathway is buffered

by another KP having partially redundant functions (Tieri et al., 2010). Notably, the degeneracy observed in top layer KPs concords well with the low number (13 %) of KPs encoded by essential genes belonging to this layer. Second, the core layer possesses the required features for generating coordinated responses: (i), it receives and integrates various inputs (high node in-degrees and enrichment for bottlenecks, pathway-shared components, and scaffold-associated KPs (Figure 2.3a, 3d, 3e and 4e); (ii), it occupies a central position in the hierarchy (Figure 2.1b); (iii), it is involved in a critical task (cell cycle and decision-making) (Figure 2.2a and Table A1.IV); and most importantly, (iv) it is excessively regulated at different levels in time and space. Without such a tightly regulated layer, coordinated responses would necessitate ample individual controls and any misregulation of the latter controls would easily impair cellular survival (Kitano, 2004). All these characteristics contribute in delineating functional principles of the KP-Net as known to date.

2.4.16 Conclusions

In this study, we built a KP-Net assembled from high quality PDIs in the budding yeast, determined its hierarchical structure and integrated the widest range of KP biological properties with elucidated hierarchical structure. This allowed us to formulate hypotheses about the functions of the KP-Net layers. As mentioned previously, the KP-Net assembled in this study represents a snapshot of the KP-Net that exists in the budding yeast. Advances in large-scale screens, in particular those exploring substrates of KPs will enhance coverage of the assembled KP-Net. Also, with the enhancement of HTP technologies, integration of other type of biological properties, such as methylation, ubiquitination, and temporal PDIs, with the KP-Net might become possible, which could reveal new functional principles of the KP-Net. A better perception of how the KP-Net functions could also open new opportunities to understand the actions of KP inhibitors on normal and pathological processes such as cancers.

2.5 Methods

2.5.1 Over-representation of various logic motifs in the KP-Net

One thousand random networks were generated by degree preserving randomization (DPR, Methods). Each of the random networks was sorted by the VS algorithm and the number of its

FFLs, FBLs, and bi-fan logic motifs was assessed. The *P*-value is the fraction of times the number of each logic motif in random networks is as large as that in the KP-Net.

2.5.2 Network randomization

In this study, we randomized the KP-Net using five types of network randomizations:

2.5.2.1 Degree preserving randomization (DPR)

We randomly selected two edges of the KP-Net and exchanged their ends. We then removed multiple edges having the same direction between two nodes by switching each of them with randomly selected edges. The rewiring procedure was repeated 10000 times to each random network.

2.5.2.2 Similar degree preserving randomization (SDPR)

We used the matching algorithm (Methods) to generate random graphs having similar degree distributions to that of the KP-Net (Milo et al., 2002; Molloy and Reed, 1995; Newman et al., 2001). We then switched network edges using the first randomization method (DPR) to make sure that the generated random networks differ from each other.

In-degree preserving randomization (IDPR): interactions were represented as a table made of two columns: “from” and “to”. We recreated the “from” column by randomly selecting KPs with replacement. We then switched network edges using the first randomization method (DPR).

2.5.2.3 Out-degree preserving randomization (ODPR)

We recreated the “to” column by randomly selecting KPs with replacement. We then switched network edges using the first randomization method (DPR).

2.5.2.4 Degree non-preserving randomization (DNPR)

We created a random network from scratch by connecting two nodes that were randomly selected with replacement.

2.5.3 The matching algorithm

In order to generate networks having a degree distribution that is similar to that of the KP-Net, we defined the degree distribution of the random network by randomly selecting three groups of

KPs: the first group had the same in- and out-degrees as the KP-Net and the second and third groups had the same in- and out-degrees as the KP-Net, but incremented and decremented by 1, respectively. Second, we connected the network using a variant of the matching algorithm (Milo et al., 2004). Briefly, each vertex of the random network was assigned a number of in- and out-stubs equal to its in- and out-degrees. In- and out-stubs were selected in pairs and joined up to make the network edges. In each step, the selection of in- and out-stubs was weighted by the square of the current in- and out-stubs that were not yet connected but should be. This procedure produced random networks that have very similar in- and out-degrees distributions to those of the KP-Net.

2.5.4 Testing whether the KP-Net GRC is bigger than Erdős–Rényi network GRCs

We generated 10000 Erdős–Rényi random networks having the same number of nodes and edges as the KP-Net and calculated their GRCs (Mones et al., 2012). The *P*-value of this test is the proportion of random network GRCs that are as large as the KP-Net GRC.

2.5.5 Comparing means of node properties in two layers using RT

Let L1 and L2 be the size of two layers of the KP-Net to be compared; S the set containing the nodes of these layers; S1 the set of L1 nodes randomly sampled without replacement from S; S2 the set of the remaining L2 nodes in S after sampling. The difference between the means of the node properties in S1 and S2 were calculated. These steps were repeated 10000 times. The *P*-value is equal to the proportion of times the difference between the means of the sampled sets are as big/small as the difference between the means of the two compared layers.

2.5.6 Generating subsampled/noisy networks and assessing their layers stability and their overlap with KP-Net layers

We generated ten sets of 100 subsampled and noisy networks from the KP-Net. The five sets of subsampled/noisy networks were produced by randomly removing/adding 40, 80, 120, 160 and 200 edges to the KP-Net, respectively. These steps were repeated 100 times, so each set contains 100 subsampled/noisy networks. We then applied the VS algorithm on each subsampled/noisy network to identify their three layers. Layers stability of the generated networks was assessed

using the Jaccard coefficient as a similarity “cluster wise” measure between the original and the subsampled/noisy layers (Henning, 2007). Overlap between original and subsampled/noisy layers were assessed using the hypergeometric test (HT).

2.5.7 Predicting kinases

First, we identified substrates in the KP-Net that contain a phosphorylated residue modulated by time after osmotic shock defined as dynamic phosphosite by Kanshin and Bergeron-Sandoval et al. (Kanshin et al., 2015). We also identified the consensus phosphorylation motives of each kinase in the KP-Net when possible from the literature (Table A1.V). We then connected each substrate containing a dynamic phosphosite to all kinases having a consensus motif matching the substrate phosphosite by edges to form kinase-substrates interactions. Using the KP-Net as a gold standard network, we retained kinase-substrate interactions that occur in the KP-Net.

2.6 Acknowledgements

The authors would like to thank Raja Jothi for helpful and instructive exchange, Abdelalli Kelil and Emmanuel Levy for helpful comments and stimulating discussions.

Chapter 3 Dissection of Cdk1-cyclin complexes *in vivo*

Po Hien Ear^{a1}, Michael J. Booth^{a,2}, **Diala Abd-Rabbo**^{a,b}, Jacqueline Kowarzyk Moreno^a, Conrad Hall^c, Daici Chen^{c,3}, Jackie Vogel^{c,4}, and Stephen W. Michnick^{a,b,4}

Article published in **PNAS**, 110(39):15716-21 (Ear et al., 2013)

Included with the permission of PNAS

^aDépartement de Biochimie, ^bCentre Robert-Cedergren, Bio-Informatique et Génomique, Université de Montréal, Montréal, QC, Canada H3C 3J7, ^cDepartment of Biology, McGill University, Montreal, QC, Canada H3G 0B1, ¹Present address actuelle: Department of Genetics, Harvard Medical School, Boston, MA 02115. ²Present address: Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, United Kingdom. ³ Present address: IRIC, Institut de recherche en immunologie et cancérologie, Université de Montréal, Montréal, QC, Canada, H3C-3J7. ⁴ To whom correspondence should be addressed : jackie.vogel@mcgill.ca or stephen.michnick@umontreal.ca.

3.1 Authors' contributions

In the present study, we used the *in vivo* PCA based on the OyCD to identify substrates of Cdc28 and to identify the cyclin or cyclins on which the interactions between Cdc28 and its substrates depend.

Po Hien Ear, Michael J. Booth, Jacqueline Kowarzyk Moreno, Conrad Hall, Jackie Vogel and Stephen W. Michnick designed research; Po Hien Ear, Michael J. Booth, Jacqueline Kowarzyk Moreno, Conrad Hall and Daici Chen performed research; Po Hien Ear, Michael J. Booth, Diala Abd-Rabbo, Jacqueline Kowarzyk Moreno, Daici Chen, Jackie Vogel and Stephen W. Michnick analyzed data; and Po Hien Ear, Michael J. Booth, Diala Abd-Rabbo, Jacqueline Kowarzyk Moreno, Conrad Hall, Daici Chen, Jackie Vogel and Stephen W. Michnick wrote the paper. My contribution to the present study is outlined in the following sub-sections.

3.1.1 Quantifying colony growth

Quantification of colony growth was automated using an ImageJ script that was developed by Kirill Tarassov in the Michnick laboratory. I enhanced this script to avoid missing the quantification of colonies at the first row or column of the analyzed plates. Colony growth is assessed by calculating colony integrated intensity. Six plates containing 384 colonies per plate, arranged in 16 rows by 24 columns, were photographed and resulting images were analyzed by the imageJ script. The script places a grid composed of 16 rows and 24 columns on each plate in a way that each colony fits in a cell of this grid; it then assesses the center and the radius of each colony in grid cells. Based on the radius, the program calculates the integrated density of each colony. The position of the grid is assessed automatically by attempting to localize the first and last rows and the first and last columns. To identify the first and last columns, 16 circles having a radius of 4 pixels and spaced by 39.85 pixels (space between centers of two adjacent colonies on a plate) are traced at five predetermined x coordinates (260, 262, 264, 266 and 268 to seek the first column and 1216.4, 1218.4, 1220.4, 1222.4 and 1224.4 to seek the last column). The pair of x coordinates for which the sum of the colony integrated intensities is the biggest is retained and is considered as the x coordinates of the first and last columns. To identify the first and last rows, 24 circles having a radius of 4 pixels and spaced by 39.85 pixels were traced at five predetermined y coordinates (235, 237, 239, 341 and 343 to seek the first row and 872.6, 874.6, 876.6, 878.6 and

880.6 to seek the last row). The pair of y coordinates for which the sum of the colony integrated intensities is the biggest is retained and is considered as the y coordinates of the first and last rows. Correctly localizing these rows and columns depends on finding the sought rows and columns on any combination of the predefined x and y coordinates. For instance, if the space between the image border and the first row does not match any of the first five predefined y coordinates, the first row will not be localized correctly. Therefore the program needed to be amended to remedy this drawback. I adjusted the program to automate the assessment of the grid position on the plate based on determining the position of the intersection between the upper border and the left hand border of the plate and this by supposing that these borders are located within the first 300 pixels of the image upper and left hand borders, respectively. Two sliding rectangles, one horizontal and another vertical, are drawn and their integrated intensities are calculated. Sliding begins from left ($x = 0$) to right for the vertical rectangle and from top ($y = 0$) to bottom for the horizontal rectangle of the image and stops at a maximum distance of 300 pixels from the border of the image. The x coordinate of the horizontal rectangle and y coordinate of the vertical rectangle that give the highest sum of their integrated intensities represent the upper left corner of the plate borders. This amendment permitted to avoid grid misalignment. After the grid is placed correctly, the enhanced script continues to assess colony growth as in the original script.

3.1.2 Preys in OyCD PCA are implicated in various biological processes

Among the 68 preys that were tested to interact with Cdc28, many were chosen from known or potential Cdc28 substrates from the literature or were chosen randomly. It was important to investigate whether the preys screened by the OyCD PCA are biased to be mostly implicated in a specific biological process and are not involved in diverse biological processes. To this end, I assessed the enriched GO terms in the list of the 68 screened preys using the conditional hypergeometric test implemented in the GOStats library (Falcon and Gentleman, 2007) of the R project for statistical computing environment (R-Development-Core-Team, 2011). GO term annotations were obtained from the org.Sc.sgd.db R package. This analysis consists of testing whether the number of screened preys associated to a GO term is larger than expected by chance. The conditional hypergeometric test is characterized by taking into consideration the hierarchical structure of the GO terms tree which causes the children to inherit their ancestors GO Terms. This test limits the list of over-represented GO terms to the most specific ones and omits all more

general GO terms that are over-represented, only because a group of their children were identified as significant.

Results of this analysis showed that the screened preys that were tested to interact with Cdc28 are enriched for diverse biological processes including: cell cycle processes, mRNA transcription, metabolic processes and mitotic cell cycle (Table 3.I). It is not surprising that the screened preys are enriched for mitosis and cell cycle processes, because some of them represent potential and already known substrates of Cdc28, the main regulator of cell cycle in the budding yeast. However, the screened preys are also enriched for other biological processes (transcription and metabolic processes). This result was briefly reported in Paragraph 3.5.1 as follows: “These proteins included those involved in transcription, cell cycle regulation, and mitotic spindle assembly”.

Table 3.I. Top enriched GO terms in the preys that were tested to interact with Cdc28

GO Term Description	Associated genes #	Fold enrichment	P-value
cell cycle phase transition	11	14.4	1.5×10^{-10}
cell cycle	19	4.7	3.75×10^{-9}
biological regulation	42	2.2	4.33×10^{-9}
regulation of RNA biosynthetic process	25	3.4	1.68×10^{-9}
regulation of cell cycle	13	6.6	4.88×10^{-8}
regulation of transcription involved in G1/S transition of mitotic cell cycle	6	21.2	2.66×10^{-7}
positive regulation of nucleic acid-templated transcription	15	4.8	2.96×10^{-7}
positive regulation of RNA metabolic process	15	4.7	4.35×10^{-7}
nucleobase-containing compound biosynthetic process	27	2.7	5.06×10^{-7}
cell division	14	4.9	5.42×10^{-7}
mitotic cell cycle process	7	13.8	5.48×10^{-7}
regulation of transcription, DNA-templated	21	3.2	7.67×10^{-7}

3.1.3 Comparison of OyCD PCA data with other datasets

In order to assess the performance of the OyCD PCA screen, I tried to compare results of each of the first and second screens of the OyCD PCA with other already published datasets.

3.1.3.1 Comparison of first screen of OyCD PCA data with other datasets

The first screen of the OyCD PCA in Ear et al. represents a novel experimental method to identify proteins interacting with Cdc28 *in vivo*. Therefore, it was important to evaluate the quality of the generated results. To this end, I attempted to assess the significance of the overlap between Cdc28 interactors identified by the OyCD PCA in Ear et al. and those reported in six other studies: (1), 299 Cdc28 interactors identified by a large screen performed, in the Michnick laboratory, using the PCA based on the murine dihydrofolate reductase, called hereafter Cdc28 2010 screen (unpublished data); (2), 600 cell cycle-regulated or periodically expressed genes that de Lichtenberg et al. integrated with interaction data to analyze the dynamics of protein complexes during cell cycle in the budding yeast (de Lichtenberg et al., 2005); (3), 1129 genes that were identified to be cell cycle-regulated by Rowicka et al. using an algorithm based on maximum-entropy deconvolution applied on expression data from highly synchronized cells in *Saccharomyces cerevisiae* (Rowicka et al., 2007); (4), 185 proteins phosphorylated by an analogue-sensitive Cdc28 complexed with the cyclin Clb2, a regulatory subunit of Cdc28, in whole-cell extracts (Ubersax et al., 2003); (5), 308 phosphoproteins having a significant decrease in their level of phosphorylation on inhibiting Cdc28 in budding yeast cells (Holt et al., 2009); and finally (6), 85 Cdc28 interactors annotated in the KID database having a KID score > 6.29 corresponding to a P -value ≤ 0.01 . Overlapping Cdc28 substrates in Ear et al. with proteins coded by cell cycle regulated genes, called hereafter periodic genes, was motivated by an enrichment of Cdc28 substrates in periodic genes that was observed by de Lichtenberg et al. [$P < 10^{-4}$, hypergeometric test, (de Lichtenberg et al., 2005)].

Table 3.II shows that there is no significant overlap between Cdc28 interactors identified by the OyCD PCA in Ear et al. and those identified in all the other studies used for the comparison. This result is not surprising for the following reasons: first, the overlap with Cdc28 2010 screen is not significant, because on one hand, the Cdc28 2010 screen was performed as a pilot study that contains only one replicate which decreases the level of confidence in these results in comparison to results of the four replicates generated by Ear et al. and on the other hand, this screen used the murine dihydrofolate reductase PCA which was not well optimized to screen Cdc28 interactors as the OyCD PCA was. Second, proteins coded by periodic genes are known to have a cyclic expression; they peak in one or more phases of the cell cycle (Rowicka et al., 2007). De Lichtenberg et al. observed that proteins coded by periodic genes are enriched for both: Cdc28

substrates and proteins containing PEST regions which represent regions rich in Proline, Glutamine, Ser and Thr serving as degradation signals [$P < 10^{-2}$, hypergeometric test, (de Lichtenberg et al., 2005)]. De Lichtenberg et al. suggested that Cdc28 phosphorylates proteins coded by periodic genes on residues in PEST regions to be targeted to the proteasome for degradation (de Lichtenberg et al., 2005). The non-significant overlap between Cdc28 interactors identified by the OyCD PCA and proteins coded by periodic genes identified in studies (2) and (3) could be because these periodic proteins might be regulated not only by Cdc28 at the posttranslational level, but also by other kinases or they might even be regulated at the posttranscriptional level (e.g. mRNA degradation). Third, Cdcd28 substrates identified by Ubersax et al. do not overlap significantly with those identified by the OyCD PCA because Ubersax et al. might have missed many of Cdc28 substrates, as Ubersax screen used only one out of the nine cyclins that represent regulatory subunits of Cdc28 and that play an important role in determining specificity of Cdc28 towards its substrates (Ubersax et al., 2003). Moreover, Cdc28 substrates identified by Ubersax et al. might include false positives, since phosphorylation of substrates by Cdc28 took place in whole cell extracts, in which spatial and temporal regulations are missing. Fourth, the overlap between Cdc28 substrates identified by Holt et al. (Holt et al., 2009) and those identified in our study is not significant probably due to the enrichment of Cdc28 substrates identified by Holt et al. for indirect substrates of Cdc28. Proteins having a decreasing level of phosphorylation on inhibiting Cdc28 include not only substrates of Cdc28, but also substrates of phosphatases that are inhibited by Cdc28 and substrates of kinases that are activated by Cdc28. Finally, the overlap between Cdc28 interactors collected from the KID database (Sharifpoor et al., 2011) do not also overlap significantly with those identified in our screen perhaps because some interactions are poorly or not well annotated in the KID database, as we previously mentioned in chapter 2 (Paragraph 2.4.1). All these reasons reveal the original nature of the OyCD PCA screen and thus the original results generated by this screen.

Given the unique nature of the OyCD PCA assay and consequently the lack of a dataset that overlaps significantly with OyCD PCA data, we evaluated the performance of the assay by assessing its false-negative rate (see Appendix 2).

Table 3.II. Significance of the overlap between Cdc28 interactors identified in Ear et al. and in other datasets.

Study (S)	# of Cdc28 interactors in Ear et al. and tested in S	# of proteins not interacting with Cdc28 in Ear et al. and tested in S	# of Cdc28 interactors in Ear et al. and in S	# of Cdc28 interactors in S and tested in Ear et al.	P-value
(1), Cdc28 2010 screen	33	30	10	17	0.369
(2), (de Lichtenberg et al., 2005)	37	32	14	24	0.375
(3), (Rowicka et al., 2007)	37	32	17	32	0.625
(4), (Ubersax et al., 2003)	37	32	15	27	0.496
(5), (Holt et al., 2009)	37	32	14	22	0.189
(6), (Sharifpoor et al., 2011)	37	32	11	17	0.219

3.1.3.2 Comparison of second screen of OyCD PCA data with other datasets

The second screen of the OyCD PCA in Ear et al. is the first to investigate cyclin specificity towards Cdc28 substrates *in vivo*. Finding a significant overlap of Cdc28 substrates that depend on specific cyclins to interact with Cdc28, identified in Ear et al., with other datasets was very difficult due to two reasons. First, similar studies are very limited. Only three screens investigated the specificity of cyclin binding towards Cdc28 substrates: (i), a large screen was performed *in vitro* by Loog and Morgan using two cyclins (Clb2 and Clb5); (ii), a LTP study was achieved *in vitro* by Koivomagi et al. using four cyclins (Cln2, Clb2, Clb3 and Clb5); and (iii), a HTP study was performed also *in vitro* by Ptacek et al. using two cyclins (Cln2 and Clb5). Second, the three latter studies were not performed *in vivo* and they did not test the specificity of Cdc28 substrates towards the nine cyclins as in Ear et al.. The three studies together investigated the binding specificity of four out of the nine cyclins. With this in mind, I assessed the sensitivity and specificity of the second screen performed in Ear et al. in an attempt to overcome the difficulty of benchmarking the data generated by this screen (see Appendix 2).

3.1.4 Other contributions

Finally, I contributed in making several figures: Figure 3.2b, Figure 3.3a (in particular the bar plots) and Figure A2.1. I also contributed in performing the analysis reported in the supporting information. Finally, I contributed in the curation and management of Table A2.IV and Table A2.V.

3.2 Abstract

Cyclin-dependent kinases (Cdks) are regulatory enzymes with temporal and spatial selectivity for their protein substrates that are governed by cell cycle-regulated cyclin subunits. Specific cyclin–Cdk complexes bind to and phosphorylate target proteins, coupling their activity to cell cycle states. The identification of specific cyclin–Cdk substrates is challenging and so far, has largely been achieved through indirect correlation or use of *in vitro* techniques. Here, we use a protein-fragment complementation assay based on the optimized yeast cytosine deaminase to systematically identify candidate substrates of budding yeast *Saccharomyces cerevisiae* Cdk1 and show dependency on one or more regulatory cyclins. We identified known and candidate cyclin dependencies for many predicted protein kinase Cdk1 targets and showed elusive Clb3–Cdk1-specific phosphorylation of γ -tubulin, thus establishing the timing of this event in controlling assembly of the mitotic spindle. Our strategy can be generally applied to identify substrates and accessory subunits of multisubunit protein complexes.

3.3 Significance

A simple strategy is described to discover cyclin dependency of general cell cycle regulatory kinase cyclin-dependent kinase 1 for substrates *in vivo*. γ -Tubulin is discovered to be a specific B cyclin Clb3–cyclin-dependent kinase 1 substrate. The strategy can link proteins to specific phases of the cell cycle, and it can be used to determine subunit dependencies for other posttranslational modifying enzymes.

3.4 Introduction

A central problem in biology is determining the functions of individual enzyme subunits and the synergistic relationships among them. For example, enzymes that perform posttranslational modifications on proteins, such as the ubiquitin ligases, protein kinases and protein phosphatases, require different subunits to perform multiple transfer steps, assure specific subcellular localization, or provide additional specificity to substrate recognition (Cohen, 1989; Janssens et al., 2008; Morgan, 1997; Thornton and Toczyski, 2006). Cdks are a case in point, and the budding yeast *Saccharomyces cerevisiae* Cdk1 is a very well-studied example of an enzyme of this category (Morgan, 1995). Cdk1 requires the association of one of nine available cyclin partner proteins to recognize and phosphorylate its substrates (Mendenhall, 1993; Mendenhall and Hodge, 1998). The different Cdk1–cyclin complexes play critical roles in orchestrating the temporal and spatial ordering of events from initiation of the G1 transcriptional program (Cln1, -2, and -3) to DNA replication (Clb5 and -6), spindle assembly (Clb3 and -4), and mitosis (Clb1 and -2) (Bloom and Cross, 2007).

The crucial role of Cdk1 in cell cycle regulation has prompted several extensive or proteome-wide studies to identify Cdk1 substrates or cyclin targets (Archambault et al., 2004; Holt et al., 2009; Loog and Morgan, 2005; Ubersax et al., 2003). To date, no experimental approach has captured interactions between Cdk1 and its substrates and the dependency of this interaction on one or more cyclins in the context of a living cell. In this study, we describe an approach that captures direct interactions between Cdk1 and its substrates and reveals the dependency of this interaction on one or more cyclins in living cells.

We devised a simple *in vivo* screening strategy to both identify potential Cdk1 substrates and establish dependencies of the Cdk1 interactions with these substrates on specific cyclins using the optimized yeast *S. cerevisiae* prodrug-converting enzyme cytosine deaminase protein-fragment complementation assay (OyCD PCA) (Figure 3.1) (Ear and Michnick, 2009). The OyCD PCA consists of two complementary N- and C-terminal fragments (OyCD-F[1] or F[2]) of the FCY1 gene (FCY1) fused individually to two genes of interest and performed in an FCY1 Δ deletion strain (Ear and Michnick, 2009). If the proteins of the two protein–OyCD fusion fragments interact, OyCD refolds from its complementary fragments, reconstituting its enzymatic activity. The assay provides two potential outputs: positive growth under cytosine-limited

conditions or no growth when cells are treated with a yCD-specific prodrug called 5-FC (Ear and Michnick, 2009). Positive selection assays depend on careful control of colony density, number of cells plated, and composition of media (Kurtz et al., 1999). Here, we exploited the 5-FC negative selection assay of the OyCD PCA to provide sensitive detection of Cdk1 interactions as well as their cyclin dependence. First, we performed the OyCD screen with Cdk1 as bait and potential substrates as prey in the negative selection assay. Second, we retested the observed Cdk1–prey interactions in strains in which individual cyclin genes were deleted.

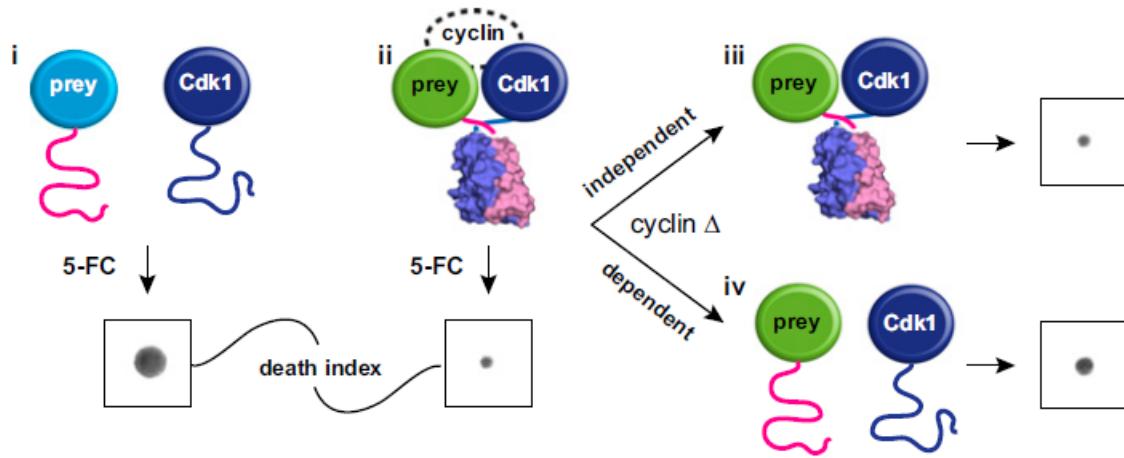


Figure 3.1. Dissecting Cdk1 complexes using the OyCD PCA.

Detecting the interaction between Cdk1 and a protein of interest using the OyCD PCA. Cdk1 and proteins of interest (prey proteins) are fused to OyCD fragments. In the death selection OyCD PCA, (i) the absence of an interaction fails to allow the OyCD reporter enzyme to fold and restore the activity of the native enzyme. Cells expressing these fusion proteins are resistant to 5-FC. (ii) If prey protein interacts with Cdk1, cells are sensitive to 5-FC. When a cyclin gene is deleted (cyclin Δ), (iii) a prey protein can still interact with Cdk1, allowing the reporter fragments to fold; consequently, cells are sensitive to 5-FC. The prey protein–Cdk1 interaction is independent of any cyclin. (iv) If the interaction is cyclin-dependent, the absence of a specific cyclin results in partial or total resistance to 5-FC.

We could predict four potential outcomes for the screens (Figure 3.1). In a primary screen, (Figure 3.1i) if Cdk1 does not interact with a prey protein in the WT strain (Figure 3.1i), cells are insensitive to 5-FC (colony growth). (Figure 3.1ii) However, if there is an interaction, OyCD PCA activity is reconstituted, and cells become sensitive to 5-FC (no colony growth). In the second screen of Cdk1–prey interactions in individual cyclin deletion strains, (Figure 3.1iii) if none of the cyclins were essential to the interaction, we would expect identical results to those observed in the WT strain (no growth in cells treated with 5-FC). (Figure 3.1iv) In cases where the Cdk1–prey

interaction depends on one or more than one cyclin, we would expect to see restoration of total or partial growth when a certain cyclin knockout strain is grown on 5-FC.

3.5 Results

3.5.1 Identifying Cdk1 complexes *in vivo*

We selected 94 known or potential Cdk1 substrates from the literature or randomly chosen proteins as controls (de Lichtenberg et al., 2005; Ubersax et al., 2003). These proteins included those involved in transcription, cell cycle regulation, and mitotic spindle assembly. Of 94 candidates, we successfully generated OyCD PCA expression vectors of 68 candidate open reading frames (ORF) fused to OyCD-F[1]. Expression plasmids were cotransformed into the FCY1 Δ deletion strain with Cdk1 fused to the complementary OyCD F[2] or a control plasmid that only expresses the OyCD-F[2]. We performed the OyCD PCA by inoculating two colonies from each transformation and growing them on medium with or without 5-FC at a concentration of 1 mg/mL; 38 of 68 preys showed a positive interaction with Cdk1–OyCD-F[2] (Figure 3.2a and Table A2.IV). One prey-expressing strain (Cdc19) among thirty-eight strains gave a false-positive signal when expressed with the fragment OyCD-F[2] alone (indicated in gray in Figure 3.2a) and was not considered further.

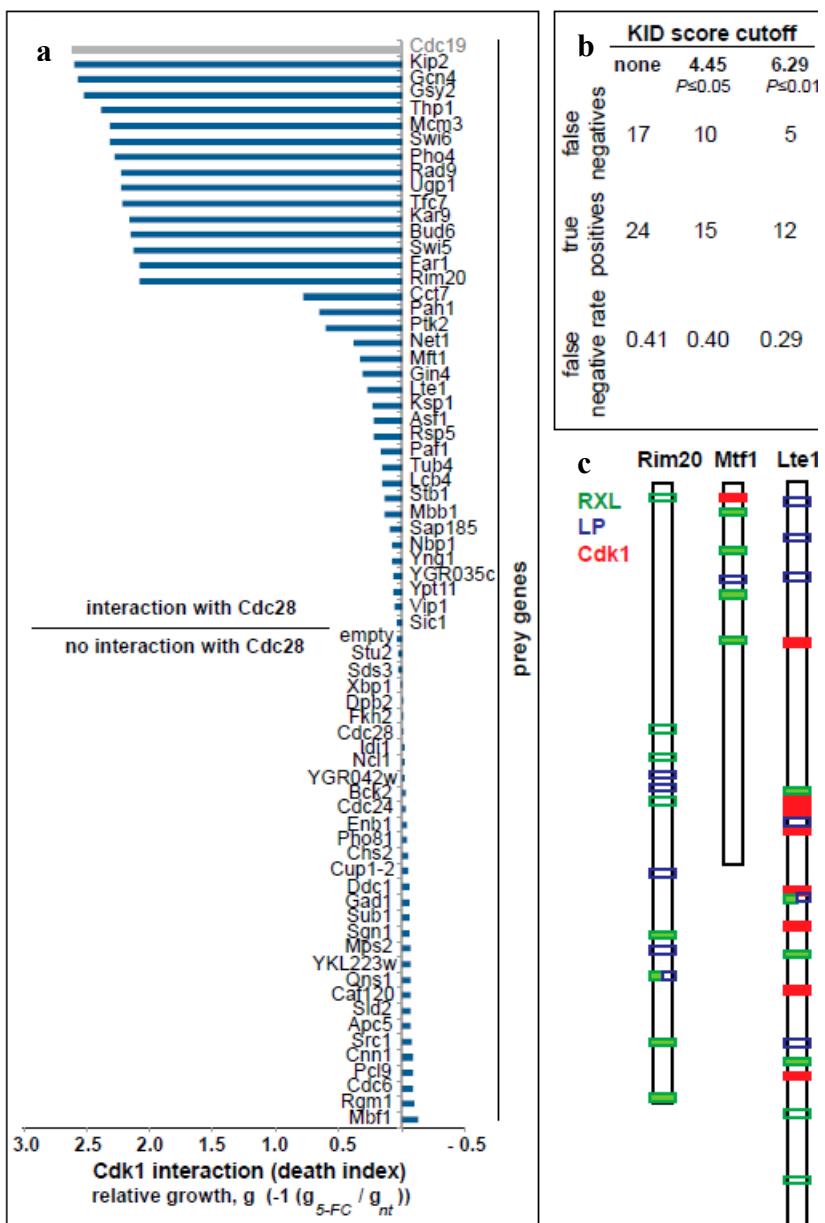


Figure 3.2. Identification of interaction partners of Cdk1.

(a) Quantification of Cdk1–prey protein interactions. Death index is calculated as the \log_{10} of the ratio of pixel mean intensities of 5-FC-treated divided by untreated colonies (5-FC/not treated). The Cdk1–Cdc19 interaction was found to be a technical false positive (see text). **(b)** Estimation of the FNR for 37 proteins identified to interact with Cdk1 by OyCD PCA (details of calculation in Appendix 2). **(c)** Examples of Cdk1 interacting partners with cyclin binding motifs and with (Lte1 and Mft1) or without (Rim20) consensus Cdk1 phosphosites. Filled bars are high-quality sites ($P \leq 0.01$, open bars are low-quality sites).

We used the KID database (Sharifpoor et al., 2011) as a gold standard to calculate false negatives (FN) and true positives (TP) in 37 prey proteins that interacted with Cdk1 in the OyCD

PCA, and in this way, we estimated the false-negative rate [FNR; $FNR = FN/(TP + FN)$] to be 29 % for the most significant KID scores (KID score for $P \leq 0.01$) (Figure 3.2b and Appendix 2). Twenty-two of 37 preys were previously identified as Cdk1 substrates, and the remaining 15 preys have not been linked to a specific cyclin–Cdk1 complex (Holt et al., 2009; Keck et al., 2011; Ptacek et al., 2005; Ubersax et al., 2003). Of 15 unique interactions observed, 13 of the prey had full [S/T]PX[K/R] or minimal [S/T]P Cdk1 consensus sites, two criteria previously used to define likely Cdk1 substrates (Figure 3.2b and Table A2.I) (Holt et al., 2009; Ubersax et al., 2003). One of these candidates, Rim20, does not have a full or minimal Cdk1 consensus site but has four high-quality cyclin binding motifs [RXL; $P \leq 0.01$, Eukaryotic Linear Motif (ELM) database] and five LP motifs that have been previously implicated in G1 cyclin–substrate binding in budding yeast (Toyoshima and Hunter, 1994). Rim20 is a regulator of Ime2, a protein kinase involved in activating meiosis (Su and Mitchell, 1993). Rim20 resembles cyclin–Cdk inhibitors, such as Sic1 and p27Kip1, and has one or more RXL cyclin binding motifs (Russo et al., 1996; Schwob et al., 1994; Toyoshima and Hunter, 1994). More typically, proteins contained various numbers of G1 and B-type cyclin binding motifs and minimal Cdk1 phosphorylation motifs (Table A2.I). For example, Mft1, a protein involved in mitotic recombination (Chavez et al., 2000), has five cyclin binding motifs (four RXL and one LP) and one minimal Cdk1 site. Lte1, a spindle-positioning checkpoint protein that regulates the Ras-like small GTPase Tem1 (Bardin et al., 2000), has many sites, including 6 RXL, 5 LP, and 8 full and 20 minimal Cdk1 sites. Phosphorylation of Lte1 by Cdk1 regulates the transition from apical to isotropic growth (Geymonat et al., 2010).

3.5.2 Cyclin dependency of Cdk1 complexes

We next tested whether the interactions between Cdk1 and prey were contingent on a particular cyclin. We performed the OyCD PCA in nine cyclin deletion strains (cln1-3 and clb1-6) for 21 of 37 proteins that interact with Cdk1 (Figure 3.3a, Figure A2.1, and Table A2.I). As a positive control, to assure that none of the cyclin deletion strains affected the performance of the OyCD PCA, we tested a constitutive homomeric GCN4 leucine zipper-forming peptide (Zip) interaction with the OyCD PCA in all the different strains. As a negative control, we tested for the activity of OyCD PCA in the different strains expressing Cdk1–OyCD-F[2] alone. All results were evaluated by taking the ratio of integrated colony intensity of each sample grown on selection medium with 1 mg/mL 5-FC divided by the intensity of colonies grown on selection medium without 5-FC.

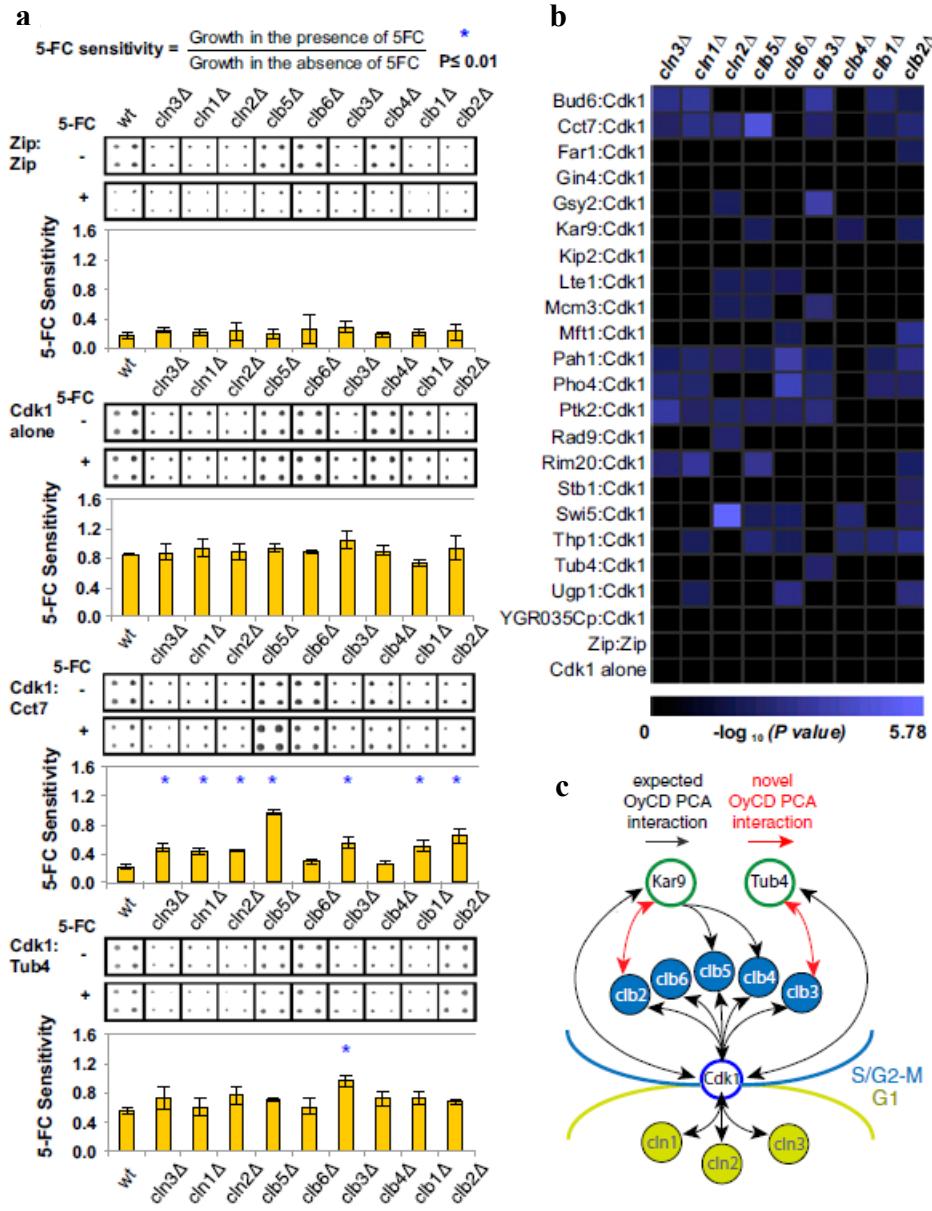


Figure 3.3. Cyclin dependence of Cdk1–prey protein interactions.

(a) Quantification of cyclin dependence. The overexpression of some fusion genes slightly affects the fitness of the yeast strain compared with yeast expressing only Cdk1 or Zip:Zip fused to OyCD fragments. The deletion of some cyclin genes affects the ability of the strain to grow compared with the WT strain. The effect of the OyCD PCA activity on growth is relatively constant in the different cyclin deletion strains (mean ± SD). Statistical significance was assessed using Student t test. $P = 0.04$ was obtained for WT and *cld3Δ* strains expressing Zip:Zip interaction. $P = 0.02$ was obtained for WT and *cld1Δ* deletion strains expressing Cdk1 alone (contingency of the Cdk1 complexes). The gene of interest and Cdk1 fused to OyCD fragments were transformed into the FCY1Δ deletion strain, which is referred to as the WT, as well as in nine different cyclins and FCY1Δ double deletion strains, which are represented by their gene name in italic followed by a Δ (e.g., *cld1Δ*). Four colonies of each transformation were assayed for OyCD PCA

(Figure 3.3 continued) activity in the presence of 1 mg/mL 5-FC in three different experiments. The growth of each sample was quantified using ImageJ. Only the results of one set of a quadruple experiment are represented. All strains expressing only Cdk1–OyCD-F[2] (Cdk1 alone) are resistant to the 5-FC death selection assay with $P < 0.02$. All strains expressing the GCN4 leucine zipper domains (Zip:Zip) fused to the OyCD fragments are sensitive to 5-FC in the death selection assay with a $P < 0.04$. A loss of interaction detected in the different cyclin deletion strains is depicted with corresponding P value. $P \leq 0.01$ was used as a cutoff for this experiment (indicated with blue asterisk). (b) Results for $P \leq 0.01$ are represented on the matrix (Tarassov and Michnick, 2005). (c) Summary of previously described (black arrows) and discovered (red arrows) interactions between Cdk1–cyclin complexes, Kar9, and Tub4 and timing of their functions in the mitotic cell cycle.

We observed that the overexpression of Cdk1 combined with some of the proteins (Thp1, Bud6, and Gin4) did affect the growth of the yeast strains compared with control strains overexpressing Cdk1 alone or the Zip:Zip complex (Figure A2.1 and Table A2.V). We also found that some cyclin deletion strains grew poorly compared with the WT strain (e.g., *CLN1-3* and *CLB1-3*) (Figure 3.3a). However, the effect of Cdk1 overexpression affected the WT and cyclin deletion strains uniformly and did not affect the OyCD PCA activity for all of the different yeast strains (Figure 3.3a and Figure A2.1). Overall, the effect of the OyCD PCA activity was dominant over the effect of strain variability as well as over the effect of overexpression of the two genes of interest.

We noted that we did not observe complete loss of 5-FC sensitivity for a Cdk1–prey protein interaction in any of the cyclin deletion strains compared with a negative control strain expressing only Cdk1–OyCD-F[2]. Among the potential reasons for this difference is that a unique cyclin was not responsible for the Cdk1–prey protein interaction in any of the cases studied here. Other cyclins could, thus, compensate for the deleted one. Also, residual binding of Cdk1 to prey proteins may always occur, despite deletion of specific cyclins. Finally, other proteins may also contribute to observed Cdk1–prey protein binding.

To compare the activity of the OyCD PCA of each interaction in the 10 different yeast strains (WT or cyclin null), a Student t test was performed using the OyCD PCA activity obtained. For the Zip:Zip control, we observed minor differences in growth in the different strains compared with the WT strain, but the strain background did not significantly affect results compared with the assay in any case (Student t test, $P < 0.04$). We did not observe significant differences among strain results for a negative control Cdk1–OyCD-F[2] expressed alone (Student t test, $P < 0.02$). We considered results with a $P \leq 0.01$ to be significant, because this cutoff is more stringent than the lowest P value for the PCA activity of cells expressing the Cdk1 alone control in the *CLB1Δ*

deletion strain (Figure 3.3a and b and Figure A2.1). Sensitivity and specificity of the test were calculated to be between 67 % and 100 % and 60 % and 63 %, respectively, depending on reference data used (Appendix 2).

Among 21 preys that interacted with Cdk1, three proteins (Gin4, Kip2, and YGR035C) remained unchanged in all of the cyclin deletion strains with respect to the WT strain, similar to the Zip:Zip interaction. The interaction between Cdk1 and the remaining 18 preys decreased in one or more cyclin deletion strains (decreased sensitivity to 5-FC). For example, growth of yeast expressing the γ -tubulin ortholog Tub4 and Cdk1 increased when the *CLB3* gene was deleted alone. We observed the same decrease in sensitivity to 5-FC for Far1 and Stb1 in the *CLB2 Δ* deletion strain and Rad9 in the *CLN2 Δ* deletion strain. In contrast, the majority of interactions between Cdk1 and prey were dependent on more than one cyclin. In some cases, multiple cyclin interactions are consistent with previous findings; for example, the Kar9–Cdk1 interaction was dependent on *CLB2*, *CLB4*, and *CLB5*. Likewise, the interaction between Rim20 and Cdk1 was dependent on G1, S-phase, and mitotic cyclins: *CLN3*, *CLN1*, *CLB5*, and *CLB2*. Efficient targeting by G1 cyclin–Cdk1 protein complexes (Cln1, -2, and -3) is promoted by LP motifs, five of which are clustered among the four RXL motifs located in the carboxyl terminal of Rim20. Indeed, all of the proteins with interaction with Cdk1 that depend on G1 cyclins (Rim20, Lte1, Bud6, Swi5, Cct7, Ugp1, Gsy2, Pah1, Rad9, Mcm3, Thp1, Ptk2, and Pho4) contain both canonical RXL cyclin binding motifs as well as LP motifs (Table A2.I), the latter residing within a patch of hydrophobic residues as previously described (Koivomagi et al., 2011). Finally, the distribution of cyclins associated with Lte1 is consistent with its phosphorylation and binding partners during the cell cycle. Phosphorylation of Lte1 during G2/M relocates the protein from the bud tip isotropically over the cortex of the bud, thereby inhibiting polarized growth as cells progress to mitosis. This inhibition is achieved through phosphorylation of Lte1 by the PAK kinase Cla4, which requires G1 cyclin activity and subsequent, hyperphosphorylation by Cdk1 in S/G2 (Geymonat et al., 2010). The association of Lte1 with both G1 and S/G2 cyclins is consistent with its multiple roles (control of polarization and spindle positioning) during the transition from S phase to mitosis.

3.5.3 γ -Tubulin is an *in vivo* target of Cdk1–Clb3

Our results provide specific molecular mechanistic insights into regulation of the function of the mitotic spindle, which has numerous component proteins that are Cdk1 targets (Figure 3.3c) (Holt et al., 2009). Cdk1 promotes early spindle positioning to the bud neck by driving the asymmetric distribution of Kar9 to the old spindle pole and associated cytoplasmic microtubules (Liakopoulos et al., 2003). Interaction and phosphorylation of Kar9 is mediated by several cyclins, including Clb3, Clb4, and Clb5 (Liakopoulos et al., 2003; Moore and Miller, 2007). Our finding that Kar9 binds to Clb2 could arise from the close association of Kar9, dynein, and dynactin at the positive ends of cytoplasmic microtubules associated with cortical actin (Grava et al., 2006). In addition, we found that the evolutionarily conserved microtubule nucleator γ -tubulin, Tub4 in budding yeast, is a target of Clb3–Cdk1 but not the other S-phase and mitotic cyclin–Cdk1 complexes Clb1 and -2 and Clb4, -5, and -6. This finding suggests that Tub4 regulation could be temporally coupled with but stay distinct from the Clb4–Cdk1 regulation that enforces Kar9 asymmetry during spindle assembly, which is concomitant with spindle orientation perpendicular to the future plane of cytokinesis at the bud neck (Liakopoulos et al., 2003) (Figure 3.3c).

Tub4 S360, which is in a conserved Cdk1 recognition motif, was previously shown to be a mitotic target in a phosphopeptide analysis of highly purified SPB (Keck et al., 2011). Phosphomimetic mutations in S360 (S360D/E) caused defects in spindle function. Keck et al. (Keck et al., 2011) showed that S360 is specifically phosphorylated *in vitro* by Clb2–Cdk1. Clb2–Cdk1, however, is the most promiscuous cyclin–Cdk1 complex, because it does not require a cyclin docking motif and thus, can phosphorylate any minimal [S/T]P Cdk1 recognition motif (Koivomagi et al., 2011). Furthermore, the SPBs used for the phosphopeptide analysis were isolated under a metaphase arrest condition produced by depletion of the anaphase promoting complex activating protein Cdc20, in which all B-type cyclins (Clb1 -2, -3, -4, -5, and -6) are expected to be active (Rahal and Amon, 2008; Tavormina and Burke, 1998). It was, thus, not possible in that study to unambiguously identify which cyclin–Cdk1 complex was specific for Tub4 S360 phosphorylation. Knowledge of the cell cycle timing and therefore, the one of six possible cyclin–Cdk1 complexes that phosphorylates S360 is critical to understanding when and how γ -tubulin phosphorylation contributes to spindle function. Our evidence suggests that it is Clb3, a B cyclin that accumulates during S phase and spindle assembly (Figure 3.4a), which is an

important candidate Cdk1 partner for recognizing and phosphorylating γ -tubulin. We, thus, decided to test whether Clb3–Cdk1 could specifically phosphorylate γ -tubulin *in vitro*.

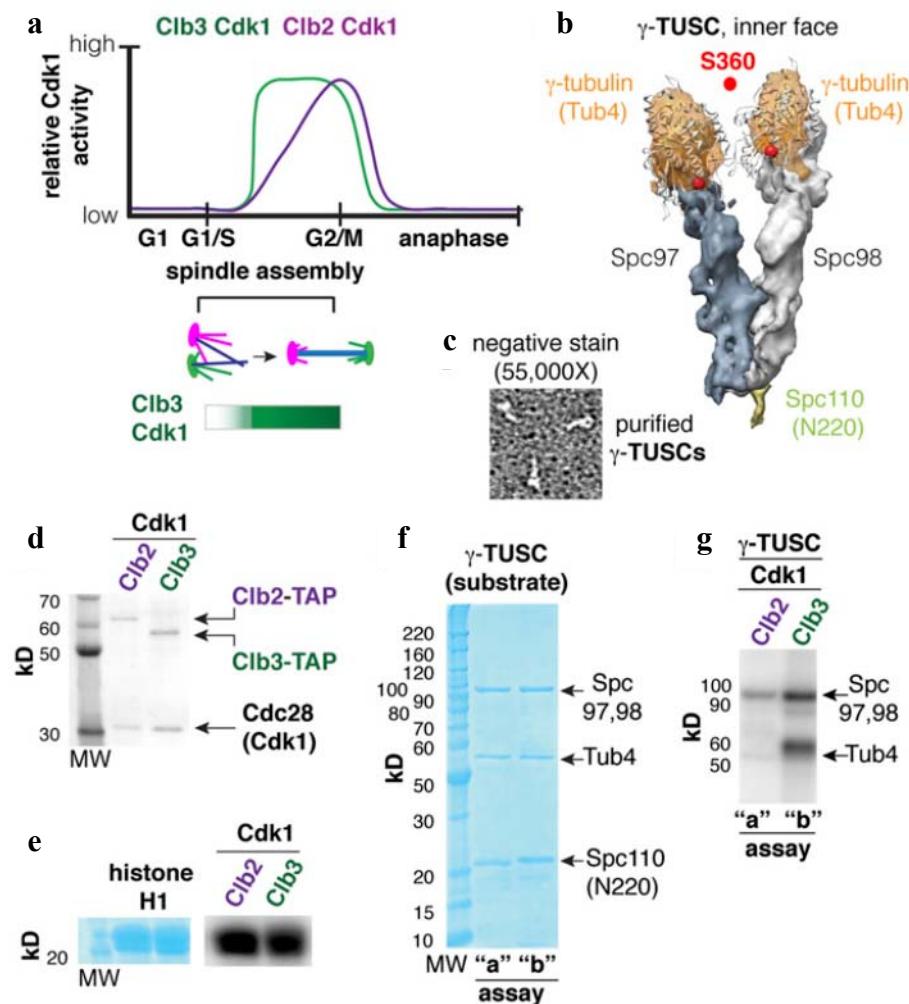


Figure 3.4. γ -Tubulin is a Clb3–Cdk1 substrate *in vitro*.

(a) Early and late B-type cyclin–Cdk complexes (e.g., Clb3 and Clb2) are active during spindle assembly, with Clb3 activity peaking before Clb2 reaches its maximum activity. (b) The γ -tubulin small complex (γ -TUSC) is composed of Tub4, Spc97, and Spc98 (2:1:1 stoichiometry). γ -TUSCs are Y-shaped complexes that form spontaneously in coinjected Sf9 cells in the presence of a fragment of Spc110 (N-terminal 220). (c) Verification of the structure of purified γ -TUSCs using negative staining and transmission electron microscopy (TEM). (d) Clb–Cdk1 complexes purified from yeast. (e) Histone H1 was used as a control to verify Clb–Cdk1 activity. (f) γ -TUSC substrate (0.4 μ M) reacted with Clb–Cdk1 in parallel kinase assays: a, 2 nM Clb2–Cdk1; b, 5 nM Clb3–Cdk1 for 30 min at 30 °C. (g) Tub4 was efficiently phosphorylated by Clb3–Cdk1 but not Clb2–Cdk1, whereas Spc97 and -98 are phosphorylated by both Clb2–Cdk1 and Clb3–Cdk1.

3.5.4 Clb3–Cdk1 preferentially targets γ -tubulin *in vitro*

Active, functional Tub4 located at SPBs is not found as a monomer but rather, is found in a large complex with two other evolutionarily conserved proteins: Spc97 and Spc98. This complex, called the γ -tubulin ring complex (γ -TURC), is composed of seven Y-shaped smaller complexes called γ -TUSCs (Figure 3.4b), each composed of two molecules of Tub4 and one molecule each of Spc97 and Spc98 (Kollman et al., 2010). γ -TUSCs are anchored to the SPBs by receptor proteins, one of which is Spc110 (Kollman et al., 2010). Using an Sf9 insect cell expression system, γ -TUSCs form spontaneously on coexpression and can be purified by affinity purification with GST-tev-Spc110^{1–220} fusion protein as the ligand (Kollman et al., 2010; Vinh et al., 2002). We reasoned that purified γ -TUSCs would be appropriate as a substrate to test the specificity of cyclin–Cdk1 complexes for two reasons: first, this approach would test *in vitro* Cdk1 targeting of Tub4 in a state that is similar to that in which it exists *in vivo*, and second, Spc97 and Spc98 serve as internal controls for relative specificity. We purified γ -TUSCs as previously described (Vinh et al., 2002) and verified their structure as Y-shaped particles using negative staining EM (Figure 3.4c). We then performed parallel ³²P (ATP) incorporation assays in the presence of purified Cdk1 (Cdc28) and cyclin (Figure 3.4d) using either histone H1 (Figure 3.4e) or γ -TUSCs (Figure 3.4f) as substrates as described in Materials and Methods and in (Keck et al., 2011). Although Spc97 and/or Spc98 were targeted by both Clb2–Cdk1 and Clb3–Cdk1, we found that Clb3–Cdk1 selectively phosphorylated Tub4 with no detectable phosphorylation by histone H1-normalized Clb2–Cdk1 (Figure 3.4g). These *in vitro* assays support our *in vivo* finding that Clb3–Cdk1 is the most likely Cdk1 complex targeting Tub4 in cells.

3.5.5 Discussion

It has proven extremely difficult to distinguish specific *versus* redundant roles of a cyclin in cyclin–Cdk1 protein complexes using functional assays. We have established a systematic method to dissect the ternary interaction between a protein of interest and the Cdk1–cyclin complex using the OyCD PCA. We used the Gateway cloning strategy to generate expression plasmids for prey–OyCD PCA fragment fusions. There are sufficient ORFs in Gateway-ready vectors available to screen a substantial portion of the yeast proteome. Interactions can be further dissected through mutagenesis of potential substrates to determine whether cyclins that bind to a prey, bind to overlapping or unique sites. The OyCD PCA strategy will better allow us to place individual

substrate phosphorylation events according to the timing of cyclin synthesis and degradation and therefore, when they occur during the mitotic cell cycle.

Our results provide insights into the mechanism of regulation of Cdk1 and its unique interaction partners. For instance, our results suggest that the Mft1–Cdk1 interaction depends on Clb6, consistent with its known function in mitotic recombination (Chavez et al., 2000). Cdk1 could interact with Clb6 and phosphorylate Mft1 to regulate its activity during the S phase of the cell cycle. Equally, as noted above, Rim20 does not have any Cdk1 consensus site but interacts with Cdk1 in complexes with G1, G1-S, and S-phase cyclins (Cln3, Cln1, and Clb5), perhaps through the cyclin binding motifs (RXL and LP) that it possesses. Because Rim20 is a regulator of Ime2, a protein kinase involved in activating meiosis, it would be worth investigating whether Rim20 could inhibit Cdk1 activity to stop the mitotic cell cycle when diploid cells are nitrogen-starved, driving them into meiosis (similar to how Far1 inhibits Cdk1 activity in the presence of α -factor for haploid MATa yeast strain) (Su and Mitchell, 1993; Tyers and Futcher, 1993).

Our identification of Cdk1–Clb3 as the specific form of Cdk1 interacting with γ -tubulin *in vivo* and *in vitro* provides critical temporal information and suggests that Tub4 phosphorylation occurs in S phase during the process of spindle assembly rather than during metaphase (Figure 3.4a). Phosphorylation of S360 was detected in the pool of γ -tubulin associated with SPBs (Keck et al., 2011) but not soluble pools (Lin et al., 2011) from which the SPB-bound fraction is recruited. Although accessible to solvent, S360 is positioned in the cleft between the two molecules of Tub4 of the γ -TUSC and projects from the interior surface of the γ -TURC (Kollman et al., 2011). Given its position inside the γ -TURC, it is unlikely that S360 would be accessible to Clb3–Cdk1 if the γ -TURC is occupied by a microtubule. Phosphorylation of S360 is, therefore, likely to occur in unoccupied γ -TURCs that result from the catastrophe of spindle microtubules, which is expected to occur as microtubules search for partners (e.g., bind to kinetochores or pair with microtubules projecting from the opposite pole) during early spindle assembly, at which time Clb3–Cdk1 activity is increasing.

The detection of a specific Clb3–Cdk1 interaction with Tub4 *in vivo* without any cell cycle arrest or perturbation reveals the exquisite sensitivity of the OyCD PCA for detecting cyclin–Cdk1 substrate complexes *in vivo*. The number of γ -tubulin molecules in a yeast cell is estimated to be \sim 7000 (Ghaemmaghami, 2003); however, the number associated with the spindle poles

(where S360 is phosphorylated) is expected to be ~600 [14 per γ -TURC and therefore, per microtubule (Kollman et al., 2010) and 40 spindle microtubules and 2–3 astral microtubules per cell (Winey and Bloom, 2012)]. However, the majority of the spindle microtubules are stabilized by their attachment to kinetochores or as a result of pairing to form polar microtubules (Gardner et al., 2008; Winey and Bloom, 2012). As a result, the conversion of an occupied γ -TURC to an open γ -TURC with S360 residues that can be phosphorylated is expected to be a rare event.

Finally, the OyCD PCA strategy can be generally applied to dissect other natural multisubunit enzyme complexes, including kinases, phosphates, and ubiquitin ligases *in vivo*, for which determining the basis of substrate specificity and localization is as difficult as for cyclin–Cdk1 substrates.

3.6 Materials and methods

3.6.1 Yeast strains

We used the MAT α BY4741 yeast strain along with FCY1 (encoding yeast cytosine deaminase) and all nine single cyclin deletion strains (*fcy1* Δ , *cln1* Δ , *cln2* Δ , *cln3* Δ , *clb1* Δ , *clb2* Δ , *clb3* Δ , *clb4* Δ , *clb5* Δ , and *clb6* Δ) (Giaever et al., 2002). The FCY1 gene was replaced by the nourseothricin resistance gene in all nine cyclin deletion strains by homologous recombination.

The Yeast ORF Collection of over 4900 plasmid-based yeast genes in Gateway expression clones was purchased from Open Biosystems (Gelperin et al., 2005).

3.6.2 Plasmid construction

The pAG413GAL1-ccdB-EGFP (HIS3 marker) Gateway destination vector (Alberti et al., 2007) was purchased from Addgene. The pAG413GAL1-ccdB-OyCD-F[1] Gateway destination vector was created by cloning an OyCD-F[1] sequence into the position of the EGFP gene in the pAG413GAL1-ccdB-EGFP destination vector using EcoRV and XhoI restriction sites. The p415Gal1-Linker-OyCD-F[2] was constructed by introducing Linker-OyCD-F[2] sequence in p415Gal1 (87330; ATCC) using BamHI and XhoI sites. The p415GAL1-Cdk1-OyCD-F[2] plasmid was obtained by cloning the *CDK1* gene upstream of OyCD-F[2] using SpeI and BamHI. The negative control p415GAL1-StartLinker-OyCD-F[2] was created by cloning a Linker-OyCD-

F[2], which has an ATG codon before the linker sequence in the position of Cdk1-OyCD-F[2] in p415GAL1-Cdk1-OyCD-F[2] using SpeI and XhoI restriction sites.

3.6.3 Gateway cloning

The selected genes from the Yeast ORF Collection were transferred into a Gateway donor vector to obtain ENTRY clones by Gateway BP reactions according to the manufacturer's protocol (Invitrogen), with the exception that the reaction were scaled down four times and the incubation time was prolonged to 16 h at 22 °C. We generated a Destination vector that carries the first fragment of OyCD (OyCD-F[1]) that we named pAG413GAL1-ccdB-OyCD-F[1]. This vector was used in an attL and attR LR reaction with an ENTRY clone that carries a gene encoding a protein of interest. The LR reactions were performed according to the manufacturer's protocol (Invitrogen), with the exception that the reactions were scaled down four times and the incubation time was prolonged to 16 h at 22 °C. The product of the LR reaction is an Expression clone that contains the gene of interest fused to OyCD-F[1] (pAG413GAL1-GeneX-OyCD-F[1]) and a byproduct plasmid that is not recovered. This strategy enabled us to easily create a large number of Gateway expression clones using the selected genes from the Yeast ORF Collection with each fused to the OyCD-F[1] sequence.

3.6.4 OyCD PCA to detect protein–protein interactions with Cdk1

The selected genes fused to the OyCD-F[1] sequence in Gateway expression clones were each separately transformed into BY4741 *fcy1Δ* yeast containing either p415GAL1-Cdk1-OyCD-F[2] or p415GAL1-Linker-OyCD-F[2]. After 3 d of growth, two colonies were picked from each transformation and grown in a 96-well plate in 400 µL synthetic complete medium without histidine and leucine and with 2 % (wt/vol) raffinose for 16 h. Galactose was added to each culture at a final concentration of 2 % (wt/vol) to induce the expression of the OyCD fusion proteins for 1 h at 30 °C before pinning the samples on selection plates with and without 1mg/mL 5-FC (Sigma) with a manual pintoil (1.58 mm, 1-mL slot pins, 45 mm, VP 408Sa; V&P Scientific Inc.). Raffinose, galactose, and glucose were purchase from Bioshop. Images of the plates were taken after 3 d of growth.

3.6.5 Detecting protein–protein interactions in the different cyclin deletion strains

Proteins that interacted with Cdk1 were screened in yeast strains expressing all nine cyclin genes or lacking one of nine cyclin genes. The potential substrate genes fused to the OyCD-F[1] sequence in Gateway expression vectors were cotransformed with p415GAL1-Cdk1-OyCD-F[2] into the *FCYI* deletion (*fcyIΔ*) strain and the nine single cyclin and *FCYIΔ* double deletion strains. Each test necessitated serial transformation of each strain. This screen is a labor-intensive process, although consisting of no more steps than an alternative approach, such as transforming MAT α and MAT α haploid strains with the bait and prey plasmids followed by mating. We consider the *FCYIΔ* yeast strain as the control strain in this screen and referred to it as the WT strain, because it expresses all nine cyclins. The controls for PPI detected in 10 yeast strains were the interactions between the homodimeric GCN4 leucine Zips fused to complementary OyCD fragments serving as positive controls and Cdk1 fused to OyCD-F[2] alone serving as negative controls. Four clones were picked and grown to saturation in synthetic complete medium lacking histidine and leucine with 2 % (wt/vol) glucose and 200 μ g/mL (wt/vol) G418 (Sigma). A glycerol stock was prepared with these cultures. All samples from the glycerol stock were printed on plates containing the same medium with 3 % (wt/vol) agar and allowed to grow for 4 d. For evaluating the OyCD PCA activity, colonies were pinned on synthetic complete medium lacking histidine and leucine with 2 % (wt/vol) raffinose, 2 % (wt/vol) galactose, 200 μ g/mL G418 (Sigma), and 3 % (wt/vol) agar plates with and without 1 mg/mL 5-FC using a robotically manipulated 384 pintoool (0.356-mm flat round-shaped pins, custom AFIX384FP8 BMP Multimek FP8N; V&P Scientific Inc.). Pictures were taken after 1, 2, 3, and 4 d of growth. This experiment was repeated three times.

3.6.6 Analysis of the cyclin deletion strains

Cell growth was quantified using ImageJ (Abramoff, 2004) by calculating the integrated intensity of each colony. This experiment was repeated three times. Results of one representative dataset, with images collected on day 4, are shown in Figure 3.3a. The activity of the OyCD PCA was measured by taking the ratio of integrated intensity of each colony grown on 1 mg/mL 5-FC over the integrated intensity of colonies grown in the absence of 5-FC. A two-tailed type 1 Student t

test was performed using the results of OyCD PCA activity obtained for the WT strain and each of nine single cyclin deletion strains. A minimum $P = 0.04$ was obtained for cells expressing the GCN4 leucine Zip:Zip in the *Cln3Δ* deletion strain. The minimum P value ($P = 0.02$) was obtained for cells expressing Cdk1 fused to OyCD-F[1] (Cdk1 alone) in the *Clb1Δ* deletion strain. Only P values ≤ 0.01 were considered as significant for the interaction between Cdk1 and test proteins in the different yeast strains. Results were represented in a matrix using iVici (Tarassov and Michnick, 2005).

3.6.7 Purification of γ -TUSC substrate and particle analysis

Constructs used for γ -TUSC expression are described in refs. (Kollman et al., 2010) and (Vinh et al., 2002). An N-terminal fragment of Spc110 fused to TEV-GST (Spc110N220-tev-GST) was used for complex purification. For purification of γ -TUSCs, Sf9 cells were coinjected with four virus stocks prepared as described in (Wasilko et al., 2009), and each virus encodes one of four components of the γ -TUSC (Tub4, Spc97, Spc98, and Spc110N220-GST). Infected Sf9 cells were incubated for 48 h, and γ -TUSC complexes formed spontaneously in the Sf9 cells. Cell lysis and all subsequent steps were performed in HB100 (40 mM K-hepes, pH 7.5, 100 mM KCL, 1 mM MgCl₂, 1 mM EGTA, 0.1 mM GTP; 1 mM DTT) as described in (Vinh et al., 2002). Whole-cell extracts containing γ -TUSC complexes were incubated with glutathione resin, the resin was washed, and the bound γ -TUSCs were eluted in HB100 by TEV cleavage. Purified complexes, which can be in the form of full or partial γ -TURCs and γ -TUSCs (Kollman et al., 2010), were diluted to 100 µg/mL and incubated on ice in HB100 buffer with KCL concentration raised to 500 mM to produce predominantly γ -TUSCs (Figure 3.4c). Negative staining was performed as described in (Kollman et al., 2010) using an FEI Tecnai G2 F20 200 kV Cryo-S/TEM Microscope.

3.6.8 Cdk1 kinase assays

Active Clb2–Cdk1 and Clb3–Cdk1 complexes were purified using the tandem affinity purification (TAP) method as described in (Koivomagi et al., 2011). Activity of the purified kinases was tested with histone H1 at a concentration of ~ 10 µM using 2 nM Clb2 or 5 nM Clb3–Cdk1 kinase. The composition of the assay was as follows. A prereaction mixture of kinase, 100 µM cold ATP, and 10 µCi γ^{32} P-ATP was prepared in kinase buffer as described in (Koivomagi et al., 2011). The

concentration of the purified γ -TUSC substrate was kept at 0.4 μ M in Clb2–Cdk1 (2 nM; a reaction) and Clb3–Cdk1 (5 nM; b reaction) kinase assays. Parallel a and b reactions were started by adding the kinase mixture to the γ -TUSC substrate, with incubation for 30 min at 30 °C. Reactions were terminated by addition of 0.1 volumes β -mercaptoethanol and 0.25 volumes 4 \times SDS sample buffer. Parallel reactions were separated on a 10 % SDS/PAGE gel, and proteins were visualized by Coomassie blue stain. Incorporation of P32 into the γ -TUSC components Spc98, Spc97, and Tub4 was measured using a TRIO Typhoon phosphoimager.

3.7 Note added in proof

Prior to publication, Nazarova et al. described the significance of Clb3-Cdk1 phosphorylation of γ -tubulin (Nazarova et al., 2013).

3.8 Acknowledgements

We thank Emmanuel Levy for the ImageJ macro used to quantify the growth of the colonies (cyclin contingency experiments) and Gary Brouhard, Justin Kollman, Susanne Bechsted, and members of the laboratories of JV and SWM for many helpful discussions. EM was performed at the Facility for Electron Microscopy Research at McGill University. The authors acknowledge support from Canadian Institutes of Health Research Grant MOP- 123335 and Natural Sciences and Engineering Research Council (Canada) Grants RGPIN 262246 (to JV), MOP-GMX-152556 (to SWM), and MOP-GMX-231013 (to SWM). PHE thanks the Faculté des Études Supérieures for scholarships. CH is supported by a Cellular Dynamics of Macromolecular Complexes-Collaborative Research and Training Experience Graduate Fellowship.

Chapter 4 Conclusions and future work

4.1 Conclusions

While proteins present the main workhorses of living cells involved in almost all biological processes, protein PDIs represent the essential means by which living cells regulate their biochemical machineries. Aberrant regulation of KPs plays a direct causal role in a wide range of diseases related to neurological, skeletal, vascular, immunological, metabolic and cellular growth rate disorders (Hendriks et al., 2013; Lahiry et al., 2010). Implication of KPs in fatal diseases returns to their profound effects in regulating a wide range of biological processes, making from these enzymes, from their regulatory mechanisms and their signaling pathways a subject for intense investigations in both prokaryotic and eukaryotic cells. Over the past decade, many studies attempted to identify KPs substrates and to elucidate the pathways in which they are implicated. With the advent of HTP phosphoproteomic techniques, databases became overwhelmed with the sheer size of data annotating KPs potential substrates as well as their phosphosite profiles in different conditions. Despite this wealthy data, it is now well established that to gain insights into how KP-Nets function, we need to move from determining functions of individual proteins to understanding interactions between these proteins on the proteome scale, we need also to understand how each of these enzymes and their substrates are assembled to form the structure of the intricate KP-Net and ultimately to apprehend how the underlying structure of this network is rewired to respond appropriately to different changes in environmental conditions resulting in various complex system behaviors. In summary, we need to move from protein individual analysis to system collective analysis. This thesis addressed some of these questions.

We assembled the largest bona fide KP-Net known to date in the budding yeast. The importance of this network returns on the one hand, to the high quality of its PDIs and on the other hand, to including phosphatases in it. Most of the previously assembled KP-Nets do not include phosphatases and if they do, their PDIs are mostly indirect (Table A1.II). To ensure that the assembled network includes high confidence data, interactions were filtered according to two criteria: first, interactions should have a KID score ≥ 4.52 (corresponding to a $P \leq 0.05$) and second, interactions should be validated by at least one HTP or LTP experiment showing the

occurrence of a biochemical PDI [HTP protein phosphorylation chip (*in vitro*), HTP solution kinase assay using analogue sensitive kinase alleles (*in vitro*), LTP protein KP assays (*in vitro*), LTP phosphorylation site mapping by mass spectrometry or using phospho-specific antibody in western blot or site directed mutagenesis (*in vitro* or *in vivo*)]. Obviously, the second selection criterion does not guarantee that the assembled PDIs are direct. But, including a KID score cutoff backs-up the assembled interactions by indirect data, such as detecting PPIs, genetic interactions, co-localization, chemical co-fitness relationship or co-expression of the enzyme and its substrates as well as observing a dependency of the substrate transcript/protein abundance or subcellular localization on the KP activity. Therefore, we think that the KP-Net of the budding yeast assembled in this thesis is, to the best of our knowledge, the first and largest assembled network containing both bona fide PDIs, proposing the KP-Net as a reliable gold standard for subsequent phosphoproteomic studies. Unfortunately, the unavailability of such high quality PDIs caused studies of the KP-Net to lag way behind that of TF-Nets. Interestingly, the KP-Net assembled in this work could be updated to include more recent data and could be extended by the integration of other biological resources, unleashing KP-Net studies in the budding yeast.

The assembled KP-Net was found to be non-randomly organized with a scale-free topology as in (Zhu et al., 2007) with a degree exponent $\gamma = 2.58$, Figure A1.1). As previously mentioned, the hierarchical structure is not an alien property of real-world networks including biological networks. The hierarchical level of the KP-Net was assessed using the GRC metric to be equal to 0.61, suggesting a moderate hierarchical amount for the KP-Net. This result coincides with the findings of Trusina et al. stating that the hierarchical level of a network having a scale-free topology shows negative correlation with the network degree exponent, γ (Trusina et al., 2004). This degree exponent ranges from a maximal hierarchy when γ approaches 2 to a minimum when γ approaches 3. Accordingly, with a degree exponent of 2.58, the KP-Net is expected to have a moderate hierarchical level; an expectation that concords also with the corporate hierarchy described by Bhardwaj et al. (Bhardwaj et al., 2010b).

Bhardwaj et al. described hierarchical structures of yeast and human KP-Nets to work in an intermediate situation between two extremes encountered in: (1), autocracies, such as military hierarchies characterized by well-defined levels and a clear cascade of commands; and (2), democracies, such as in a scientific collaboration networks lacking a clear definition of levels and enriched in collaborative efforts to guide the discussions (Bhardwaj et al., 2010b). The

authors compared this intermediate hierarchy to a corporate organization in which middle managers collaboratively work and interact extensively with other employees to ensure an efficient management of the organization (Floyd and Wooldridge, 1997; Pappas, 2001). The authors extrapolated the intermediate hierarchical structure of corporate organizations on the studied networks in which regulators belonging to the middle layer are the most among other layers to control their substrates in collaboration with other regulators. Interestingly, our findings are consistent with and support these observations in different ways. First, the SCC making up the KP-Net core layer represents a structure that highly promotes the collaboration between core layer KPs in controlling the network (Paragraph 2.4.2). Second, the higher in-degree, out-degree and the higher number of subcellular localizations of KPs in the core layer and the enrichment of this layer for KPs representing bottlenecks, pathway-shared components and essential genes in comparison with KPs in top and bottom layers reveal the critical role that the core layer KPs play in managing and more importantly in coordinating the information hovering the network from top to bottom (Figure 2.3 and Figure 2.4).

On the other hand, autocratic layouts are characterized to have a strict and rigid structure in which commands go exclusively through midlevel regulators generating bottlenecks in the middle level of the network. In contrast to autocratic hierarchies, corporate layouts are characterized to have more flexible structures in which shortcuts between upper regulators and bottom targets avoid going through regulators in the middle level. This alleviates the high level of stress occurring in midlevel regulators, a result that is concordant with the shortcuts observed in the KP-net, where 4.5 % of KP-Net interactions are between the top and the bottom layers and 30 % of KP-Net interactions are between the top layer and the layer containing substrates of KPs that are not KPs, without going through the core layer (Figure 2.2b).

Because of its outperformance over other algorithms, the VS algorithm (Jothi et al., 2009) was applied on the KP-Net in order to elucidate its hierarchical structure (Figure 2.2). In addition, the VS algorithm is the only network decomposition method that does not force a node to be classified in one level, but allows nodes to span among many levels. This feature is important in order to provide a global solution, that encapsulates many solutions having less nodes that span more than one level. We are not sure that forcing nodes in one level is feasible from a biological point of view, due to the considerable number of uncovered phosphorylation interactions till today. Another feature of the VS algorithm is that it classifies all nodes of a

network in different levels in a way to prevent upward arrows. However, it is important to note that the representation of the hierarchical structure of the KP-Net as a hierarchy that is devoid of upward arrows does not completely represent all events that occur in living cells. This could be better understood by imagining, for instance, a TF belonging to the zero layer of the KP-Net that regulates the expression of genes encoding KPs in any layer of the hierarchy. The zero layer is the lowest layer in the hierarchy and is composed of substrates of KPs or of KPs that have no known substrates. Such examples add upward arrows to the real KP-Nets and these upward arrows form FBLs that are not represented in the KP-Net. FBLs emerging from the exchanged effects of KP-Nets and TF-Nets on each other are not represented for two reasons: first, arrows in the KP-Net represent PDIs and second, this study focuses on studying KP-Nets exclusively.

Given that KPs in top and core layers regulate those in the bottom layer, the backbone of the KP-Net could appear to resemble a FFL (Figure 2.2b). However, this observation represents an artifact generated by the application of the VS algorithm. Actually, the way the VS algorithm classifies network nodes in three layers could rarely generate a linear structure, unless no KPs of the top layer regulate any of the KPs of the bottom layer ($P < 10^{-4}$, DPR, Paragraph 2.5.2.1).

The KP-Net was rather found to have a bow tie structure which describes not only the small number of KPs in the core layer in comparison to that of KPs in top and bottom layers, but also the large number of inputs received by and of outputs generated by the core layer in comparison to those of the other layers (Figure 2.3a,b). A bow tie organizational structure permits thus the KP-Net to deal with a wide variety of inputs that might be received simultaneously by its strongly connected core layer, where these inputs could be processed and extensively exchanged among core layer KPs in order to make a decision and choose which output or outputs among the many outputs should be generated as a response or a set of harmonized cellular responses. With this concept in mind, a bow tie shape sustains the dynamic nature of the KP-Net.

Inspired by Jothi et al. (Jothi et al., 2009), after elucidating the hierarchical structure of the KP-Net, we assessed topological properties for KPs in each of its layers and overlaid the widest range of biological properties on top of the KP-Net hierarchy. As in Jothi et al., the topological and biological properties of the network controllers, KPs in our case and TFs in Jothi et al case, are not equally distributed among the three layers of the network hierarchical

structure. Importantly, the distribution of TF properties identified among the three layers of a TF-Net hierarchical structure in the budding yeast (Jothi et al., 2009) differs from the distribution of KP properties found among the three layers of the KP-Net assembled in this work (Abd-Rabbo and Michnick, 2017). Discussing these distinctions in light of the particular biological nature of each network might uncover some potential strategies employed by each of the networks to achieve their functions.

Although biological networks could have similar architectural structure, similar topological properties and more strikingly similar general biological functions, they could use different mechanisms to achieve their principal goals. For instance, both the TF-Nets and KP-Nets share the same organizational structure and global topological properties (Abd-Rabbo and Michnick, 2017; Jothi et al., 2009), as they are scale-free and characterized by the small-world property, a non-linear hierarchical structure and are enriched for almost same logic motifs (SIM, FFLs, FBLs, bi-fans and diamonds) (Alon, 2007; Jothi et al., 2009; Zhu et al., 2007). Moreover, both networks have the same principal biological functions consisting of information processing and cell regulation. TFs regulate genes expression depending on information received by either a kinase or another signaling molecule related to the extracellular environment via a signaling transduction network. As for KPs, they regulate protein activity, localization and interaction with other proteins depending on information received from signaling molecules such as transmembrane or cytoplasmic receptors. Despite all the mentioned similarities between TF-Nets and KP-Nets, these two networks differ considerably when comparing the distribution of some of their biological properties along their network hierarchical structure, an observation that could be tightly related to their different natures.

How does the nature of the TF-Net differ from that of the KP-Net? In fact, since proteins represent the main workhorses that carry out most cellular biological processes, usually a living cell will primarily depend on the TF-Net either to produce more mRNA copies that will be subsequently translated into proteins to perform certain tasks when needed or to repress the expression of certain genes when their tasks are not required. Definitely, this system is relatively slow to reach its goals and is not reversible, unless the produced proteins stop functioning by targeting them for degradation. On the other hand, a living cell opts for the KP-Net as a much faster alternative to regulate biological processes when needed. This system is faster than the TF-Net, because of the rapid kinetics nature of a phosphorylation interaction compared with those of

a transcription event; it is also reversible, because each phosphorylation interaction could be reversed by a dephosphorylation interaction and *vice versa*.

Differences in the nature of these two networks could probably result in distinct distributions of biological properties of networks constituents along each network structure. Interestingly, proteins in the top layer of TF-Nets and KP-Nets were found to be more abundant than those in the bottom layer. Moreover, TF essentiality was found to be positively correlated with the position of TFs in the TF-Net hierarchy, whereas essential KPs were found enriched in the core layer of the KP-Net. Furthermore, proteins in the top layer were found to be noisier than those in the bottom layer of the TF-Net. In contrast, KPs in the top layer were found less noisy than those in core and bottom layers. The observed distributions of TF protein abundance and noise among the three layers of the TF-Net hierarchical structure were associated with enhancing the capacity of living cells to survive under different conditions (Jothi et al., 2009). The distributions of KP protein abundance and noise in the KP-Net hierarchy might also serve the same goal. Trying to interpret these results in the context of the way each network functions could help uncovering biological principles employed by these networks in cellular response.

Since the TF-Net is irreversible and relatively slow to achieve its goals, a wrong cellular decision could not only cost the living cell excessive unnecessary charges (cellular engagement into the long process of genes and then proteins expression), but could sometimes lead to cellular death. Importantly, the noisy protein expression of key TFs was shown to allow a fraction of clonal cell populations to survive or die upon cancer treatment (Cohen et al., 2008; Spencer et al., 2009). Therefore, high protein noise at the top layer of the TF-Net was suggested to allow certain key TFs to utilize the same underlying network structure differently when exposed to various stimuli. Actually, a minimum expression level of a given TF will be required to activate a cellular pathway. Hence, only the proportion of cells expressing a TF required to activate a vital pathway under a given condition would respond effectively to a given stimulus and the remaining population proportion would just die under the latter stimulus. Yet, the situation could be reversed under another condition. Cells under-expressing this TF could favor activation of another pathway representing an effective response to stimulus 2 (Jothi et al., 2009). For instance, the fluctuating level of expression of TFs would permit a big proportion of cells clonal population to continue to proliferate under normal conditions and only a small fraction of this population to survive under severe conditions (e.g. the bacterial resistance to certain antibiotics).

Finally, the relatively high proportion of essential TFs in the top layer of the TF-Net might reflect the occurrence of relatively a bigger number of key TFs dictating cell death and survival in this layer.

On the other hand, the fast nature of events mediated by the KP-Net could privilege KPs occurring at the top layer of the network to be characterized by two features: a relatively high protein abundance and minimal protein noise. This arrangement could be ideal to guarantee that signaling molecules will find KPs available at invariable relatively higher abundances in the input layer compared with the output layer of the network and will activate them, triggering appropriate signaling cascades in response to different environmental cues. Observing a higher abundance of KP proteins in the top layer compared to those in the bottom layer might lead to conclude that upper layer KPs are not saturated by their substrates and could not achieve ultrasensitivity or switch like behaviors (Ferrell, 1996). However, protein abundance data used in this work are not sufficient to deduce such conclusions, as these data represent an amalgamation of protein abundances produced by different studies that were accomplished under various conditions. Consequently, further investigations will be needed to examine whether each of the KPs belonging to the top layer are saturated or not by their substrates. In case our observations are supported by evidence, top layer KPs might not exhibit a switch-like behavior. This represents a plausible scenario which concords with our findings. A switch-like behavior might not be required at the top layer representing the input level of the KP-Net. The top layer might instead be important to receive inputs and transmit them to KPs involved in decisions making and strongly interacting with each other to ensure processing of signals before any decision could be taken. This is supported by our findings showing that KPs in the top layer are enriched in signaling regulation and response to stimulus, that those at the core layer are enriched for cell cycle and decision-making and more importantly that the number of phosphosites predicted to act as molecular switches was found to be higher in KPs in core and bottom layers than in those in the top layer (Figure 2.2a, Figure 2.4d and Table A1.IV). On another hand, a high abundance of KP proteins might be necessary at the top layer for two reasons: First, the observed degeneracy of many KPs in the top layer (e.g. PKAs, Tel1-Mec1 and calcineurins, (Figure 2.1a) might suggest that KPs having overlapping functions could substitute each other in certain conditions; therefore, a high number of copies of a KP permits them to act on their substrates and the substrates of the other unavailable enzymes with whom they share some functions. Second,

since KPs play a pivotal role in signaling pathways and since they are upstream of TFs in these pathways, the KP-Net could not afford the risk to dismiss a signal due to the unavailability of the enzyme responsible for launching a signaling pathway in a particular condition. At the same time, the KP-Net could not afford to be affected by noisy expression of its components at the top layer for relaying information. Finally, the enrichment of the KP-Net core layer for KPs representing essential genes might highlight the important and critical role core layer KPs play in cell fate decisions. Therefore, the low protein noise accompanied with the relatively high protein abundance in the top layer and the over-representation of essential KPs occurring in the core layer of the KP-Net might make an appropriate layout for the KP-Net to ensure a robust signaling and consequently to enhancing the capacity of living cells to survive under different conditions

Since results and suggestions reported in this thesis depends basically on the application of the VS algorithm, it was extremely important to make sure that these results were not biased. To this end, parameters affecting VS algorithm results were unprecedentedly identified. Our investigations showed that both node in- and out-degrees play an important role in classifying nodes across the network hierarchy (Paragraph 2.4.11). Although this finding might describe the VS algorithm as a biased algorithm, distribution of biological properties across the three layers of the network were shown to be unbiased, consequently unaffected by KPs degree distribution among these layers (Figure 2.6, black line). Moreover, this bias, might also be equally encountered in most of other decomposition algorithms (Bhardwaj et al., 2010a; Gulsoy et al., 2012; Hartsperger et al., 2010; Ma et al., 2004), because it coincides perfectly with hierarchical structure properties, in which nodes having a zero in-degree are automatically classified in the top of the hierarchy and those having a zero out-degree in the bottom of the hierarchy. In contrast, no correlation was found between most of the biological properties that has been studied in this thesis and in- and out-degrees of nodes respectively (Figure A1.5-8). Only some biological properties of KPs were found to be moderately correlated with either KP in- or out-degrees. The latter observations might be tightly related to the ideas introduced previously in (Paragraph 1.5.1) suggesting that nodes topological properties encapsulate a limited amount of information about nodes biological functions, but together nodes biological and topological properties could better define their functions.

Considering the outperformance of the VS algorithm over other algorithms and the valuable insights such network analysis reveal, an R package called the VertexSort was developed and made available (Supplementary methods in) for the scientific community to perform similar studies using the VS algorithm on other networks having the command-execution organizational aspect. This package also includes the five randomization algorithms used in (Abd-Rabbo and Michnick, 2017). One of these algorithms, the SDPR, was developed in this work to generate random networks having similar in- and out-degree distributions to those of the network in question. SDPR uses a variant of the matching algorithm (Milo et al., 2004). Randomization functions developed in the VertexSort use the Snowfall package (Knaus et al., 2009) to benefit from the parallel programming technology. This technology permits to run a code on multiple computer cores simultaneously, which decreases the execution time significantly. This feature permits the generation of 1000 random networks having the same size of the KP-Net (616 nodes and 1087 edges) in 1 minute 34 seconds instead of 7 minutes and 33 seconds on a 3.2GHz Intel Core i7 using 11 processors.

After presenting advantages of this research project and its impact on the field of biological network analysis in particular on regulatory and signaling networks, the adopted project design imposed certain limitations, such as: missing PDIs, missing biological properties (the protein noise of a considerable number of KPs was missing), and bias in the VS algorithm. In fact, characterizing the KP-Net properties, representing one of the goals of this thesis, imposed on the KP-Net, to be assembled from high quality PDIs, leading to the exclusion of low confidence PDIs from the KP-Net and to the inclusion of high quality PDIs biased towards choosing interactions implicating KPs that caught the interest of the scientific community (e.g. Cdc28, Cdc14, Hog1 and TPKs). In contrast, including low quality PDIs could have introduced much more unwanted indirect PDIs than genuine direct PDIs in the KP-Net, leading to meaningless interpretations. Furthermore, including only bona fide PDIs in the KP-Net permitted us to describe a network that reflects our current knowledge in the field, meeting perfectly the goals of this project. Moreover, missing biological properties of KPs return to technical limitations of the chosen experiments generating the biological properties that were integrated within the KP-Net. To the best of our knowledge, however, those experiments provided the most complete and accurate data on the time this study was accomplished. Indeed, improving the accuracy of the current proteomic experimental techniques will make more data

available, covering a larger proportion of the proteome. However, despite the latter drawback, characteristics of the KP-Net revealed in this study are not expected to be very different from real-world KP-Nets for the significant overlap between the three layers of the KP-Net and the three layers of the noisy KP-Nets after adding noise (Figure 2.7b et d). Finally, although classifying nodes within a network using the VS algorithm was shown to be affected by node degree distribution, biological properties were also shown to be unaffected by this bias (Figure 2.6, black line).

As stated in the introduction, functional redundancy and compensatory behaviors of certain KPs represent one of the factors contributing to the complexity of KP-Nets. Such behaviors are not exhibited only by KPs, but also by some of KP regulatory subunits, making thus the decipherment of KP-Net complexity even more difficult. For example, functional redundancy of Cdc28 cyclins was reported by various studies in early 1990s. The approach of these studies consisted of characterizing and comparing the phenotypes of budding yeast cells containing null mutation of one or more cyclins (Fitch et al., 1992; Richardson et al., 1992; Schwob and Nasmyth, 1993). Very few approaches were interested in identifying the specificity of cyclins towards Cdc28 substrates *in vitro* (Koivomagi et al., 2011; Loog and Morgan, 2005; Ubersax and Ferrell, 2007). Theoretically, cyclins showing specificity to the same set of Cdc28-substrates could exhibit compensatory behaviors of each other towards their shared substrates. Yet, there were no *in vivo* experimental methods differing from the phenotype based approaches to confirm cyclins functional redundancy at large scale. Moreover, there were no *in vivo* HTP assay to identify which cyclin or cyclins are involved in interactions between Cdc28 and its substrates.

Our OyCD PCA screen employed to identify Cdc28 substrates and cyclin dependency confirmed that cyclins exhibit wide compensatory behaviors on knocking out one of them (Ear et al., 2013). Such compensatory behaviors can be deduced from observing that many of the proteins (Bud6, Cct7, Gsy2, Kar9, Lte1, Mcm3, Pah1, Pho4, Ptk2, Rim20, Swi5, Thp1 and Ugp1, Figure 3.3b), that were assayed to interact with Cdc28 in our screen, show dependency on more than one cyclin belonging to the same cyclin class (G1/S class: Cln1-3; S class: Clb5-6 and G2/M class: Clb1-4) and even on more than one cyclin belonging to a different class (Figure 3.3b). For instance, while the chaperonin subunit Cct7 was found to be phosphorylated by

Cdc28-Cln2 in (Ptacek et al., 2005), it was identified in our OyCD PCA as a potential substrate of Cdc28 complexed with most of the cyclins, but particularly with G1/S class cyclins that are well characterized to fulfil redundant functions.

The dependency of Cdc28-substrate interactions on different phase specific cyclins suggests a temporal organization of phosphorylation events during cell cycle progression. For instance, our screen showed that Clb2, Clb4 and Clb5 are implicated in the phosphorylation of Kar9, a microtubule-associated protein, by Cdc28 (Figure 3.3b). These findings are supported by literature evidence, where it has been shown that Cdc28-Clb5 phosphorylates Ser 496 of Kar9 to prevent Kar9 from loading to the mother bound pole, an important event for asymmetric position of Kar9 to the spindle pole and to the microtubules (Moore and Miller, 2007). It has also been shown that Cdc28-Clb4 phosphorylates Ser 196 of Kar9 in order to direct cytoplasmic mitotic microtubules to the bud (Liakopoulos et al., 2003; Moore and Miller, 2007). Phosphorylation of these two residues is needed for proper alignment of spindle along the mother-bud axis in pre-anaphase (Liakopoulos et al., 2003; Maekawa et al., 2003; Moore and Miller, 2007). Although, there is no evidence of the implication of Cdc28-Clb2 in phosphorylating Kar9, many observations suggested the existence of at least another phosphosite on Kar9, making this implication possible (Maekawa and Schiebel, 2004; Moore et al., 2006; Moore and Miller, 2007).

Finally, very few potential substrates of Cdc28 (Far1, Rad9, Stb1 and Tub4) showed dependency on only one cyclin in the OyCD PCA screen (Figure 3.3b). Because of the cyclic expression of cyclins, the latter finding could shed light on the time during which Cdc28 phosphorylates its substrate in the cell cycle, and sometimes on the circumstances under which this event takes place. For instance, the phosphorylation of γ -tubulin Tub4 by Cdc28 was found exclusively dependent on Clb3 *in vivo* (Figure 3.3b) and *in vitro* (Figure 3.4g). Tub4 in the form of a dimer with two other proteins Spc97 and Spc98 form a Y shaped complex, the γ -tubulin complex (γ -TUSC), which is responsible for microtubule nucleation (Kollman et al., 2011). Seven γ -TUSCs complexes assemble in the form of a ring to make the gamma-tubulin ring (γ -TURC) serving as a template for microtubule nucleation (Kollman et al., 2010). The three-dimensional conformation of the γ -TURC showed that Tub4 can be phosphorylated from the interior surface of the ring. Hence, Tub4 cannot be accessed and phosphorylated by the Cdc28-

Clb3 complex unless the γ -TURC is unoccupied by a microtubule. On another hand, dependency of the Cdc28-Tub4 interaction on Clb3 suggests that phosphorylation of Tub4 by Cdc28 occurs between the mid and the end of the S phase when the expression of Clb3 begins to peak. During the S phase, the γ -TURC is theoretically occupied by microtubules that are seeking a kinetochore to bind to, unless one of these microtubules experience a catastrophe that causes its complete denucleation. Taken together, this suggests that Tub4 is phosphorylated in S phase on a microtubule catastrophe before the microtubules are connected with any kinetochore. In summary, a large scale OyCD PCA to identify Cdc28 substrates and cyclin dependency could provide critical temporal information on cell cycle events, bringing a better understanding of cell cycle progression.

4.2 Future directions

Among the most challenging obstacles that hinder the study and the reverse engineering of KP-Nets is the association of each phosphorylated site or protein to the KPs responsible for catalyzing the PDI. The recent revolutionary advances in HTP technologies enhanced exponentially our knowledge about phosphorylation networks, expanding the coverage of phosphoproteins and suggesting a list of potential KPs responsible for acting directly or indirectly on these phosphoproteins. Although such experiments are by themselves insufficient to elucidate KP-Nets, this trend is expected to continue in the future. Importantly, bioinformatics tools based on data integration could be of considerable help in this area. Among the interesting integrative computation approaches that help in discovering kinases-substrates relationships, Networkin combined specifically human data from various experimental resources (consensus recognition motifs, PPIs, coexpression data, colocalization data, literature, and genomic co-occurrence data) and had successfully identified the kinases responsible for phosphorylating 53BP1 and Rad50 in the DNA damage pathway in human (Linding et al., 2007). Various computational techniques having a similar objective were developed for yeast; however, they did not recognize a similar success. The major drawback of these tools is that they do not benefit from the wealth of biological data generated from different HTP experiments and use only some of these data such as protein structural information and protein linear sequence information to predict kinases phosphosites, permitting consequently to predict kinases substrates (Trost and Kusalik, 2011). PhosphoChain is one of the rare and interesting tools that integrate biological

data (mRNA and phosphorylation profiles on KPs deletion with phosphorylation motifs) to predict condition-specific yeast KP-Nets (Chen et al., 2013). But, since mRNA profiles is not the ideal indicator of protein activity, an interesting enhancement of PhosphoChain could be to use protein expression instead of mRNA expression. Another important strategy also combining different multi-omics data introduced a score called HeRS that permitted to successfully unravel many hetero-regulatory modules made of KPs, TFs and their substrates in MAPK pathways. Tools aiming at predicting phosphorylation sites, KP substrates, and KP-Nets will continue to be developed based on integrating different dynamic biological data in the next years.

As phosphorylation interactions play a crucial role in regulating cellular behaviors, identifying the consequences of phosphorylation interactions could enhance enormously our understanding about how cells behave in response to changes in its surrounding environment. An extremely overwhelming but difficult goal to reach is to identify the biological consequence of each functional phosphorylation event on substrates subcellular localization, stability, activity and interaction with other proteins. Mutation of these phosphosites could bring a better understanding about their functions; however, the irreproducibility of the observed changes to a protein caused by phosphorylation makes the interpretation of these mutations very difficult. A step further that could help in tackling these questions consists of performing HTP experiments that simultaneously quantify changes in protein phosphorylation profiles and protein subcellular localization, protein phosphorylation profiles and protein abundance or changes in protein phosphorylation profiles and PPIs on the knockdown of KPs one at a time. Such studies will allow the association of the deleted KP and hence the phosphosites on the latter KP to a set of events and consequently the association of this set of events to the pathway in which the knocked down KP is implicated.

Most of the efforts dedicated for studying KP-Nets used data generated from experiments in which cells were exposed to one environmental condition, resulting in static maps. To explore the dynamic nature of such biological networks, an applied common strategy was to integrate biological data reflecting the dynamic part of these networks with the networks mapped under normal conditions. To the best of our knowledge this approach was not applied on the yeast phosphorylome as it was on the yeast PPI network (de Lichtenberg et al., 2005) or as on the human phosphorylome (Olsen et al., 2010). Our study (Abd-Rabbo and Michnick, 2017) and Cheng and colleagues study (Cheng et al., 2015) represent the first attempt to integrate static

biological data with the KP-Net and the K-Net in the budding yeast, respectively. It will be interesting to overlay biological data (e.g. mRNA and protein expression) measured in different conditions or time points on the KP-Net assembled in this thesis to investigate whether the KP-Net will maintain the distribution of its biological properties during its dynamic behavior. In case the KP-Net is characterized by different biological properties, it will be important to examine how the KP-Net biological properties will be distributed along its hierarchical structure and whether this new distribution unveils new insights on the regulatory mechanisms shaping out KP-Net responses.

In the last decade, many studies have been devoted to examine the KP-Net under different conditions, such as cells exposed to alpha factor, DNA damage agents, rapamycin and hyperosmotic stress or cells deficient in a KP (Chen et al., 2010; Gruhler et al., 2005; Iesmantavicius et al., 2014; Li et al., 2007; Smolka et al., 2007; Soufi et al., 2009). The latter studies enhanced our understanding about the pathways implicated in different stress conditions. But, much more research is needed to be achieved to identify the dynamic parts of the KP-Net, to discover how this network is rewired to adapt to the new surrounding environmental conditions and to uncover the sequence of the PDIs and that of the conditional ones. It might be interesting to integrate already existing phosphoproteomic data in order to get the most from the immense body of data that has previously been generated during the last decade. For instance, we think that integrating a valuable resource such as data generated from the Bodenmiller study (Bodenmiller et al., 2010) with other experimental data and pathway literature knowledge could uncover valuable findings about dynamic parts of the phosphorylome and might associate new phosphoproteins to existing pathways. Furthermore, the current graph framework used to analyze phosphorylation networks represents a protein by a node and a PDI by an edge. This representation need to be enhanced to permit the study of various phosphosites belonging to one protein and to take into account the conditional events that depend on each other on large scale.

There is emerging evidence of the interplay between various PTMs (Venne et al., 2014). For instance, it has been shown that phosphorylation interactions could represent signal for protein-ubiquitin ligases E3 to target a protein for ubiquitination in a phospho-dependent manner under various circumstances (e.g. on exit of mitosis, in response to DNA damage on double-stranded DNA breaks, and in response to nutrient deficiency) (Iesmantavicius et al., 2014; Polo and Jackson, 2011; Watanabe et al., 2004). Other evidence showed that methylation and

acetylation could regulate phosphorylation (Dihazi et al., 2005; Yun and Fu, 2000). The advances of LC-MS/MS techniques and posttranslational modified peptides enrichment strategies will allow the identification of more than 450 posttranslational remaining modifications currently annotated in Uniprot on large scale in the future decade (Magrane and Consortium, 2011). With the availability of ubiquitination (Venancio et al., 2009), methylation (Pang et al., 2010) and acetylation (Henriksen et al., 2012; Pang et al., 2010) on large scale more efforts will be oriented towards characterizing the co-occurrence and influence of each of these posttranslational events on phosphorylation on large scale to decipher their biological roles and to examine whether they target a specific class of proteins, in particular phases of the cell cycle, in specific subcellular localizations or under various environmental conditions.

Jia et al. has recently addressed an important question related to the nature of controllability in complex systems (Jia et al., 2013). The authors showed that real networks could adopt a centralized or a distributed control mode. Centralized systems are characterized by a high level of security and enhanced efficiency suitable for task executions, whereas distributed systems are characterized by requiring a large number of control configurations suitable for yielding innovation. Importantly, the authors suggested that these networks could also switch from one controllability mode to another through few structural perturbations such as switching the direction of some edges, making from this switch a modification or an adaptation of their functional strategies. The authors classified the nodes into critical, redundant and intermittent nodes. Critical or driver nodes are those needed to control the dynamics of a network and this by continuously controlling these nodes in order to determine their state. In contrast, redundant nodes are never required for the control of the system. Finally, intermittent nodes could belong to either of the two classes depending on the system condition. Determining the controllability mode of the KP-Net, investigating whether the KP-Net could switch its control mode and on which cost, classifying its nodes into critical, redundant and intermittent nodes categories and characterizing the topological and biological properties of these classes could provide valuable information about the principles used by the KP-Net in its dynamic rewiring to response to different stimuli.

Finally, an important continuation of research performed in this thesis could be to apply the VS algorithm on KP-Nets and TF-Nets in human and in other organisms and to compare

distribution of properties of each network with those seen in yeast in order to investigate whether the KP-Net preserves the same properties across species.

References

- Abd-Rabbo, D., and Michnick, S.W. (2017). Delineating functional principles of the bow-tie structure of a kinase-phosphatase network in the budding yeast. *BMC bioinformatics* *11*, 14.
- Abramoff, M.D., Magelhaes, P.J., Ram, S.J. (2004). Image Processing with ImageJ. *Biophotonics International* *11*, 36-42.
- Ahn, J.H., McAvoy, T., Rakhilin, S.V., Nishi, A., Greengard, P., and Laird, A.C. (2007). Protein kinase A activates protein phosphatase 2A by phosphorylation of the B56delta subunit. *Proceedings of the National Academy of Sciences of the United States of America* *104*, 2979-2984.
- Albert, R. (2005). Scale-free networks in cell biology. *Journal of cell science* *118*, 4947-4957.
- Albert, R., Jeong, H., and Barabasi, A.L. (2000). Error and attack tolerance of complex networks. *Nature* *406*, 378-382.
- Alberti, S., Gitler, A.D., and Lindquist, S. (2007). A suite of Gateway cloning vectors for high-throughput genetic analysis in *Saccharomyces cerevisiae*. *Yeast* *24*, 913-919.
- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002). Signaling through G-Protein-Linked Cell-Surface Receptors. In *Molecular biology of the cell* (New York: Garland Science).
- Alon, U. (2007). Network motifs: theory and experimental approaches. *Nat Rev Genet* *8*, 450-461.
- Aragues, R., Sander, C., and Oliva, B. (2008). Predicting cancer involvement of genes from heterogeneous data. *BMC bioinformatics* *9*, 172.
- Arava, Y., Wang, Y., Storey, J.D., Liu, C.L., Brown, P.O., and Herschlag, D. (2003). Genome-wide analysis of mRNA translation profiles in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America* *100*, 3889-3894.
- Archambault, V., Chang, E.J., Drapkin, B.J., Cross, F.R., Chait, B.T., and Rout, M.P. (2004). Targeted proteomic study of the cyclin-Cdk module. *Molecular cell* *14*, 699-711.
- Ba, A.N., and Moses, A.M. (2010). Evolution of characterized phosphorylation sites in budding yeast. *Molecular biology and evolution* *27*, 2027-2037.
- Bader, G.D., Betel, D., and Hogue, C.W. (2003). BIND: the Biomolecular Interaction Network Database. *Nucleic Acids Res* *31*, 248-250.
- Barabasi, A.L., and Albert, R. (1999). Emergence of scaling in random networks. *Science* *286*, 509-512.
- Barabasi, A.L., and Oltvai, Z.N. (2004). Network biology: understanding the cell's functional organization. *Nat Rev Genet* *5*, 101-113.
- Bardin, A.J., Visintin, R., and Amon, A. (2000). A mechanism for coupling exit from mitosis to partitioning of the nucleus. *Cell* *102*, 21-31.
- Barford, D., Das, A.K., and Egloff, M.P. (1998). The structure and mechanism of protein phosphatases: insights into catalysis and regulation. *Annual review of biophysics and biomolecular structure* *27*, 133-164.
- Barral, Y., Parra, M., Bidlingmaier, S., and Snyder, M. (1999). Nim1-related kinases coordinate cell cycle progression with the organization of the peripheral cytoskeleton in yeast. *Genes & development* *13*, 176-187.
- Basehoar, A.D., Zanton, S.J., and Pugh, B.F. (2004). Identification and distinct regulation of yeast TATA box-containing genes. *Cell* *116*, 699-709.
- Basu, S., Mehreja, R., Thibierge, S., Chen, M.T., and Weiss, R. (2004). Spatiotemporal control of gene expression with pulse-generating networks. *Proceedings of the National Academy of Sciences of the United States of America* *101*, 6355-6360.

- Belle, A., Tanay, A., Bitincka, L., Shamir, R., and O'Shea, E.K. (2006). Quantification of protein half-lives in the budding yeast proteome. *Proceedings of the National Academy of Sciences of the United States of America* *103*, 13004-13009.
- Bhalla, U.S., and Iyengar, R. (1999). Emergent properties of networks of biological signaling pathways. *Science* *283*, 381-387.
- Bhardwaj, N., Kim, P.M., and Gerstein, M.B. (2010a). Rewiring of transcriptional regulatory networks: hierarchy, rather than connectivity, better reflects the importance of regulators. *Science signaling* *3*, ra79.
- Bhardwaj, N., Yan, K.K., and Gerstein, M.B. (2010b). Analysis of diverse regulatory networks in a hierarchical context shows consistent tendencies for collaboration in the middle levels. *Proceedings of the National Academy of Sciences of the United States of America* *107*, 6841-6846.
- Bhattacharyya, R.P., Remenyi, A., Yeh, B.J., and Lim, W.A. (2006). Domains, motifs, and scaffolds: the role of modular interactions in the evolution and wiring of cell signaling circuits. *Annual review of biochemistry* *75*, 655-680.
- Biondi, R.M., and Nebreda, A.R. (2003). Signalling specificity of Ser/Thr protein kinases through docking-site-mediated interactions. *The Biochemical journal* *372*, 1-13.
- Bloom, J., Cristea, I.M., Procko, A.L., Lubkov, V., Chait, B.T., Snyder, M., and Cross, F.R. (2011). Global analysis of Cdc14 phosphatase reveals diverse roles in mitotic processes. *The Journal of biological chemistry* *286*, 5434-5445.
- Bloom, J., and Cross, F.R. (2007). Multiple levels of cyclin specificity in cell-cycle control. *Nature reviews Molecular cell biology* *8*, 149-160.
- Bodenmiller, B., Wanka, S., Kraft, C., Urban, J., Campbell, D., Pedrioli, P.G., Gerrits, B., Picotti, P., Lam, H., Vitek, O., et al. (2010). Phosphoproteomic analysis reveals interconnected system-wide responses to perturbations of kinases and phosphatases in yeast. *Science signaling* *3*, rs4.
- Bollobàs, B., and Riordan, O. (2003). Mathematical results on scale-free graphs. In *Handbook of graphs and networks*, S.B.a.H. Schuster, ed. (Wiley-VCH), pp. 1-34.
- Bray, D. (1990). Intracellular signalling as a parallel distributed process. *Journal of theoretical biology* *143*, 215-231.
- Bray, D. (1995). Protein molecules as computational elements in living cells. *Nature* *376*, 307-312.
- Breitkreutz, A., Choi, H., Sharom, J.R., Boucher, L., Nedveda, V., Larsen, B., Lin, Z.Y., Breitkreutz, B.J., Stark, C., Liu, G., et al. (2010). A global protein kinase and phosphatase interaction network in yeast. *Science* *328*, 1043-1046.
- Carlson, S.M., Chouinard, C.R., Labadoff, A., Lam, C.J., Schmelzle, K., Fraenkel, E., and White, F.M. (2011). Large-scale discovery of ERK2 substrates identifies ERK-mediated transcriptional regulation by ETV3. *Science signaling* *4*, rs11.
- Carmel, L., Harel, D., and Koren, Y. (2004). Combining hierarchy and energy for drawing directed graphs. *IEEE transactions on visualization and computer graphics* *10*, 46-57.
- Caydası, A.K., Kurtulmus, B., Orrico, M.I., Hofmann, A., Ibrahim, B., and Pereira, G. (2010). Elm1 kinase activates the spindle position checkpoint kinase Kin4. *The Journal of cell biology* *190*, 975-989.
- Chan, S.W., and Dedon, P.C. (2010). The biological and metabolic fates of endogenous DNA damage products. *Journal of nucleic acids* *2010*, 929047.
- Charif, D., and Lobry, J.R. (2007). SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In *Structural approaches to sequence evolution: Molecules, networks, populations*, U.B.a.M.P.a.H.E.R.a.M. Vendruscolo, ed. (New York: Springer Verlag), pp. 207-232.
- Chavez, S., Beilharz, T., Rondon, A.G., Erdjument-Bromage, H., Tempst, P., Svejstrup, J.Q., Lithgow, T., and Aguilera, A. (2000). A protein complex containing Tho2, Hpr1, Mft1 and a novel protein, Thp2,

- connects transcription elongation with mitotic recombination in *Saccharomyces cerevisiae*. *The EMBO journal* **19**, 5824-5834.
- Chen, S.H., Albuquerque, C.P., Liang, J., Suhandynata, R.T., and Zhou, H. (2010). A proteome-wide analysis of kinase-substrate network in the DNA damage response. *The Journal of biological chemistry* **285**, 12803-12812.
- Chen, W.M., Danziger, S.A., Chiang, J.H., and Aitchison, J.D. (2013). PhosphoChain: a novel algorithm to predict kinase and phosphatase networks from high-throughput expression data. *Bioinformatics* **29**, 2435-2444.
- Cheng, C., Andrews, E., Yan, K.K., Ung, M., Wang, D., and Gerstein, M. (2015). An approach for determining and measuring network hierarchy applied to comparing the phosphorylome and the regulome. *Genome Biol* **16**, 63.
- Cherry, J.M., Hong, E.L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E.T., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S.R., et al. (2012). *Saccharomyces Genome Database*: the genomics resource of budding yeast. *Nucleic Acids Res* **40**, D700-705.
- Chi, A., Huttenhower, C., Geer, L.Y., Coon, J.J., Syka, J.E., Bai, D.L., Shabanowitz, J., Burke, D.J., Troyanskaya, O.G., and Hunt, D.F. (2007). Analysis of phosphorylation sites on proteins from *Saccharomyces cerevisiae* by electron transfer dissociation (ETD) mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 2193-2198.
- Chirolì, E., Rancati, G., Catusi, I., Lucchini, G., and Piatti, S. (2009). Cdc14 inhibition by the spindle assembly checkpoint prevents unscheduled centrosome separation in budding yeast. *Molecular biology of the cell* **20**, 2626-2637.
- Chong, Y.T., Koh, J.L., Friesen, H., Duffy, K., Cox, M.J., Moses, A., Moffat, J., Boone, C., and Andrews, B.J. (2015). Yeast Proteome Dynamics from Single Cell Imaging and Automated Analysis. *Cell* **161**, 1413-1424.
- Choudhary, C., Kumar, C., Gnad, F., Nielsen, M.L., Rehman, M., Walther, T.C., Olsen, J.V., and Mann, M. (2009). Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* **325**, 834-840.
- Ciesla, J., Fraczyk, T., and Rode, W. (2011). Phosphorylation of basic amino acid residues in proteins: important but easily missed. *Acta biochimica Polonica* **58**, 137-148.
- ClauSET, A., Shalitz, C., and Newman, M.E. (2009). Power-law distributions in empirical data. arXiv:07061062v2 [physicsdata-an].
- Clotet, J., Escote, X., Adrover, M.A., Yaakov, G., Gari, E., Aldea, M., de Nadal, E., and Posas, F. (2006). Phosphorylation of Hsl1 by Hog1 leads to a G2 arrest essential for cell survival at high osmolarity. *The EMBO journal* **25**, 2338-2346.
- Cohen, A.A., Geva-Zatorsky, N., Eden, E., Frenkel-Morgenstern, M., Issaeva, I., Sigal, A., Milo, R., Cohen-Saidon, C., Liron, Y., Kam, Z., et al. (2008). Dynamic proteomics of individual cancer cells in response to a drug. *Science* **322**, 1511-1516.
- Cohen, P. (1989). The structure and regulation of protein phosphatases. *Annual review of biochemistry* **58**, 453-508.
- Cohen, R., and Havlin, S. (2003). Scale-free networks are ultrasmall. *Physical review letters* **90**, 058701.
- Csikasz-Nagy, A., Kapuy, O., Toth, A., Pal, C., Jensen, L.J., Uhlmann, F., Tyson, J.J., and Novak, B. (2009). Cell cycle regulation by feed-forward loops coupling transcription and phosphorylation. *Molecular systems biology* **5**, 236.
- de Lichtenberg, U., Jensen, L.J., Brunak, S., and Bork, P. (2005). Dynamic complex formation during the yeast cell cycle. *Science* **307**, 724-727.
- Dephoure, N., Howson, R.W., Blethrow, J.D., Shokat, K.M., and O'Shea, E.K. (2005). Combining chemical genetics and proteomics to identify protein kinase substrates. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 17940-17945.

- Dihazi, H., Kessler, R., Muller, G.A., and Eschrich, K. (2005). Lysine 3 acetylation regulates the phosphorylation of yeast 6-phosphofructo-2-kinase under hypo-osmotic stress. *Biological chemistry* *386*, 895-900.
- Dinkel, H., Chica, C., Via, A., Gould, C.M., Jensen, L.J., Gibson, T.J., and Diella, F. (2011). Phospho.ELM: a database of phosphorylation sites--update 2011. *Nucleic Acids Res* *39*, D261-267.
- Dittrich, M.T., Klau, G.W., Rosenwald, A., Dandekar, T., and Muller, T. (2008). Identifying functional modules in protein-protein interaction networks: an integrated exact approach. *Bioinformatics* *24*, i223-231.
- Dosztanyi, Z., Csizmok, V., Tompa, P., and Simon, I. (2005). The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *Journal of molecular biology* *347*, 827-839.
- Duch, A., Felipe-Abrio, I., Barroso, S., Yaakov, G., Garcia-Rubio, M., Aguilera, A., de Nadal, E., and Posas, F. (2013). Coordinated control of replication and transcription by a SAPK protects genomic integrity. *Nature* *493*, 116-119.
- Ear, P.H., Booth, M.J., Abd-Rabbo, D., Kowarzyk Moreno, J., Hall, C., Chen, D., Vogel, J., and Michnick, S.W. (2013). Dissection of Cdk1-cyclin complexes in vivo. *Proceedings of the National Academy of Sciences of the United States of America* *110*, 15716-15721.
- Ear, P.H., and Michnick, S.W. (2009). A general life-death selection strategy for dissecting protein functions. *Nat Methods* *6*, 813-816.
- Elia, A.E., Rellos, P., Haire, L.F., Chao, J.W., Ivins, F.J., Hoepker, K., Mohammad, D., Cantley, L.C., Smerdon, S.J., and Yaffe, M.B. (2003). The molecular basis for phosphodependent substrate targeting and regulation of Plks by the Polo-box domain. *Cell* *115*, 83-95.
- Enserink, J.M., Hombauer, H., Huang, M.E., and Kolodner, R.D. (2009). Cdc28/Cdk1 positively and negatively affects genome stability in *S. cerevisiae*. *The Journal of cell biology* *185*, 423-437.
- Enserink, J.M., and Kolodner, R.D. (2010). An overview of Cdk1-controlled targets and processes. *Cell division* *5*, 11.
- Erdős, P., and Rényi, A. (1960). On the evolution of random graphs. *Publication of the Mathematical Institute of the Hungarian Academy of Sciences* *5*, 17-61.
- Eser, P., Demel, C., Maier, K.C., Schwalb, B., Pirk, N., Martin, D.E., Cramer, P., and Tresch, A. (2014). Periodic mRNA synthesis and degradation co-operate during cell cycle gene expression. *Molecular systems biology* *10*, 717.
- Falcon, S., and Gentleman, R. (2007). Using GOstats to test gene lists for GO term association. *Bioinformatics* *23*, 257-258.
- Ferrell, J.E., Jr. (1996). Tripping the switch fantastic: how a protein kinase cascade can convert graded inputs into switch-like outputs. *Trends Biochem Sci* *21*, 460-466.
- Ficarro, S.B., McCleland, M.L., Stukenberg, P.T., Burke, D.J., Ross, M.M., Shabanowitz, J., Hunt, D.F., and White, F.M. (2002). Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nat Biotechnol* *20*, 301-305.
- Fiedler, D., Braberg, H., Mehta, M., Chechik, G., Cagney, G., Mukherjee, P., Silva, A.C., Shales, M., Collins, S.R., van Wageningen, S., et al. (2009). Functional organization of the *S. cerevisiae* phosphorylation network. *Cell* *136*, 952-963.
- Fischer, H.P. (2008). Mathematical modeling of complex biological systems: from parts lists to understanding systems behavior. *Alcohol Res Health* *31*, 49-59.
- Fitch, I., Dahmann, C., Surana, U., Amon, A., Nasmyth, K., Goetsch, L., Byers, B., and Futcher, B. (1992). Characterization of four B-type cyclin genes of the budding yeast *Saccharomyces cerevisiae*. *Molecular biology of the cell* *3*, 805-818.

- Flake, G.W., Lawrence, S., and Giles, L.C. (2000). Efficient Identification of Web Communities. Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining, 150-160.
- Floyd, S.W., and Wooldridge, B. (1997). Middle management's strategic influence and organizational performance. *J Manage Stud* 34, 465-485.
- Foiani, M., Pellicoli, A., Lopes, M., Lucca, C., Ferrari, M., Liberi, G., Muzi Falconi, M., and Plevani, P. (2000). DNA damage checkpoints and DNA replication controls in *Saccharomyces cerevisiae*. *Mutation research* 451, 187-196.
- Frame, S., Cohen, P., and Biondi, R.M. (2001). A common phosphate binding site explains the unique substrate specificity of GSK3 and its inactivation by phosphorylation. *Molecular cell* 7, 1321-1327.
- Frigyes, K. (1929). "Láncszemek," in *Minden másképpen van*. 85-90.
- Gardner, M.K., Bouck, D.C., Paliulis, L.V., Meehl, J.B., O'Toole, E.T., Haase, J., Soubry, A., Joglekar, A.P., Winey, M., Salmon, E.D., et al. (2008). Chromosome congression by Kinesin-5 motor-mediated disassembly of longer kinetochore microtubules. *Cell* 135, 894-906.
- Garfinkel, R.S., and Nemhauser, G.L. (1972). Integer programming (New York: John Wiley & Sons).
- Gasch, A.P., and Werner-Washburne, M. (2002). The genomics of yeast responses to environmental stress and starvation. *Functional & integrative genomics* 2, 181-192.
- Gelperin, D.M., White, M.A., Wilkinson, M.L., Kon, Y., Kung, L.A., Wise, K.J., Lopez-Hoyo, N., Jiang, L., Piccirillo, S., Yu, H., et al. (2005). Biochemical and genetic analysis of the yeast proteome with a movable ORF collection. *Genes & development* 19, 2816-2826.
- Gerstein, M.B., Kundaje, A., Hariharan, M., Landt, S.G., Yan, K.K., Cheng, C., Mu, X.J., Khurana, E., Rozowsky, J., Alexander, R., et al. (2012). Architecture of the human regulatory network derived from ENCODE data. *Nature* 489, 91-100.
- Geymonat, M., Spanos, A., Jensen, S., and Sedgwick, S.G. (2010). Phosphorylation of Lte1 by Cdk prevents polarized growth during mitotic arrest in *S. cerevisiae*. *The Journal of cell biology* 191, 1097-1112.
- Ghaemmaghami, S.H., W.K.; Bower, K; Howson, R. W.; Belle, A.; Dephoure, N.; O'Shea, E. K.; Weissman, J. S (2003). Global analysis of protein expression in yeast. . *Nature* 425(6959) 737-741.
- Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Veronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., Andre, B., et al. (2002). Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* 418, 387-391.
- Goh, K.I., Cusick, M.E., Valle, D., Childs, B., Vidal, M., and Barabasi, A.L. (2007). The human disease network. *Proceedings of the National Academy of Sciences of the United States of America* 104, 8685-8690.
- Granovetter, M.S. (1973). the strength of weak ties. *American Journal of Sociology* 78, 1360-1380.
- Grava, S., Schaefer, F., Faty, M., Philippijen, P., and Barral, Y. (2006). Asymmetric recruitment of dynein to spindle poles and microtubules promotes proper spindle orientation in yeast. *Developmental cell* 10, 425-439.
- Graves, J.D., and Krebs, E.G. (1999). Protein phosphorylation and signal transduction. *Pharmacology & therapeutics* 82, 111-121.
- Gruhler, A., Olsen, J.V., Mohammed, S., Mortensen, P., Faergeman, N.J., Mann, M., and Jensen, O.N. (2005). Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. *Molecular & cellular proteomics : MCP* 4, 310-327.
- Guare, J. (1990). Six Degrees of Separation: A Play, First edition edn (New York: Random House).
- Guerrero, C., Milenkovic, T., Przulj, N., Kaiser, P., and Huang, L. (2008). Characterization of the proteasome interaction network using a QTAX-based tag-team strategy and protein interaction network analysis. *Proceedings of the National Academy of Sciences of the United States of America* 105, 13333-13338.

- Gulsoy, G., Bandhyopadhyay, N., and Kahveci, T. (2012). HIDEN: Hierarchical decomposition of regulatory networks. *BMC bioinformatics* 13, 250.
- Hakes, L., Pinney, J.W., Robertson, D.L., and Lovell, S.C. (2008). Protein-protein interaction networks and biology--what's the connection? *Nat Biotechnol* 26, 69-72.
- Han, J.D., Dupuy, D., Bertin, N., Cusick, M.E., and Vidal, M. (2005). Effect of sampling on topology predictions of protein-protein interaction networks. *Nat Biotechnol* 23, 839-844.
- Hao, D., Ren, C., and Li, C. (2012). Revisiting the variation of clustering coefficient of biological networks suggests new modular structure. *BMC systems biology* 6, 34.
- Hartsperger, M.L., Strache, R., and Stumpflen, V. (2010). HiNO: an approach for inferring hierarchical organization from regulatory networks. *PloS one* 5, e13698.
- Hartwell, L.H., Hopfield, J.J., Leibler, S., and Murray, A.W. (1999). From molecular to modular cell biology. *Nature* 402, C47-52.
- Harvey, S.L., Charlet, A., Haas, W., Gygi, S.P., and Kellogg, D.R. (2005). Cdk1-dependent regulation of the mitotic inhibitor Wee1. *Cell* 122, 407-420.
- Heinrich, R., Neel, B.G., and Rapoport, T.A. (2002). Mathematical models of protein kinase signal transduction. *Mol Cell* 9, 957-970.
- Helikar, T., Konvalina, J., Heidel, J., and Rogers, J.A. (2008). Emergent decision-making in biological signal transduction networks. *Proceedings of the National Academy of Sciences of the United States of America* 105, 1913-1918.
- Hendriks, W.J., Elson, A., Harroch, S., Pulido, R., Stoker, A., and den Hertog, J. (2013). Protein tyrosine phosphatases in health and disease. *The FEBS journal* 280, 708-730.
- Henning, C. (2007). Cluster-wise assessment of cluster stability. *Comput Stat Data An* 52, 258 - 271.
- Henriksen, P., Wagner, S.A., Weinert, B.T., Sharma, S., Bacinskaia, G., Rehman, M., Juffer, A.H., Walther, T.C., Lisby, M., and Choudhary, C. (2012). Proteome-wide analysis of lysine acetylation suggests its broad regulatory scope in *Saccharomyces cerevisiae*. *Molecular & cellular proteomics : MCP* 11, 1510-1522.
- Hertz, J., Krogh, A., and Palmer, R. (1991). *Introduction to the theory of neural computation* (United States of America: Westview Press).
- Holt, L.J., Tuch, B.B., Villen, J., Johnson, A.D., Gygi, S.P., and Morgan, D.O. (2009). Global analysis of Cdk1 substrate phosphorylation sites provides insights into evolution. *Science* 325, 1682-1686.
- Hornberg, J.J., Bruggeman, F.J., Binder, B., Geest, C.R., de Vaate, A.J., Lankelma, J., Heinrich, R., and Westerhoff, H.V. (2005). Principles behind the multifarious control of signal transduction. ERK phosphorylation and kinase/phosphatase control. *FEBS J* 272, 244-258.
- Huber, W., Carey, V.J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B.S., Bravo, H.C., Davis, S., Gatto, L., Girke, T., et al. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Methods* 12, 115-121.
- Hunter, T. (2000). Signaling--2000 and beyond. *Cell* 100, 113-127.
- Hunter, T., and Plowman, G.D. (1997). The protein kinases of budding yeast: six score and more. *Trends Biochem Sci* 22, 18-22.
- Hynne, F., Dano, S., and Sorensen, P.G. (2001). Full-scale model of glycolysis in *Saccharomyces cerevisiae*. *Biophys Chem* 94, 121-163.
- Iesmantavicius, V., Weinert, B.T., and Choudhary, C. (2014). Convergence of Ubiquitylation and Phosphorylation Signaling in Rapamycin-treated Yeast Cells. *Molecular & cellular proteomics : MCP* 13, 1979-1992.
- Iglesias, P.A. (2013). Systems biology: the role of engineering in the reverse engineering of biological signaling. *Cells* 2, 393-413.
- Ingalls, B.P. (2013). Signal Transduction Pathways. In *Mathematical Modeling in Systems Biology An introduction* (The MIT Press), pp. 149-155.

- Janssens, V., Longin, S., and Goris, J. (2008). PP2A holoenzyme assembly: in cauda venenum (the sting is in the tail). *Trends Biochem Sci* 33, 113-121.
- Jeong, H., Mason, S.P., Barabasi, A.L., and Oltvai, Z.N. (2001). Lethality and centrality in protein networks. *Nature* 411, 41-42.
- Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N., and Barabasi, A.L. (2000). The large-scale organization of metabolic networks. *Nature* 407, 651-654.
- Jia, T., Liu, Y.Y., Csoka, E., Posfai, M., Slotine, J.J., and Barabasi, A.L. (2013). Emergence of bimodality in controlling complex networks. *Nature communications* 4, 2002.
- Jin, F., Liu, H., Liang, F., Rizkallah, R., Hurt, M.M., and Wang, Y. (2008). Temporal control of the dephosphorylation of Cdk substrates by mitotic exit pathways in budding yeast. *Proceedings of the National Academy of Sciences of the United States of America* 105, 16177-16182.
- Jonsson, P.F., and Bates, P.A. (2006). Global topological features of cancer proteins in the human interactome. *Bioinformatics* 22, 2291-2297.
- Jothi, R., Balaji, S., Wuster, A., Grochow, J.A., Gsponer, J., Przytycka, T.M., Aravind, L., and Babu, M.M. (2009). Genomic analysis reveals a tight link between transcription factor dynamics and regulatory network architecture. *Molecular systems biology* 5, 294.
- Kanshin, E., Bergeron-Sandoval, L.P., Isik, S.S., Thibault, P., and Michnick, S.W. (2015). A cell-signaling network temporally resolves specific versus promiscuous phosphorylation. *Cell reports* 10, 1202-1214.
- Kantarci, B., and Labatut, V. (2014). Classification of Complex Networks Based on Topological Properties. 3rd Conference on Social Computing and its Applications, Karlsruhe : Germany (2013) *arXiv:1402.0238 [cs.SI]*.
- Keck, J.M., Jones, M.H., Wong, C.C., Binkley, J., Chen, D., Jaspersen, S.L., Holinger, E.P., Xu, T., Niepel, M., Rout, M.P., et al. (2011). A cell cycle phosphoproteome of the yeast centrosome. *Science* 332, 1557-1561.
- Kerrien, S., Aranda, B., Breuza, L., Bridge, A., Broackes-Carter, F., Chen, C., Duesbury, M., Dumousseau, M., Feuermann, M., Hinz, U., et al. (2012). The IntAct molecular interaction database in 2012. *Nucleic Acids Res* 40, D841-846.
- Khammash, M. (2008). Reverse engineering: the architecture of biological networks. *Biotechniques* 44, 323-329.
- Khanin, R., and Wit, E. (2006). How scale-free are biological networks. *Journal of computational biology : a journal of computational molecular cell biology* 13, 810-818.
- Kim, D., Kim, M.S., and Cho, K.H. (2012). The core regulation module of stress-responsive regulatory networks in yeast. *Nucleic Acids Res* 40, 8793-8802.
- King, K., Kang, H., Jin, M., and Lew, D.J. (2013). Feedback control of Swe1p degradation in the yeast morphogenesis checkpoint. *Molecular biology of the cell* 24, 914-922.
- Kitano, H. (2004). Biological robustness. *Nat Rev Genet* 5, 826-837.
- Kitazono, A.A., Garza, D.A., and Kron, S.J. (2003). Mutations in the yeast cyclin-dependent kinase Cdc28 reveal a role in the spindle assembly checkpoint. *Molecular genetics and genomics : MGG* 269, 672-684.
- Knaus, J., Porzelius, C., Binder, H., and Schwarzer, G. (2009). Easier parallel computing in R with snowfall and sfCluster. *The R Journal* 1, 54-59.
- Knight, J.D., Pawson, T., and Gingras, A.C. (2013). Profiling the kinome: current capabilities and future challenges. *Journal of proteomics* 81, 43-55.
- Koivomagi, M., Valk, E., Venta, R., Iofik, A., Lepiku, M., Morgan, D.O., and Loog, M. (2011). Dynamics of Cdk1 substrate specificity during the cell cycle. *Molecular cell* 42, 610-623.
- Kollman, J.M., Merdes, A., Mourey, L., and Agard, D.A. (2011). Microtubule nucleation by gamma-tubulin complexes. *Nature reviews Molecular cell biology* 12, 709-721.

- Kollman, J.M., Polka, J.K., Zelter, A., Davis, T.N., and Agard, D.A. (2010). Microtubule nucleating gamma-TuSC assembles structures with 13-fold microtubule-like symmetry. *Nature* 466, 879-882.
- Krackhardt, D. (1994). Graph theoretical dimensions of informal organizations. In Computational organization theory (NJ, USA: L. Erlbaum Associates Inc), pp. 89-111.
- Kreegipuu, A., Blom, N., Brunak, S., and Jarv, J. (1998). Statistical analysis of protein kinase specificity determinants. *FEBS letters* 430, 45-50.
- Kriegenburg, F., Ellgaard, L., and Hartmann-Petersen, R. (2012). Molecular chaperones in targeting misfolded proteins for ubiquitin-dependent degradation. *The FEBS journal* 279, 532-542.
- Kurtz, J.E., Exinger, F., Erbs, P., and Jund, R. (1999). New insights into the pyrimidine salvage pathway of *Saccharomyces cerevisiae*: requirement of six genes for cytidine metabolism. *Current genetics* 36, 130-136.
- Lage, K., Karlberg, E.O., Storling, Z.M., Olason, P.I., Pedersen, A.G., Rigina, O., Hinsby, A.M., Turner, Z., Pociot, F., Tommerup, N., et al. (2007). A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nat Biotechnol* 25, 309-316.
- Lahiry, P., Torkamani, A., Schork, N.J., and Hegele, R.A. (2010). Kinase mutations in human disease: interpreting genotype-phenotype relationships. *Nat Rev Genet* 11, 60-74.
- Lai, W., and Aldous, D. (2012). Fitting Power Law Distributions to Data.
- Landry, C.R., Levy, E.D., and Michnick, S.W. (2009). Weak functional constraints on phosphoproteomes. *Trends in genetics : TIG* 25, 193-197.
- Lane, D. (2006). Hierarchy, Complexity, Society. In *Hierarchy in Natural and Social Sciences Methodos Series*, D. Pumain, ed. (Dordrecht, The Netherlands: Springer), pp. 81-119.
- Lasalde, C., Rivera, A.V., Leon, A.J., Gonzalez-Feliciano, J.A., Estrella, L.A., Rodriguez-Cruz, E.N., Correa, M.E., Cajigas, I.J., Bracho, D.P., Vega, I.E., et al. (2014). Identification and functional analysis of novel phosphorylation sites in the RNA surveillance protein Upf1. *Nucleic Acids Res* 42, 1916-1929.
- Lauffenburger, D.A. (2000). Cell signaling pathways as control modules: complexity for simplicity? *Proceedings of the National Academy of Sciences of the United States of America* 97, 5031-5033.
- Levy, E.D., Michnick, S.W., and Landry, C.R. (2012). Protein abundance is key to distinguish promiscuous from functional phosphorylation based on evolutionary information. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* 367, 2594-2606.
- Li, X., Gerber, S.A., Rudner, A.D., Beausoleil, S.A., Haas, W., Villen, J., Elias, J.E., and Gygi, S.P. (2007). Large-scale phosphorylation analysis of alpha-factor-arrested *Saccharomyces cerevisiae*. *Journal of proteome research* 6, 1190-1197.
- Liakopoulos, D., Kusch, J., Grava, S., Vogel, J., and Barral, Y. (2003). Asymmetric loading of Kar9 onto spindle poles and microtubules ensures proper spindle alignment. *Cell* 112, 561-574.
- Lianga, N., Williams, E.C., Kennedy, E.K., Dore, C., Pilon, S., Girard, S.L., Deneault, J.S., and Rudner, A.D. (2013). A Wee1 checkpoint inhibits anaphase onset. *The Journal of cell biology* 201, 843-862.
- Licata, L., Briganti, L., Peluso, D., Perfetto, L., Iannuccelli, M., Galeota, E., Sacco, F., Palma, A., Nardozza, A.P., Santonico, E., et al. (2012). MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res* 40, D857-861.
- Lienhard, G.E. (2008). Non-functional phosphorylations? *Trends Biochem Sci* 33, 351-352.
- Lin, T.C., Gombos, L., Neuner, A., Sebastian, D., Olsen, J.V., Hrle, A., Benda, C., and Schiebel, E. (2011). Phosphorylation of the yeast gamma-tubulin Tub4 regulates microtubule function. *PloS one* 6, e19700.
- Linding, R., Jensen, L.J., Osthimer, G.J., van Vugt, M.A., Jorgensen, C., Miron, I.M., Diella, F., Colwill, K., Taylor, L., Elder, K., et al. (2007). Systematic discovery of in vivo phosphorylation networks. *Cell* 129, 1415-1426.

- Lipshtat, A., Purushothaman, S.P., Iyengar, R., and Ma'ayan, A. (2008). Functions of bifans in context of multiple regulatory motifs in signaling networks. *Biophysical journal* 94, 2566-2579.
- Loog, M., and Morgan, D.O. (2005). Cyclin specificity in the phosphorylation of cyclin-dependent kinase substrates. *Nature* 434, 104-108.
- Ma, H.W., Buer, J., and Zeng, A.P. (2004). Hierarchical structure and modules in the Escherichia coli transcriptional regulatory network revealed by a new top-down approach. *BMC bioinformatics* 5, 199.
- Ma, H.W., and Zeng, A.P. (2003). The connectivity structure, giant strong component and centrality of metabolic networks. *Bioinformatics* 19, 1423-1430.
- Maeda, T., Tsai, A.Y., and Saito, H. (1993). Mutations in a protein tyrosine phosphatase gene (PTP2) and a protein serine/threonine phosphatase gene (PTC1) cause a synthetic growth defect in *Saccharomyces cerevisiae*. *Molecular and cellular biology* 13, 5408-5417.
- Maekawa, H., and Schiebel, E. (2004). Cdk1-Clb4 controls the interaction of astral microtubule plus ends with subdomains of the daughter cell cortex. *Genes & development* 18, 1709-1724.
- Maekawa, H., Usui, T., Knop, M., and Schiebel, E. (2003). Yeast Cdk1 translocates to the plus end of cytoplasmic microtubules to regulate bud cortex interactions. *The EMBO journal* 22, 438-449.
- Magrane, M., and Consortium, U. (2011). UniProt Knowledgebase: a hub of integrated protein data. *Database (Oxford)* 2011, bar009.
- Mah, A.S., Elia, A.E., Devgan, G., Ptacek, J., Schutkowski, M., Snyder, M., Yaffe, M.B., and Deshaies, R.J. (2005). Substrate specificity analysis of protein kinase complex Dbf2-Mob1 by peptide library and proteome array screening. *BMC biochemistry* 6, 22.
- Malathi, K., Xiao, Y., and Mitchell, A.P. (1999). Catalytic roles of yeast GSK3beta/shaggy homolog Rim11p in meiotic activation. *Genetics* 153, 1145-1152.
- Malik, R., Nigg, E.A., and Korner, R. (2008). Comparative conservation analysis of the human mitotic phosphoproteome. *Bioinformatics* 24, 1426-1432.
- Mangan, S., Zaslaver, A., and Alon, U. (2003). The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks. *Journal of molecular biology* 334, 197-204.
- Manning, G., Plowman, G.D., Hunter, T., and Sudarsanam, S. (2002a). Evolution of protein kinase signaling from yeast to man. *Trends Biochem Sci* 27, 514-520.
- Manning, G., Whyte, D.B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002b). The protein kinase complement of the human genome. *Science* 298, 1912-1934.
- Mattison, C.P., and Ota, I.M. (2000). Two protein tyrosine phosphatases, Ptp2 and Ptp3, modulate the subcellular localization of the Hog1 MAP kinase in yeast. *Genes & development* 14, 1229-1235.
- Memon, N., Larsen, H.L., Hicks, D.L., and Harkiolakis, N. (2008). Detecting hidden hierarchy in terrorist networks: some case studies. Paper presented at: Proceedings of the IEEE ISI 2008 PAISI, PACCF, and SOCO international workshops on Intelligence and Security Informatics (New York: Springer-Verlag).
- Mendenhall, M.D. (1993). An inhibitor of p34CDC28 protein kinase activity from *Saccharomyces cerevisiae*. *Science* 259, 216-219.
- Mendenhall, M.D., and Hodge, A.E. (1998). Regulation of Cdc28 cyclin-dependent protein kinase activity during the cell cycle of the yeast *Saccharomyces cerevisiae*. *Microbiol Mol Biol Rev* 62, 1191-1243.
- Meszaros, B., Simon, I., and Dosztanyi, Z. (2009). Prediction of protein binding regions in disordered proteins. *PLoS Comput Biol* 5, e1000376.
- Mewes, H.W., Frishman, D., Guldener, U., Mannhaupt, G., Mayer, K., Mokrejs, M., Morgenstern, B., Munsterkotter, M., Rudd, S., and Weil, B. (2002). MIPS: a database for genomes and protein sequences. *Nucleic Acids Res* 30, 31-34.

- Milenkovic, T., Memisevic, V., Ganesan, A.K., and Przulj, N. (2010). Systems-level cancer gene identification from protein interaction network topology applied to melanogenesis-related functional genomics data. *Journal of the Royal Society, Interface / the Royal Society* 7, 423-437.
- Milenkovic, T., and Przulj, N. (2008). Uncovering biological network function via graphlet degree signatures. *Cancer informatics* 6, 257-273.
- Milgram, S. (1967). The small world problem. *Psychol Today* 2, 60-67.
- Miller, M.E., and Cross, F.R. (2000). Distinct subcellular localization patterns contribute to functional specificity of the Cln2 and Cln3 cyclins of *Saccharomyces cerevisiae*. *Molecular and cellular biology* 20, 542-555.
- Milo, R., Kashtan, N., Itzkovitz, S., Newman, M.E.J., and Alon, U. (2004). On the uniform generation of random graphs with prescribed degree sequences. arXiv:cond-mat/0312028v2 [cond-matstat-mech].
- Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network motifs: simple building blocks of complex networks. *Science* 298, 824-827.
- Mitchell, D.A., and Sprague, G.F., Jr. (2001). The phosphotyrosyl phosphatase activator, Ncs1p (Rrd1p), functions with Cla4p to regulate the G(2)/M transition in *Saccharomyces cerevisiae*. *Molecular and cellular biology* 21, 488-500.
- Miura, F., Kawaguchi, N., Yoshida, M., Uematsu, C., Kito, K., Sakaki, Y., and Ito, T. (2008). Absolute quantification of the budding yeast transcriptome by means of competitive PCR between genomic and complementary DNAs. *BMC Genomics* 9, 574.
- Mok, J., Im, H., and Snyder, M. (2009). Global identification of protein kinase substrates by protein microarray analysis. *Nature protocols* 4, 1820-1827.
- Mok, J., Kim, P.M., Lam, H.Y., Piccirillo, S., Zhou, X., Jeschke, G.R., Sheridan, D.L., Parker, S.A., Desai, V., Jwa, M., et al. (2010). Deciphering protein kinase specificity through large-scale analysis of yeast phosphorylation site motifs. *Science signaling* 3, ra12.
- Mok, J., Zhu, X., and Snyder, M. (2011). Dissecting phosphorylation networks: lessons learned from yeast. *Expert review of proteomics* 8, 775-786.
- Molloy, M., and Reed, B. (1995). A critical point for random graphs with a given degree sequence. *Random Structures and Algorithms* 6, 161–179.
- Mones, E. (2013). Hierarchy in directed random networks. *Physical review E, Statistical, nonlinear, and soft matter physics* 87, 022817.
- Mones, E., Vicsek, L., and Vicsek, T. (2012). Hierarchy measure for complex networks. *PloS one* 7, e33799.
- Moore, J.K., D'Silva, S., and Miller, R.K. (2006). The CLIP-170 homologue Bik1p promotes the phosphorylation and asymmetric localization of Kar9p. *Molecular biology of the cell* 17, 178-191.
- Moore, J.K., and Miller, R.K. (2007). The cyclin-dependent kinase Cdc28p regulates multiple aspects of Kar9p function in yeast. *Molecular biology of the cell* 18, 1187-1202.
- Morgan, D.O. (1995). Principles of CDK regulation. *Nature* 374, 131-134.
- Morgan, D.O. (1997). Cyclin-dependent kinases: engines, clocks, and microprocessors. *Annu Rev Cell Dev Biol* 13, 261-291.
- Morrison, D.K., Murakami, M.S., and Cleghon, V. (2000). Protein kinases and phosphatases in the *Drosophila* genome. *The Journal of cell biology* 150, F57-62.
- Muller, H.M., Kenny, E.E., and Sternberg, P.W. (2004). Textpresso: an ontology-based information retrieval and extraction system for biological literature. *PLoS biology* 2, e309.
- Muzzey, D., Gomez-Uribe, C.A., Mettetal, J.T., and van Oudenaarden, A. (2009). A systems-level analysis of perfect adaptation in yeast osmoregulation. *Cell* 138, 160-171.
- Nagy, M., Akos, Z., Biro, D., and Vicsek, T. (2010). Hierarchical group dynamics in pigeon flocks. *Nature* 464, 890-893.

- Nazarova, E., O'Toole, E., Kaitna, S., Francois, P., Winey, M., and Vogel, J. (2013). Distinct roles for antiparallel microtubule pairing and overlap during early spindle assembly. *Molecular biology of the cell* 24, 3238-3250.
- Newman, J.R., Ghaemmaghami, S., Ihmels, J., Breslow, D.K., Noble, M., DeRisi, J.L., and Weissman, J.S. (2006). Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* 441, 840-846.
- Newman, M.E.J., Strogatz, S.H., and Watts, D.J. (2001). Random graphs with arbitrary degree distribution and their applications. *Phys Rev* 6.
- Nise, N.S. (2015). Control Systems Engineering (United States of America: John Wiley and Sons, Inc.).
- Novak, B., Kapuy, O., Domingo-Sananes, M.R., and Tyson, J.J. (2010). Regulated protein kinases and phosphatases in cell cycle decisions. *Current opinion in cell biology* 22, 801-808.
- Ogi, H., Wang, C.Z., Nakai, W., Kawasaki, Y., and Masumoto, H. (2008). The role of the *Saccharomyces cerevisiae* Cdc7-Dbf4 complex in the replication checkpoint. *Gene* 414, 32-40.
- Olsen, J.V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P., and Mann, M. (2006). Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* 127, 635-648.
- Olsen, J.V., Vermeulen, M., Santamaria, A., Kumar, C., Miller, M.L., Jensen, L.J., Gnad, F., Cox, J., Jensen, T.S., Nigg, E.A., et al. (2010). Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Science signaling* 3, ra3.
- Oti, M., and Brunner, H.G. (2007). The modular nature of genetic diseases. *Clinical genetics* 71, 1-11.
- Oti, M., Snel, B., Huynen, M.A., and Brunner, H.G. (2006). Predicting disease genes using protein-protein interactions. *Journal of medical genetics* 43, 691-698.
- Pang, C.N., Gasteiger, E., and Wilkins, M.R. (2010). Identification of arginine- and lysine-methylation in the proteome of *Saccharomyces cerevisiae* and its functional implications. *BMC Genomics* 11, 92.
- Pappas, J. (2001). Strategic knowledge, social structure, and middle management activities: A study of strategic renewal. In Isenberg School of Management (Amherst: University of Massachusetts).
- Pawson, T., and Scott, J.D. (1997). Signaling through scaffold, anchoring, and adaptor proteins. *Science* 278, 2075-2080.
- Pawson, T., and Scott, J.D. (2005). Protein phosphorylation in signaling--50 years and counting. *Trends Biochem Sci* 30, 286-290.
- Perkins, T.J., Jaeger, J., Reinitz, J., and Glass, L. (2006). Reverse engineering the gap gene network of *Drosophila melanogaster*. *PLoS Comput Biol* 2, e51.
- Pinna, L.A., and Ruzzene, M. (1996). How do protein kinases recognize their substrates? *Biochimica et biophysica acta* 1314, 191-225.
- Polo, S.E., and Jackson, S.P. (2011). Dynamics of DNA damage response proteins at DNA breaks: a focus on protein modifications. *Genes & development* 25, 409-433.
- Przulj, N., Corneil, D.G., and Jurisica, I. (2004). Modeling interactome: scale-free or geometric? *Bioinformatics* 20, 3508-3515.
- Ptacek, J., Devgan, G., Michaud, G., Zhu, H., Zhu, X., Fasolo, J., Guo, H., Jona, G., Breitkreutz, A., Sopko, R., et al. (2005). Global analysis of protein phosphorylation in yeast. *Nature* 438, 679-684.
- Putnam, C.D., Jaehnig, E.J., and Kolodner, R.D. (2009). Perspectives on the DNA damage and replication checkpoint responses in *Saccharomyces cerevisiae*. *DNA repair* 8, 974-982.
- R-Development-Core-Team (2011). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing {ISBN} 3-900051-07-0.
- R Core Team (2012). R: A Language and Environment for Statistical Computing.
- Rahal, R., and Amon, A. (2008). Mitotic CDKs control the metaphase-anaphase transition and trigger spindle elongation. *Genes & development* 22, 1534-1548.
- Ravasz, E., and Barabasi, A.L. (2003). Hierarchical organization in complex networks. *Physical review E, Statistical, nonlinear, and soft matter physics* 67, 026112.

- Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N., and Barabasi, A.L. (2002). Hierarchical organization of modularity in metabolic networks. *Science* 297, 1551-1555.
- Richardson, H., Lew, D.J., Henze, M., Sugimoto, K., and Reed, S.I. (1992). Cyclin-B homologs in *Saccharomyces cerevisiae* function in S phase and in G2. *Genes & development* 6, 2021-2034.
- Ronen, M., Rosenberg, R., Shraiman, B.I., and Alon, U. (2002). Assigning numbers to the arrows: parameterizing a gene regulation network by using accurate expression kinetics. *Proceedings of the National Academy of Sciences of the United States of America* 99, 10555-10560.
- Rosso, V., and Yoshida, S. (2011). Spatial regulation of Cdc55-PP2A by Zds1/Zds2 controls mitotic entry and mitotic exit in budding yeast. *The Journal of cell biology* 193, 445-454.
- Rowe, R., Creamer, G., Herskowitz, S., and Stolfo, S.J. (2007). Automated social hierarchy detection through email network analysis 2007 workshop on Web mining and social network analysis. Paper presented at: Proceedings of the 9th WebKDD and 1st SNA-KDD.
- Rowicka, M., Kudlicki, A., Tu, B.P., and Otwinowski, Z. (2007). High-resolution timing of cell cycle-regulated gene expression. *Proceedings of the National Academy of Sciences of the United States of America* 104, 16892-16897.
- Royer, C. (1999). Outline of the thermodynamic and structural principles governing the ways that proteins interact with other proteins. Previously published in the Biophysics Textbook Online (BTOL). In Protein-protein interactions.
- Russo, A.A., Jeffrey, P.D., Patten, A.K., Massague, J., and Pavletich, N.P. (1996). Crystal structure of the p27Kip1 cyclin-dependent-kinase inhibitor bound to the cyclin A-Cdk2 complex. *Nature* 382, 325-331.
- Salwinski, L., Miller, C.S., Smith, A.J., Pettit, F.K., Bowie, J.U., and Eisenberg, D. (2004). The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res* 32, D449-451.
- Sambour, L., and Thierry-Mieg, N. (2010). New insights into protein-protein interaction data lead to increased estimates of the *S. cerevisiae* interactome size. *BMC bioinformatics* 11, 605.
- Santini, C.C., Bath, J., Turberfield, A.J., and Tyrrell, A.M. (2012). A DNA network as an information processing system. *Int J Mol Sci* 13, 5125-5137.
- Schwartz, M.A., and Madhani, H.D. (2004). Principles of MAP kinase signaling specificity in *Saccharomyces cerevisiae*. *Annual review of genetics* 38, 725-748.
- Schwob, E., Bohm, T., Mendenhall, M.D., and Nasmyth, K. (1994). The B-type cyclin kinase inhibitor p40SIC1 controls the G1 to S transition in *S. cerevisiae*. *Cell* 79, 233-244.
- Schwob, E., and Nasmyth, K. (1993). CLB5 and CLB6, a new pair of B cyclins involved in DNA replication in *Saccharomyces cerevisiae*. *Genes & development* 7, 1160-1175.
- Seet, B.T., Dikic, I., Zhou, M.M., and Pawson, T. (2006). Reading protein modifications with interaction domains. *Nature reviews Molecular cell biology* 7, 473-483.
- SGD Project (Aug. 2015). <http://downloads.yeastgenome.org/curation/literature/archive/>.
- SGD Project (Nov. 2014). http://downloads.yeastgenome.org/sequence/S288C_reference/orf_protein/archive/.
- Shah, K., Liu, Y., Deirmengian, C., and Shokat, K.M. (1997). Engineering unnatural nucleotide specificity for Rous sarcoma virus tyrosine kinase to uniquely label its direct substrates. *Proceedings of the National Academy of Sciences of the United States of America* 94, 3565-3570.
- Sharifpoor, S., Nguyen Ba, A.N., Young, J.Y., van Dyk, D., Friesen, H., Douglas, A.C., Kurat, C.F., Chong, Y.T., Founk, K., Moses, A.M., et al. (2011). A quantitative literature-curated gold standard for kinase-substrate pairs. *Genome Biol* 12, R39.
- Shen-Orr, S.S., Milo, R., Mangan, S., and Alon, U. (2002). Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature genetics* 31, 64-68.
- Shi, Y. (2009). Serine/threonine phosphatases: mechanism through structure. *Cell* 139, 468-484.

- Simon, H. (1973). The organization of complex systems. In *Hierarchy Theory: The Challenge of Complex Systems*, H. Pattee, ed. (New York: George Braziller).
- Smolka, M.B., Albuquerque, C.P., Chen, S.H., and Zhou, H. (2007). Proteome-wide identification of *in vivo* targets of DNA damage checkpoint kinases. *Proceedings of the National Academy of Sciences of the United States of America* *104*, 10364-10369.
- Sneppen, K., Krishna, S., and Semsey, S. (2010). Simplified models of biological networks. *Annual review of biophysics* *39*, 43-59.
- Song, X., Chi, Y., Hino, K., and Tseng, B.L. (2007). Identifying opinion leaders in the blogosphere. Paper presented at: Proceedings of the sixteenth ACM conference on Conference on information and knowledge management ACM, CIKM.
- Soufi, B., Kelstrup, C.D., Stoehr, G., Frohlich, F., Walther, T.C., and Olsen, J.V. (2009). Global analysis of the yeast osmotic stress response by quantitative proteomics. *Molecular bioSystems* *5*, 1337-1346.
- Spencer, S.L., Gaudet, S., Albeck, J.G., Burke, J.M., and Sorger, P.K. (2009). Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nature* *459*, 428-432.
- Stark, C., Breitkreutz, B.J., Reguly, T., Boucher, L., Breitkreutz, A., and Tyers, M. (2006). BioGRID: a general repository for interaction datasets. *Nucleic Acids Res* *34*, D535-539.
- Stark, C., Su, T.C., Breitkreutz, A., Lourenco, P., Dahabieh, M., Breitkreutz, B.J., Tyers, M., and Sadowski, I. (2010). PhosphoGRID: a database of experimentally verified *in vivo* protein phosphorylation sites from the budding yeast *Saccharomyces cerevisiae*. *Database (Oxford)* *2010*, bap026.
- Stark, M.J. (1996). Yeast protein serine/threonine phosphatases: multiple roles and diverse regulation. *Yeast* *12*, 1647-1675.
- Steen, M.v. (2010). Graph Theory and Complex Networks An Introduction, M.v. Steen, ed., pp. 285.
- Strogatz, S.H. (2001). Exploring complex networks. *Nature* *410*, 268-276.
- Stumpf, M.P., Ingram, P.J., Nouvel, I., and Wiuf, C. (2005a). Statistical model selection methods applied to biological network data. *Trans Comp Syst Biol* *3*, 65-77.
- Stumpf, M.P., Wiuf, C., and May, R.M. (2005b). Subnets of scale-free networks are not scale-free: sampling properties of networks. *Proceedings of the National Academy of Sciences of the United States of America* *102*, 4221-4224.
- Su, S.S., and Mitchell, A.P. (1993). Identification of functionally related genes that stimulate early meiotic gene expression in yeast. *Genetics* *133*, 67-77.
- Sugiyama, K., Tagawa, S., and Troda, M. (1981). Methods for visual understanding of hierarchical system structures. In *IEEE Transactions in Systems, Man and Cybernetics*, pp. 100-125.
- Sun, M., Schwalb, B., Schulz, D., Pirk, N., Etzold, S., Lariviere, L., Maier, K.C., Seizl, M., Tresch, A., and Cramer, P. (2012). Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. *Genome Res* *22*, 1350-1359.
- Szomolay, B., and Shahrezaei, V. (2012). Bell-shaped and ultrasensitive dose-response in phosphorylation-dephosphorylation cycles: the role of kinase-phosphatase complex formation. *BMC systems biology* *6*, 26.
- Tan, C.S., Jorgensen, C., and Linding, R. (2010). Roles of "junk phosphorylation" in modulating biomolecular association of phosphorylated proteins? *Cell cycle* *9*, 1276-1280.
- Tarassov, K., and Michnick, S.W. (2005). iVici: Interrelational Visualization and Correlation Interface. *Genome Biol* *6*, R115.
- Tavormina, P.A., and Burke, D.J. (1998). Cell cycle arrest in cdc20 mutants of *Saccharomyces cerevisiae* is independent of Ndc10p and kinetochore function but requires a subset of spindle checkpoint genes. *Genetics* *148*, 1701-1713.
- Tenenbaum, D. KEGGREST: Client-side REST access to KEGG. R package version 1.8.0.

- Thomas, P.D., Campbell, M.J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A., and Narechania, A. (2003). PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res* 13, 2129-2141.
- Thornton, B.R., and Toczyski, D.P. (2006). Precise destruction: an emerging picture of the APC. *Genes & development* 20, 3069-3078.
- Tieri, P., Grignolio, A., Zaikin, A., Mishto, M., Remondini, D., Castellani, G.C., and Franceschi, C. (2010). Network, degeneracy and bow tie. Integrating paradigms and architectures to grasp the complexity of the immune system. *Theoretical biology & medical modelling* 7, 32.
- Tonks, N.K. (2006). Protein tyrosine phosphatases: from genes, to function, to disease. *Nature reviews Molecular cell biology* 7, 833-846.
- Tonks, N.K. (2013). Protein tyrosine phosphatases--from housekeeping enzymes to master regulators of signal transduction. *FEBS J* 280, 346-378.
- Toyoshima, H., and Hunter, T. (1994). p27, a novel inhibitor of G1 cyclin-Cdk protein kinase activity, is related to p21. *Cell* 78, 67-74.
- Trinkle-Mulcahy, L., Andersen, J., Lam, Y.W., Moorhead, G., Mann, M., and Lamond, A.I. (2006). Repo-Man recruits PP1 gamma to chromatin and is essential for cell viability. *The Journal of cell biology* 172, 679-692.
- Trockenbacher, A., Suckow, V., Foerster, J., Winter, J., Krauss, S., Ropers, H.H., Schneider, R., and Schweiger, S. (2001). MID1, mutated in Opitz syndrome, encodes an ubiquitin ligase that targets phosphatase 2A for degradation. *Nature genetics* 29, 287-294.
- Trost, B., and Kusalik, A. (2011). Computational prediction of eukaryotic phosphorylation sites. *Bioinformatics* 27, 2927-2935.
- Trusina, A., Maslov, S., Minnhagen, P., and Sneppen, K. (2004). Hierarchy measures in complex networks. *Physical review letters* 92, 178702.
- Tyers, M., and Futcher, B. (1993). Far1 and Fus3 link the mating pheromone signal transduction pathway to three G1-phase Cdc28 kinase complexes. *Molecular and cellular biology* 13, 5659-5669.
- Ubersax, J.A., and Ferrell, J.E., Jr. (2007). Mechanisms of specificity in protein phosphorylation. *Nature reviews Molecular cell biology* 8, 530-541.
- Ubersax, J.A., Woodbury, E.L., Quang, P.N., Paraz, M., Blethrow, J.D., Shah, K., Shokat, K.M., and Morgan, D.O. (2003). Targets of the cyclin-dependent kinase Cdk1. *Nature* 425, 859-864.
- Van Roey, K., Gibson, T.J., and Davey, N.E. (2012). Motif switches: decision-making in cell regulation. *Curr Opin Struct Biol* 22, 378-385.
- Venancio, T.M., Balaji, S., Iyer, L.M., and Aravind, L. (2009). Reconstructing the ubiquitin network: cross-talk with other systems and identification of novel functions. *Genome Biol* 10, R33.
- Venne, A.S., Kollipara, L., and Zahedi, R.P. (2014). The next level of complexity: crosstalk of posttranslational modifications. *Proteomics* 14, 513-524.
- Vidanes, G.M., Sweeney, F.D., Galicia, S., Cheung, S., Doyle, J.P., Durocher, D., and Toczyski, D.P. (2010). CDC5 inhibits the hyperphosphorylation of the checkpoint kinase Rad53, leading to checkpoint adaptation. *PLoS biology* 8, e1000286.
- Vinh, D.B., Kern, J.W., Hancock, W.O., Howard, J., and Davis, T.N. (2002). Reconstitution and characterization of budding yeast gamma-tubulin complex. *Molecular biology of the cell* 13, 1144-1157.
- Virshup, D.M., and Shenolikar, S. (2009). From promiscuity to precision: protein phosphatases get a makeover. *Molecular cell* 33, 537-545.
- Visintin, R., Craig, K., Hwang, E.S., Prinz, S., Tyers, M., and Amon, A. (1998). The phosphatase Cdc14 triggers mitotic exit by reversal of Cdk-dependent phosphorylation. *Molecular cell* 2, 709-718.
- Wang, L., Hou, L., Qian, M., and Deng, M. (2012a). Integrating phosphorylation network with transcriptional network reveals novel functional relationships. *PloS one* 7, e33160.

- Wang, M., Weiss, M., Simonovic, M., Haertinger, G., Schrimpf, S.P., Hengartner, M.O., and von Mering, C. (2012b). PaxDb, a database of protein abundance averages across all three domains of life. *Molecular & cellular proteomics : MCP* 11, 492-500.
- Wasilko, D.J., Lee, S.E., Stutzman-Engwall, K.J., Reitz, B.A., Emmons, T.L., Mathis, K.J., Bienkowski, M.J., Tomasselli, A.G., and Fischer, H.D. (2009). The titerless infected-cells preservation and scale-up (TIPS) method for large-scale production of NO-sensitive human soluble guanylate cyclase (sGC) from insect cells infected with recombinant baculovirus. *Protein Expr Purif* 65, 122-132.
- Wasserman, S., and Faust, C. (1994). Social Network Analysis: Methods and Applications (Cambridge University Press).
- Watanabe, N., Arai, H., Nishihara, Y., Taniguchi, M., Watanabe, N., Hunter, T., and Osada, H. (2004). M-phase kinases induce phospho-dependent ubiquitination of somatic Wee1 by SCF β -TrCP. *Proceedings of the National Academy of Sciences of the United States of America* 101, 4419-4424.
- Watts, D.J., and Strogatz, S.H. (1998). Collective dynamics of 'small-world' networks. *Nature* 393, 440-442.
- Whinston, E., Omerza, G., Singh, A., Tio, C.W., and Winter, E. (2013). Activation of the Smk1 mitogen-activated protein kinase by developmentally regulated autophosphorylation. *Molecular and cellular biology* 33, 688-700.
- Whitacre, J.M. (2012). Biological robustness: paradigms, mechanisms, and systems principles. *Frontiers in genetics* 3, 67.
- Winey, M., and Bloom, K. (2012). Mitotic spindle form and function. *Genetics* 190, 1197-1224.
- Winterbach, W., P.V., M., Reinders, M., Wang, H., and Ridder, D. (2013). Local topological signatures for network-based prediction of biological function. In *Pattern recognition in bioinformatics*, P.P.a.M.W. S. Istrail, ed. (London New York: Springer), pp. 23-34.
- Wu, J.Q., Guo, J.Y., Tang, W., Yang, C.S., Freel, C.D., Chen, C., Nairn, A.C., and Kornbluth, S. (2009). PP1-mediated dephosphorylation of phosphoproteins at mitotic exit is controlled by inhibitor-1 and PP1 phosphorylation. *Nature cell biology* 11, 644-651.
- Wu, R., Haas, W., Dephoure, N., Huttlin, E.L., Zhai, B., Sowa, M.E., and Gygi, S.P. (2011). A large-scale method to measure absolute protein phosphorylation stoichiometries. *Nat Methods* 8, 677-683.
- Wu, X., Jiang, R., Zhang, M.Q., and Li, S. (2008). Network-based global inference of human disease genes. *Molecular systems biology* 4, 189.
- Yachie, N., Saito, R., Sugiyama, N., Tomita, M., and Ishihama, Y. (2011). Integrative features of the yeast phosphoproteome and protein-protein interaction map. *PLoS Comput Biol* 7, e1001064.
- Yi, T.M., Kitano, H., and Simon, M.I. (2003). A quantitative characterization of the yeast heterotrimeric G protein cycle. *Proceedings of the National Academy of Sciences of the United States of America* 100, 10764-10769.
- Yoshimi, M., Goyama, S., Kawazu, M., Nakagawa, M., Ichikawa, M., Imai, Y., Kumano, K., Asai, T., Mulloy, J.C., Kraft, A.S., et al. (2012). Multiple phosphorylation sites are important for RUNX1 activity in early hematopoiesis and T-cell differentiation. *European journal of immunology* 42, 1044-1050.
- Yu, H., and Gerstein, M. (2006). Genomic analysis of the hierarchical structure of regulatory networks. *Proceedings of the National Academy of Sciences of the United States of America* 103, 14724-14731.
- Yun, C.Y., and Fu, X.D. (2000). Conserved SR protein kinase functions in nuclear import and its action is counteracted by arginine methylation in *Saccharomyces cerevisiae*. *The Journal of cell biology* 150, 707-718.
- Zhu, H., Klemic, J.F., Chang, S., Bertone, P., Casamayor, A., Klemic, K.G., Smith, D., Gerstein, M., Reed, M.A., and Snyder, M. (2000). Analysis of yeast protein kinases using protein chips. *Nature genetics* 26, 283-289.

Zhu, X., Gerstein, M., and Snyder, M. (2007). Getting connected: analysis and principles of biological networks. *Genes & development* 21, 1010-1024.

Appendix 1. Methods, tables and supplementary figures of chapter 2

Supplementary materials

Gene essentiality

The list of essential genes was downloaded from the *Saccharomyces* gene database (SGD) as of Aug. 2015 (SGD Project, Aug. 2015). A gene was considered to be essential if it has an inviable phenotype on its deletion, that is the ‘phenotype’ and ‘mutant_type’ columns in the downloaded table are equal to ‘inviable’ and ‘null’, respectively.

Pathway-shared components

A pathway-shared component is a KP in the KP-Net that participates in more than one pathway. Pathways were accessed from the Bioconductor KEGGREST package and queried using the R scripting language (Huber et al., 2015; R Core Team, 2012; Tenenbaum).

mRNA synthesis and degradation rate

mRNA synthesis and degradation rates were obtained from a large scale study that used metabolic labelling to quantify total and newly transcribed mRNAs at 41 time points in *Saccharomyces cerevisiae* cells that were synchronized at G1 phase using the alpha factor (Eser et al., 2014). Absolute mRNA synthesis and degradation rates were estimated by this study using the comparative dynamic transcriptome analysis (cDTA) (Sun et al., 2012).

mRNA and protein abundance

mRNA expression profiles were taken from a HTP study using a “generalized” adaptor-tagged competitive PCR (GATC-PCR) between genomic DNA and complementary DNA (Miura et al., 2008). Protein abundances were obtained from the PaxDB database in particle per million (ppm) unit (Wang et al., 2012b). To convert these abundances into mol/cell, we fit a linear regression model between protein abundance from Ghaemmaghami (mol/cel) and integrated abundance from PaxDB (ppm). This resulted in the following equation:

$$\log_{10}(\text{ppm_abund}) = \log_{10}(\text{mol_cel_abund}) - 1.9858.$$

To get protein abundances in mol/cel unit, we multiplied the ppm integrated abundances by 96.78.

mRNA translation rates

mRNA translation rates were taken from a genome scale study that provides a profile of ribosome association with mRNAs in rapidly growing budding yeast cells. mRNAs were separated using velocity sedimentation on a sucrose gradient and 14 fractions across this gradient were collected and quantified using DNA microarrays (Arava et al., 2003).

Protein half-lives

Protein half-lives were obtained from a large scale study that used western blot analysis to assess the time that a TAP-tagged protein takes to be degraded after blocking protein synthesis by cycloheximide (Belle et al., 2006).

mRNA and protein noise

mRNA noise was obtained from a previous study assessing the probability that a messenger contains a TATA-box consensus sequence (Basehoar et al., 2004). The protein noise was taken from the Newman et al. study (Newman et al., 2006) in which the authors measured noise as the coefficient of variation (CV) of protein abundance (the ratio of the protein standard deviation to its mean abundance) in a population of *Saccharomyces cerevisiae* cells growing under a stress condition (in a minimal (SD) media). For every protein, DM represents the difference between the noise value of that protein and a running median of proteins noise. Lower DM values represent less noise in protein abundance.

Protein sequence of KPs

Protein sequence of KPs were retrieved from the SGD database as of Nov 2014 and analysed using the seqinR R library (Charif and Lobry, 2007; SGD Project, Nov. 2014).

Phosphosites in KPs

Phosphorylation sites in KPs were downloaded from the PhosphoGRID database as of Dec. 2014 (Stark et al., 2010). Only phosphorylation sites that were validated by at least two independent studies were retained.

Scaffolds

Scaffolds were annotated from the SGD database (SGD Project, Aug. 2015).

Supplementary methods

Bioinformatics analysis

All analysis were performed using in-house written scripts using the R project for statistical computing environment unless it was indicated otherwise (R Core Team, 2012).

The VertexSort R package

An R package called the VertexSort was developed in this study to permit the application of the VS algorithm on other networks (The source code with the user manual are provided as “VertexSort R package.zip” on CD-ROM).

The pipeline used to assemble the KP-Net

Approximately 75 % of the kinase-protein interactions of the KP-Net were collected from the KID database (Sharifpoor et al., 2011), a kinase specialized database that annotates and scores kinase-protein interactions validated by a broad number of low and large scale experimental methods. Despite the high quality of the KID data, many kinase-protein interactions were reported in literature and missing or not fully annotated in the KID database. Thus, we assembled these interactions as well as phosphatase-protein interactions from BioGRID, PhosphoGrid, Uniprot, two literature curation efforts and using the Textpresso and CiteXplore text mining tools (Ba and Moses, 2010; Fiedler et al., 2009; Magrane and Consortium, 2011; Muller et al., 2004; Stark et al., 2006; Stark et al., 2010) (<http://www.ebi.ac.uk/citexplore>). We annotated and scored the collected Kinase-protein and phosphatase-protein interactions according to the KID database annotation pipeline (Sharifpoor et al., 2011) with adapting the experimental methods to take into account the dephosphorylation interactions representing the reverse of phosphorylation interactions. For instance, we replaced the experimental method “*in vivo* reduction in phosphopeptide detected by mass spectrometry on mutating the kinase” by “*in vivo* increase in phosphopeptide detected by mass spectrometry on mutating the phosphatase”. Moreover, we added the experimental method “trapping phospho-substrate by dead phosphatase catalytic domain” to the

KID validation experimental methods and we associated to it the lowest score (1.2) associated to experimental methods considered as biochemical experiment showing a phosphorylation interaction. These biochemical experiments are marked with a star on top of the column headers in Table S3. Then, in order to collect bona fide direct PDIs, we selected interactions having a confidence score ≥ 4.52 (corresponding to a $P \leq 5 \times 10^{-2}$) and validated by at least one HTP or LTP biochemical experiment showing the occurrence of a PDI event.

Inference of the KP-Net degree distribution and finding the function that best fitted the inferred distribution

We tested whether the degree distribution of the KP-Net nodes follows a power law distribution by using the procedure suggested in (Clauset et al., 2009). Briefly, the scaling parameter (alpha) and the lower bound on the scaling region (x_{\min}) of the power law distribution were estimated using the maximum likelihood method. Then, a goodness-of-fit test was performed between a large number (5,000) of datasets generated by the theoretical power law distribution defined by the estimated parameters and the observed data using the Kolmogorov-Smirnov test. The P -value of the goodness-of-fit test represents the percentage of number of times where the Kolmogorov-Smirnov test rejected the suggestion that the power law distribution is a plausible hypothesis for the observed data. The same pipeline was used to test whether the empirical data follows an exponential or a normal distribution. The analyses were done using the R code developed by (Lai and Aldous, 2012). Our results show that the empirical data fits: (i), a power law distribution defined by an alpha = 2.58 and $x_{\min} = 9$ with a goodness-of-fit test $P = 1.3 \times 10^{-2}$; (ii), an exponential distribution defined by a lambda = 0.05 with a goodness-of-fit test $P = 7.8 \times 10^{-1}$; and (iii), a normal distribution defined by a mean of 2.3 and standard deviation of 0.5 with a goodness-of-fit test $P = 1$. In summary, the degree distribution of the KP-Net best fits a power law distribution or in other words is said scale free (Figure S1).

Testing whether random networks have a bow tie structure on sorting them by the VS algorithm

We applied the VS algorithm on 1,000 random networks generated from the KP-Net by DPR and identified the three layers in each random network. We then tested whether each of these networks possesses a bow tie structure or in other words whether its core layer has fewer nodes

than its top and bottom layers. The *P*-value of this test is the fraction of random networks in which the number of nodes in the core layer is smaller than that of top and bottom layers.

Assessing enriched/depleted GO terms associated with KPs in each layer of the KP-Net

This analysis was performed using the conditional hypergeometric test implemented in the Bioconductor GOStats library of the R project for statistical computing environment (Falcon and Gentleman, 2007; Huber et al., 2015; R Core Team, 2012). It consists of testing whether the number of KPs associated to a given GO term is larger/lower than expected by chance in a particular layer compared to the number of KPs in the KP-Net. The conditional hypergeometric test is characterized by taking into consideration the hierarchical structure of the GO terms tree which causes ancestors to inherit their children GO terms, hence limiting the list of over- and under-represented GO terms to the most specific ones.

Features of in- and out-degrees of nodes in top and bottom layers in random networks sorted by the VS algorithm

We applied the VS algorithm on 1000 random networks generated from the KP-Net by DNPR and identified the three layers in each random network. We then tested the occurrence of two features: (i), whether the in-degree of nodes in the top layer is smaller than that of nodes in core and bottom layers; and (ii), whether the out-degree of nodes in the bottom layer is smaller than that of nodes in top and core layers. The *P*-value of each test is equal to the fraction of random networks having the corresponding tested feature. The first test generated a *P* of 10^{-3} , showing that the VS algorithm classifies nodes having the smallest in-degrees in the top layer, and the second test generated a *P* of 0.7, showing that the VS algorithm does not necessarily classify nodes having the smallest out-degrees in the bottom layer.

Predicting disordered regions in KPs

We used the IUPred algorithm with the long mode option to predict disordered regions in KPs (Dosztanyi et al., 2005). This prediction is based on amino acid interactions energy, as amino acids in disordered regions cannot establish sufficient stabilizing interactions. Each amino acid of KP sequences was associated an IUPred score, indicating whether it occurs in disordered

regions or not. Scores greater than or equal to 0.5 are associated to amino acids predicted to be in disordered regions. Disordered regions were defined as sequences composed of amino acids having an IUPred score above 0.5 as used in (Dosztanyi et al., 2005).

Predicting linear binding motifs in disordered regions in KPs

We used the ANCHOR algorithm to predict linear binding motifs in disordered regions in KPs (Meszaros et al., 2009). This prediction is also based on amino acid interactions energy, such that amino acids in binding motifs residing in disordered regions are not only unable to establish sufficient stabilizing intra-chain interactions, but are also able to gain stabilizing energy on interacting with a globular domain of a protein partner. Each amino acid of KP sequences was associated an ANCHOR score. Scores above 0.5 are associated to amino acids predicted to be in binding motifs in disordered regions. Putative binding motifs were defined as sequences composed of at least 6 consecutive amino acids having an ANCHOR score above 0.5 as used in (Meszaros et al., 2009).

Investigating whether properties characterizing the KP-Net layers are retained in noisy networks

Since KP-Net interactions are of high confidence, we limited our analysis to investigate whether the properties characterizing the KP-Net layers were retained in noisy networks. On adding 200 edges (~18 % of the KP-Net edges) to the KP-Net, most of the topological (in-degree, out-degree and bottlenecks) and 6 out of 18 biological properties (phosphosites, molecular switches, scaffold-associated KPs, mRNA synthesis rate, mRNA noise and protein degradation) were retained in the noisy networks (Figures S9a, S9b, S9d, S9j, S9l, S9m, S9o, S9r and S9t), 1 out of 4 topological properties and 11 out of 18 biological properties were mildly retained (hubs, pathway-shared components, phosphatases, intrinsically disordered regions, LBMs, number of phosphosites in LBMs, cellular localizations, mRNA abundance, mRNA translation rate, protein abundance and protein noise, Figures S9c, S9e, S9g, S9h, S9i, S9k, S9n, S9q, S9s, S9u and S9v) and only two biological properties were not retained (essential genes and mRNA degradation, Figures S9f and S9p) in the noisy networks. In summary, properties characterizing the KP-Net were retained to different levels in the noisy networks, showing that the characteristics of the KP-Net elucidated in this study represent the best of our knowledge to date about KP-Nets.

Supplementary tables

Table A1.I. Annotation of dephosphorylation interactions.

The number of dephosphorylation interactions that were annotated in different public databases and in this study.

Database	Dephosphorylation interactions (#) as to March-2012	Dephosphorylation interactions (#) as to June-2016
Biogrid (Stark et al., 2006)	54	86
IntAct (Kerrien et al., 2012)	6	8
BIND*/BOND (Bader et al., 2003)	6	-*
MINT (Licata et al., 2012)	2	2
MIIPS (Mewes et al., 2002)	1	1
DIP (Salwinski et al., 2004)	0	0
This study	0	169

* The BIND database was combined with the BOND database. Data is not accessible (technical problems with the BOND database server).

Table A1.II. K-Nets and KP-Nets studies.

The list of HTP studies that were interested in K-Nets and KP-Nets, in characteristics of these networks and in the biological properties that were investigated in these studies.

Study	Network type	Network characteristics				Investigated biological properties																					
		Causality	Directionality	In vivo interactions	Includes phosphatases	GO terms	Phenotype	Phosphorylation consensus motifs	Logic motifs	Disordered regions	Phosphosites	Evolution rank	Pathways	Essentiality	Conservation	Physical interaction	Genetic interaction	mRNA abundance	Protein abundance	mRNA degradation	Protein degradation	Protein noise	Linear binding motifs	Molecular switches	Interaction with scaffolds	mRNA synthesis rate	mRNA translation rate
Ptacek et al. 2005 (2)	Phosphorylation network	1	1	0	0	1	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Fiedler et al. 2009 (3)	Genetic interaction network	0	0	1	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Bodenmiller et al. 2010 (4)	Genetic deletion network	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Breitkreutz et al. 2010 (5)	Protein-protein interaction network	0	0	1	1	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Bhardwaj et al. 2010 (6)	Co-Phosphorylation network	1	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sharifpoor et al. 2011 (7)	Kinase interaction network	0	0	0/1	0	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
Cheng et al. 2015 (1)	Phosphorylation network	1	1	0	0	1	0	1	0	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0
This study	Phosphorylation network	1	1	1	1	1	0	1	1	1	1	0	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1

Table A1.III. Phosphorylation and dephosphorylation interactions used to assemble the KP-Net (Supplementary file on CD-ROM)

The PDIs that were included in the KP-Net and all the experimental methods that validated them with the pubmed references of the articles in which they were reported. The KID database pipeline was used to score and annotate these interactions.

Table A1.IV. Core layer KPs implicated in decision-making.

KPs in the core layer that are implicated in decision-making processes and the list of these processes.

KP gene name	Biological role
Cdc14	<ul style="list-style-type: none"> - Promotes the mitotic exit by counteracting mitotic Cdc28 activity (Visintin et al., 1998). - Regulates various microtubule-related processes at the metaphase/anaphase transition (Chirolí et al., 2009).
Cdc28	<ul style="list-style-type: none"> - Drives the cell cycle in the budding yeast (Enserink and Kolodner, 2010). - Prevents catastrophic M-phase progression after DNA replication arrest (Enserink et al., 2009). - Arrests the spindle assembly checkpoint on spindle damage (Kitazono et al., 2003).
Cdc5	Attenuates the DNA damage checkpoint enabling cells to adapt to this damage (Vidanes et al., 2010).
Cdc7	Promotes the replication checkpoint (Ogi et al., 2008).
Elm1	Regulates spindle position checkpoint (Caydasi et al., 2010).
Gin4 and Hsl1	Link entry into mitosis to proper septin organization (Barral et al., 1999).
Hog1	Coordinates replication and transcription under osmotic stress (Duch et al., 2013).
Mih1	Promotes anaphase onset (Liang et al., 2013).
Rad53	Is involved in the three DNA damage checkpoints (G1/S transition, during the S phase and before mitosis) (Foiani et al., 2000).
Swe1	Enables the morphogenesis checkpoint to delay nuclear division in cells that have not budded (King et al., 2013).

Table A1.V. The kinase-substrate interactions that were predicted to be implicated in osmotic shock in this study (Supplementary file on CD-ROM)

Interactions relating each protein containing dynamic phosphosites on high osmotic shock as defined in Kanshin et al. to their potential kinases (Kanshin et al., 2015).

Supplementary figures

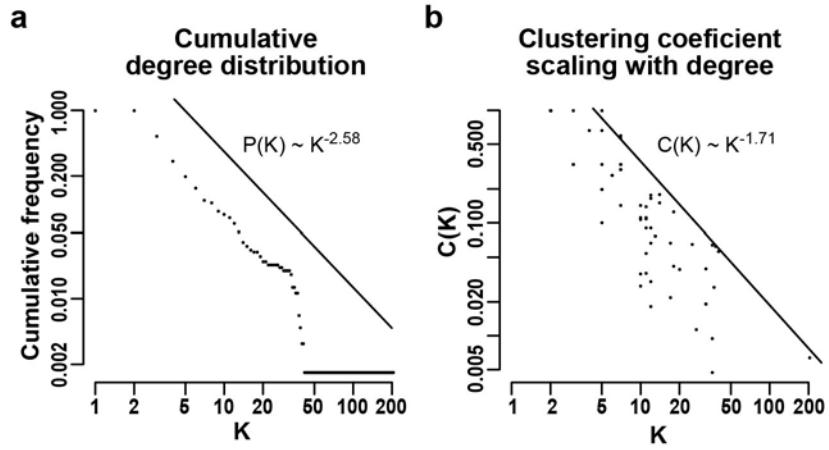


Figure A1.1. Distributions of the cumulative degree and of the clustering coefficient of KP-Net nodes follow a power law distribution.

(a) The cumulative distribution of KP degrees (K) of the KP-Net plotted with logarithmically scaled axis showing that K follows a power law distribution: $P(K) \sim K^{-2.58}$. (b) The scaling of KP clustering coefficients ($C(K)$) with the corresponding KP degrees (K) in the KP-Net plotted with a logarithmically scaled axis also showing that $C(K)$ follows a power law distribution $C(K) \sim K^{-1.71}$.

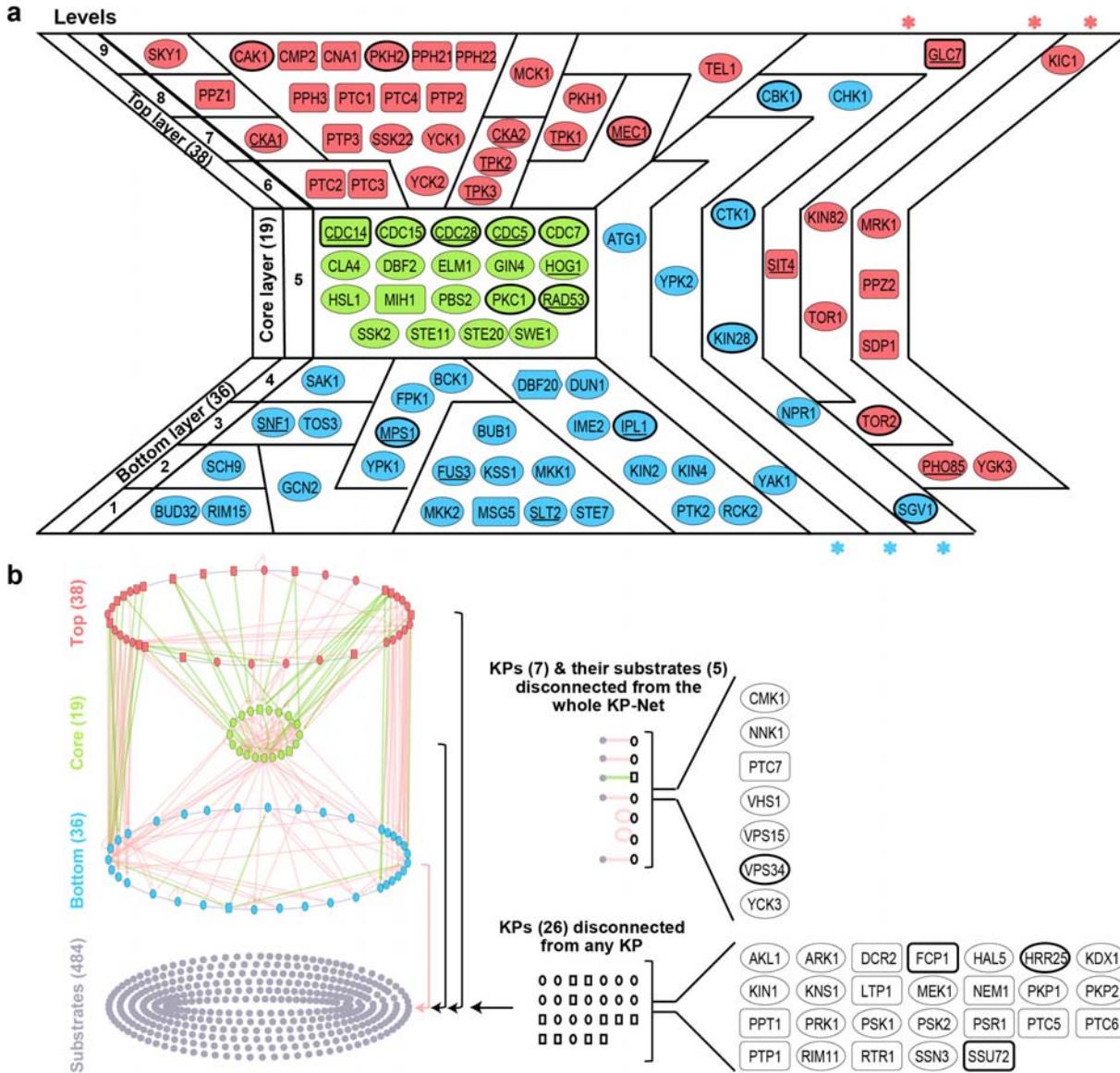


Figure A1.2. Detailed hierarchical structure of the KP-Net.

(a) The detailed hierarchical structure of the KP-Net. The classification of each KP in each level or range of levels after applying the VS algorithm on the KP-Net. Kinases are represented by an ellipse and phosphatases by a rectangle. Eleven and 9 nodes spanning the three layers are classified in top and bottom layers, respectively, because they are not regulated by and do not regulate any KP, respectively (nodes in columns with red and blue stars, respectively). Underlined KPs represent hubs and those with a thicker border designate essential KPs. (b) The relative position of the excluded KPs to the KP-Net hierarchy. The KP-Net hierarchy is shown at the left side of the figure and nodes that were excluded from this study analysis are shown with their identity at the right side of the figure. Seven KPs and their substrates (five proteins that are not KPs) were excluded from KP-Net analyses, because they are disconnected from the whole KP-Net. Twenty six KPs were also excluded from KP-Net analyses, because they do not regulate any KP and they are not regulated by any KP. Numbers between parentheses indicate the number of nodes and arrows represent directed interactions (red: phosphorylation, green: dephosphorylation and black: both)

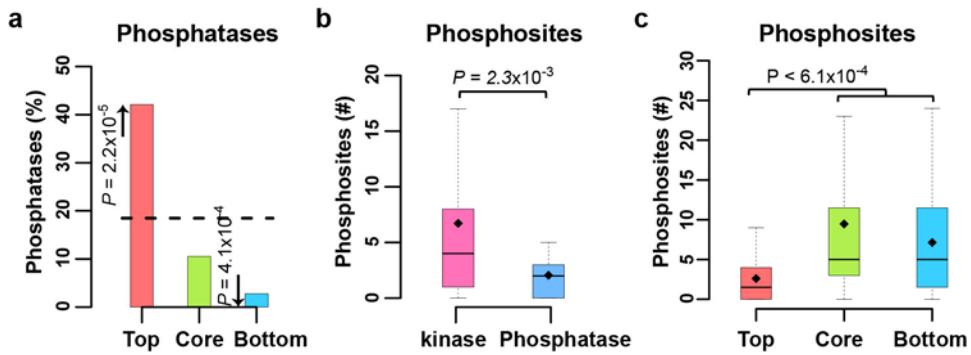


Figure A1.3. Percentage of phosphatases in each of the three layers of the KP-Net and distribution of phosphosites among KPs and among the different layers of the KP-Net.

(a) Percentage of phosphatases in the different layers of the KP-Net. Distribution of phosphosites among **(b)** sequences of KPs belonging to the KP-Net and among **(c)** sequences of KPs in the different layers of the KP-Net. Phosphosites validated by at least two studies were taken from the PhosphoGRID database. The broken line in the bar plot represents the expected mean of phosphatases percentage in each layer. Black diamonds in box plots designate the average of the number of phosphosites in KPs and in each layer, respectively. Outliers were omitted from boxplots to simplify data representation. *P*-values were calculated by comparing property means of two layers and the enrichment/depletion of a property within a layer using the randomization test (RT, Methods) and the hypergeometric test (HT), respectively. For description of the used datasets see Supplementary materials..

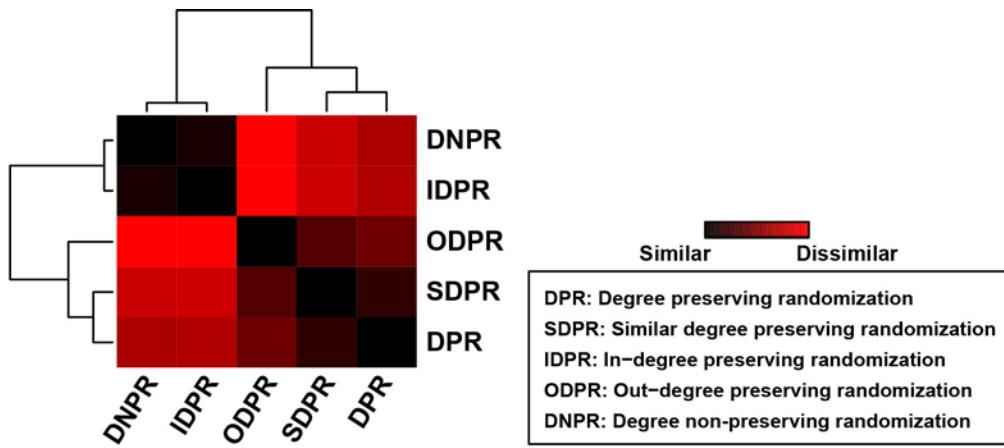


Figure A1.4. Clustering of the different sets of randomized networks.

One thousand random networks were generated using each of the following randomization methods: degree preserving randomization (DPR, Methods), similar degree preserving randomization (SDPR, Methods), in-degree preserving randomization (IDPR, Methods), out-degree preserving randomization (ODPR, Methods) and degree non-preserving randomization (DNPR, Methods). Random networks were sorted using the VS algorithm. The Euclidean distance was then used to calculate dissimilarities between the different sets of random networks in terms of the studied properties of KPs in each layer of these random networks. Agglomerative clustering was performed using the average linkage criteria.

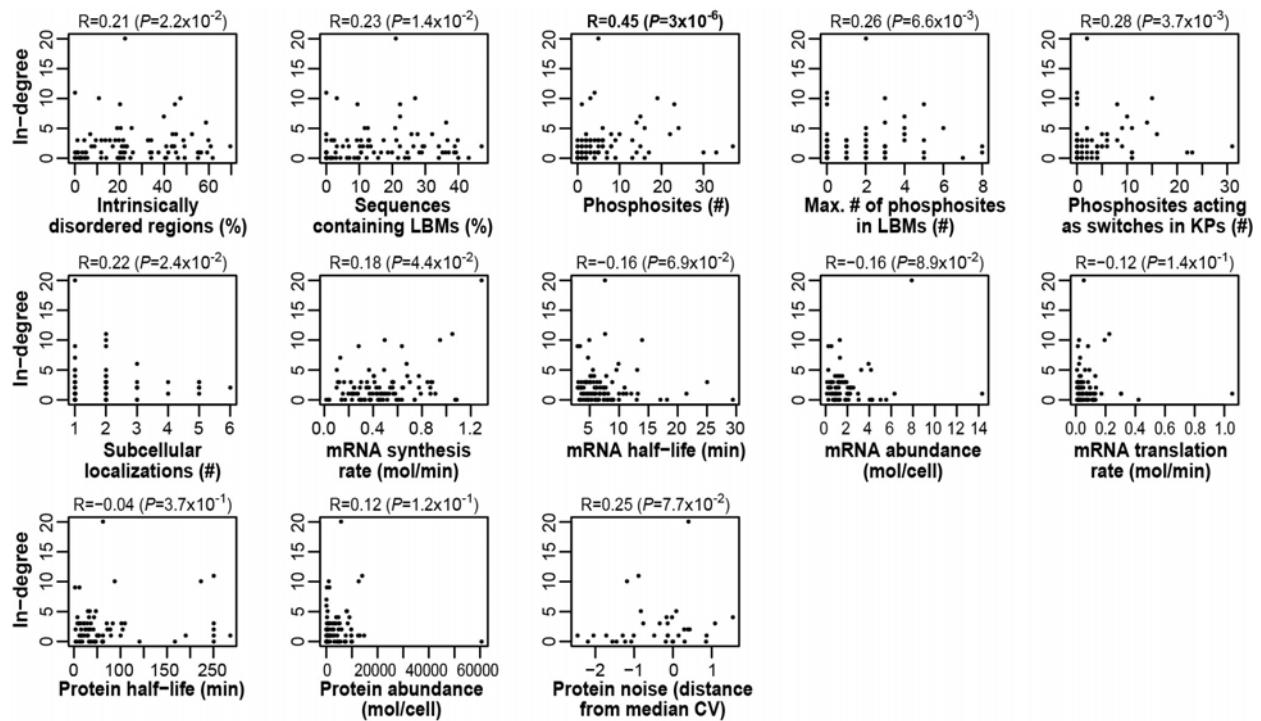


Figure A1.5. Correlation between KP in-degrees and KP numerical properties.

Correlations were calculated using the Spearman correlation coefficient and *P*-values of these correlations are reported between parentheses. Values in bold indicate the presence of moderate correlations (between 0.3 and 0.6) with a *P*-value < 0.05.

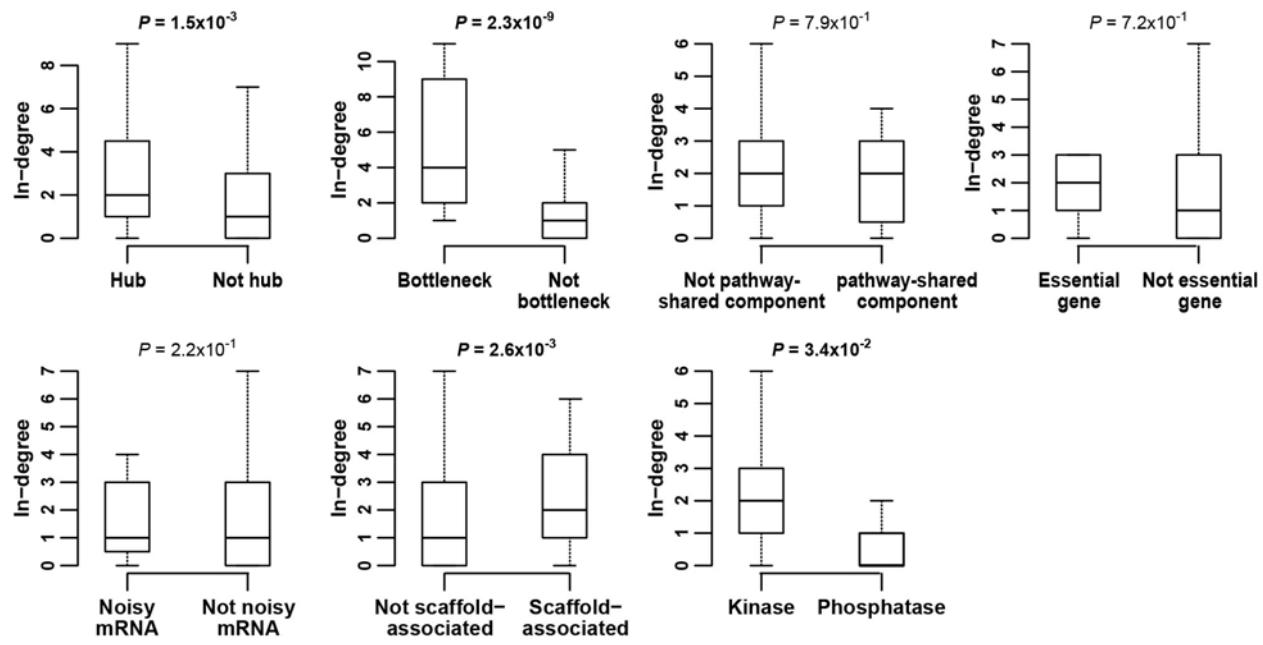


Figure A1.6. In-degree distribution of KPs characterized and not characterized by each of the studied categorical properties.

The studied categorical properties are: hub, bottleneck, pathway-shared component, essential gene, noisy mRNA, scaffold-associated KP and type of enzyme (KP). *P*-value of the significance of the difference between in-degree means of KPs characterized by and not characterized by each of these properties were calculated using the one-way ANOVA test. Values in bold indicate significant *P*-values (< 0.05).

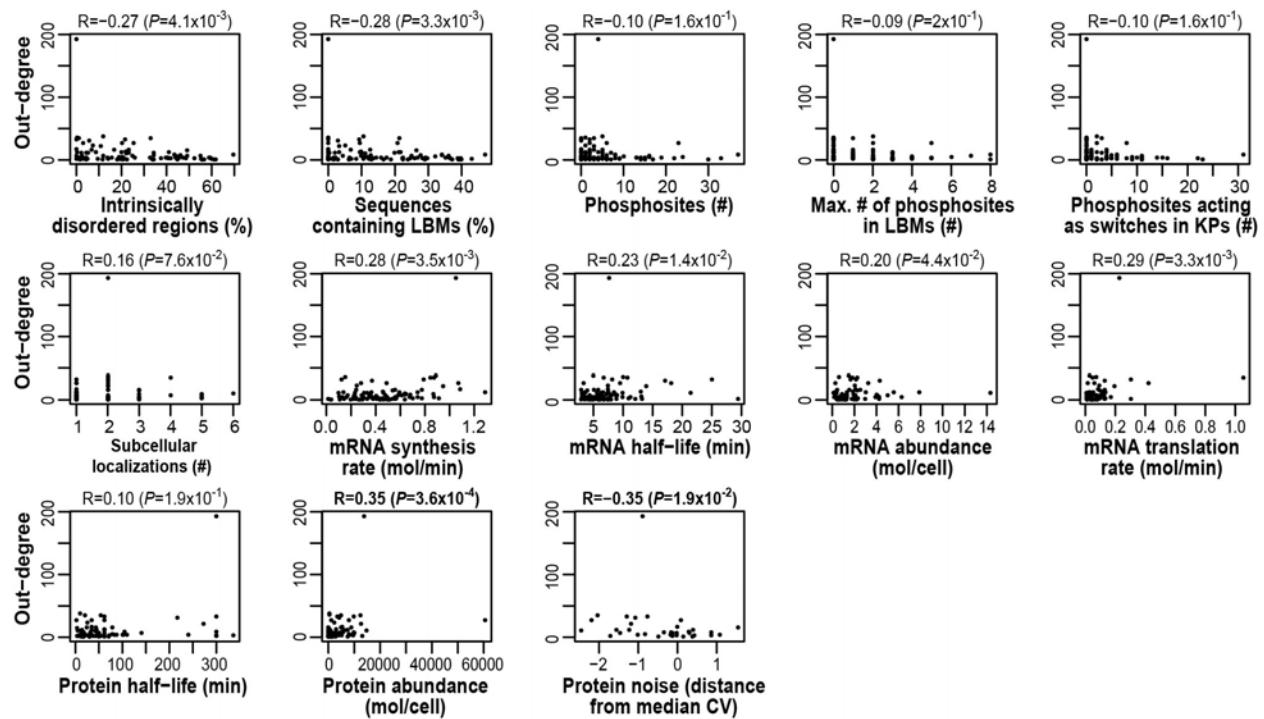


Figure A1.7. Correlation between KP out-degrees and KP numerical properties.

Correlations were calculated using the Spearman correlation coefficient and P -values of these correlations are reported between parentheses. Values in bold indicate the presence of moderate correlations (between 0.3 and 0.6) with a P -value < 0.05 .

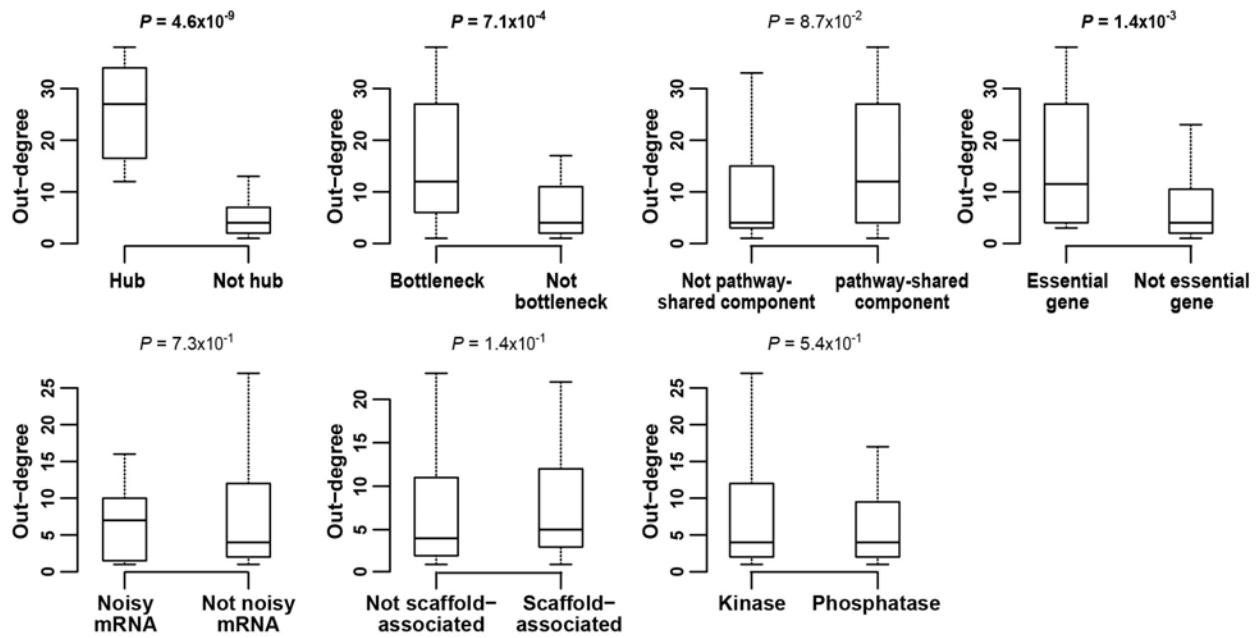


Figure A1.8. Out-degree distribution of KPs characterized and not characterized by each of the studied categorical properties.

The studied categorical properties are: hub, bottleneck, pathway-shared component, essential gene, noisy mRNA, scaffold-associated and type of enzyme (KP). *P*-value of the significance of the difference between out-degree means of KPs characterized by and not characterized by each of these properties were calculated using the one-way ANOVA test. Values in bold indicate significant *P*-values (< 0.05).

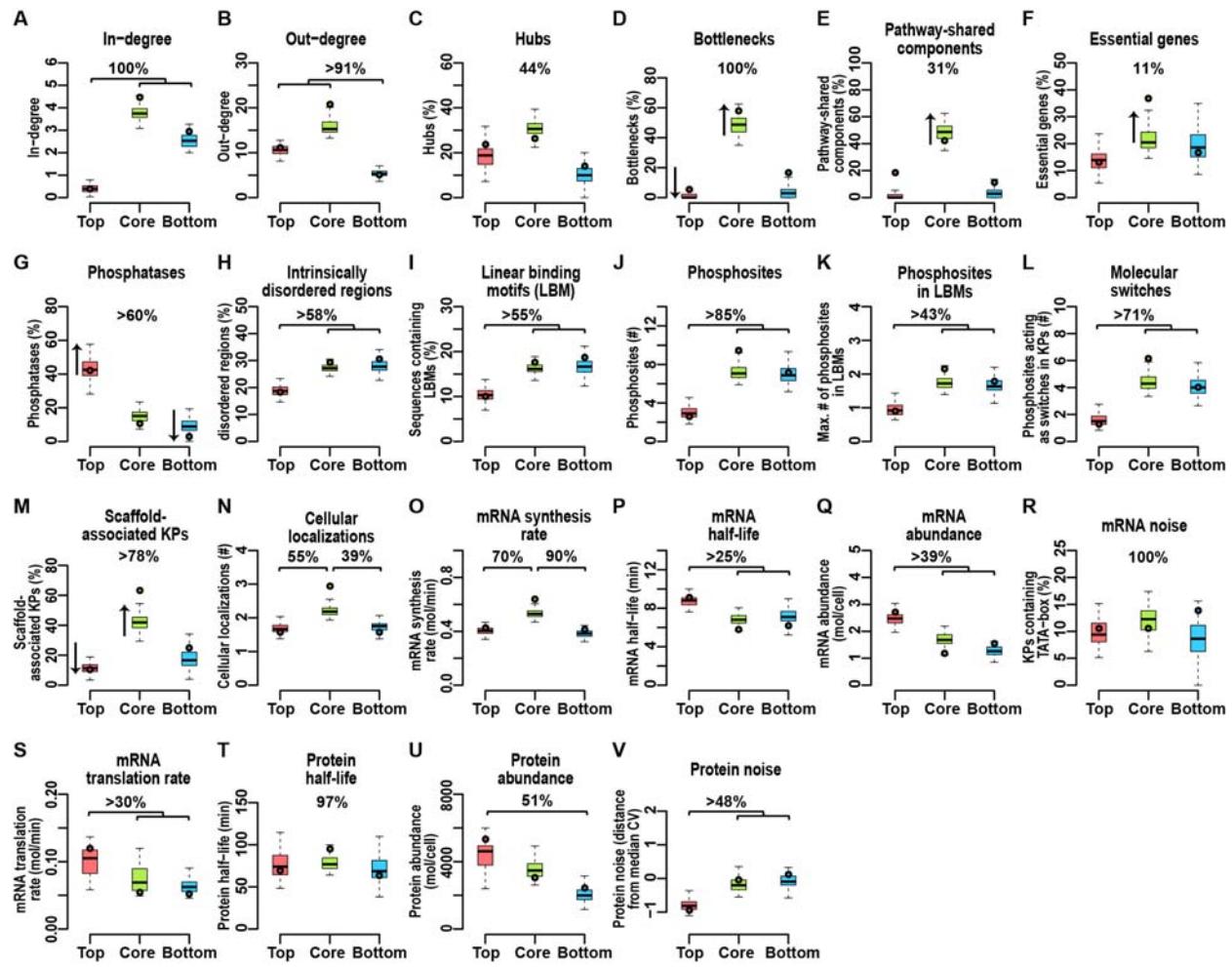


Figure A1.9. Distribution of the different properties of KPs in each layer of 100 noisy networks.

Box plots represent the distribution of the means of KP properties in each layer of the 100 noisy networks that were generated from the KP-Net by randomly adding 200 edges. Coloured circles represent the mean of KP properties of KP-Net layers. Percentages at the top of each box plot represent the fraction of noisy networks that have significant (e.g. out-degree) and non-significant (e.g. hubs) differences among their layers. In other words percentages represent the fraction of noisy networks that their layer properties have the same pattern as layer properties of the KP-Net assembled in this study.

Appendix 2. Supplementary methods, tables and figures of chapter 3

Supplementary methods

Calculation of false-negative rates (FNRs) for the Cdc28 screen

We used the KID database to assess the performance of Cdk1 screen (Sharifpoor et al., 2011). We found 17 Cdc28–protein interactions that were not detected in our screen. However, 7 of these 17 interactions have a KID score less than 4.45, which is the minimum score that corresponds to a $P \leq 0.05$ and that the work by Sharifpoor et al. (Sharifpoor et al., 2011) recommended as reliable. Moreover, the work by Sharifpoor et al. (Sharifpoor et al., 2011) recommended a KID score of 6.29 (P value ≤ 0.01). This cutoff leaves us with only five false negatives (FNs; Figure 3.2b). Taking the different cutoffs, our FNR ranged from 29 % to 41 % Table A2.II. These values compare favorably with other large-scale PPI screens with FNRs in the range of 60 % (Sambourg and Thierry-Mieg, 2010). We note, however, that the rather low FNR that we obtain does not necessarily reflect superior performance of our assay compared with other methods but rather, reflects the limited sampling that we performed.

$$\text{FNR} = FN / (TP + FN).$$

Calculation of sensitivity and specificity of the Cdk1–protein interaction screen in cyclin deletion strain backgrounds

There are limited data available on cyclin binding specificity, but there are a limited number of studies that examined cyclin specificity in great detail: the study by Loog and Morgan (Loog and Morgan, 2005) and recently, a LTP study by Koivomagi et al. (Koivomagi et al., 2011) and a HTP study by Ptacek et al. (Ptacek et al., 2005). Taken together, these studies investigated four of nine cyclins (Cln2, Clb5, Clb3, and Clb2). To compare our results with the results of these three studies (Koivomagi et al., 2011; Loog and Morgan, 2005; Ptacek et al., 2005), we took into consideration only proteins that were tested in our study and each of these three studies [2 proteins were tested in our screen and in (Koivomagi et al., 2011), and 21 proteins were tested in our screen and both in (Loog and Morgan, 2005) and in (Ptacek et al., 2005)]. We estimated

the FNR to be 33 %, sensitivity to be 67 %, and specificity to be 60 % (Table A2.III). As noted for the Cdk1 interactions discussed above, we think that these values likely underestimate false-positive and FNRs, but would not likely increase beyond other methods in a large-scale analysis.

Supplementary tables

Table A2.I. List of prey proteins that interact with Cdk1 identified by OyCD PCA

Cdk1 target (unique)*	No. of full Cdk1 sites [S/T]PX[K/R]		No. of minimal Cdk1 sites [S/T]P	No. of cyclin binding motifs		Cyclin contingency‡	
	RXL	LP [†]		Clb3-6	Cln		
Asf1	0	1	0	0	—	—	—
Bud6	2	10	4	5	Yes	Yes	—
Cct7	0	1	5	5	Yes	Yes	—
Far1	3	15	4	4	No	No	—
Gcn4	0	5	1	3	—	—	—
Gin4	2	11	7	7	No	No	—
Gsy2	0	5	1	3	Yes	Yes	—
Kar9	2	15	9	3	Yes	No	—
Kip2	2	8	3	1	No	No	—
Ksp1	0	10	3	5	—	—	—
Lcb4	5	6	3	3	—	—	—
Lte1	8	20	4	6	Yes	Yes	—
Mbb1	0	0	0	0	—	—	—
Mcm3	5	11	3	7	Yes	Yes	—
Mft1	0	1	4	1	Yes	No	—
Nbp1	3	7	1	0	—	—	—
Net1	3	17	3	9	—	—	—
Paf1	0	2	4	3	—	—	—
Pah1	2	15	6	2	Yes	Yes	—
Pho4	0	7	1	2	Yes	Yes	—
Ptk2	2	16	2	5	Yes	Yes	—
Rad9	9	20	8	3	No	Yes	—
Rim20	0	0	4	5	Yes	Yes	—
Rsp5	0	4	2	9	—	—	—
Sap185	1	8	7	3	—	—	—
Sic1	3	9	4	1	—	—	—
Stb1	5	18	2	2	No	No	—
Swi5	8	20	3	2	Yes	Yes	—
Swi6	2	5	4	2	—	—	—
Tfc7	0	4	0	2	—	—	—
Thp1	0	2	2	3	Yes	Yes	—
Tub4	0	5	1	2	Yes	No	—
Ugp1	0	2	4	2	Yes	Yes	—
Vip1	1	11	12	8	—	—	—
Ygr035c	2	2	0	0	No	No	—
Yng2	0	3	1	1	—	—	—
Ypt11	0	5	3	1	—	—	—

* Unique Cdk1 target identified in this study.

† Located within a region of primarily hydrophobic residues.

‡ Figure 3.1b.

Table A2.II. False-negative rates (FNR) of the Cdk1 screen

The FNRs of the Cdk1 screen were calculated by using a the KID as a gold standard (interactions between Cdk1 and proteins annotated in the KID and interactions between Cdk1 and proteins annotated in the KID having a KID score larger than 4.45 and 6.29)

KID score cutoff	False negative (FN)	True positive (TP)	False negative rate FN / (TP+FN)
None	17	24	41 %
4.45	10	15	40 %
6.29	5	12	29 %

Table A2.III. Sensitivity and specificity of the Cdk1–protein interaction screen in cyclin deletion strain backgrounds

Benchmark	TP	FP	FN	TN	FNR FN / (TP+FN)	Sensitivity TP / (TP+FN)	Specificity TN / (TN+FP)
Loog et al., 2005	2	7	1	11	33 %	67 %	61 %
Ptacek et al., 2005	2	15	0	25	0 %	100 %	63 %
Koivomagi et al., 2011	2	2	1	3	33 %	67 %	60 %

Table A2.IV. Cdk1 interaction death index for all prey genes (Supplementary file on CD-ROM).**Table A2.V.** Cyclin contingency of 21 Cdk1 interacting proteins (Supplementary file on CD-ROM).

Supplementary figures



Figure A2.1. Optimized yeast *Saccharomyces cerevisiae* prodrug-converting enzyme cytosine deaminase (OyCD) protein fragment complementation assay cyclin-dependent kinase (Cdk)-prey interaction assays in nine cyclin deletion strains.

(Figure A2.1 continued) The interactions of Cdk1 and its interacting partners were tested in a yeast strain expressing all of the nine cyclins (WT) and strains with one of nine cyclin and *FCY1* genes deleted. The cyclin deletion strains are labeled by the gene name followed by Δ . For example, the *CLN1* Δ deletion strain is named *cln1* Δ . The positive control corresponds to yeast expressing the homodimeric GCN4 coil-coil leucine zipper (Zip:Zip) fused to the OyCD fragments. This interaction is constitutive and independent of any cyclin. The negative control corresponds to yeast expressing only Cdk1 fused to OyCD F[2]. 5-FC, 5-fluorocytosine. $P \leq 0.01$ was used as a cutoff for this experiment. A blue asterisk indicates $P \leq 0.01$.

Appendix 3. Scientific contributions

Publications

2017

Stynen B*, **Abd-Rabbo D***, Kowarzyk J, Miller-Fleming L, Ralser M, and Michnick SW
Drug profiling based on a protein-fragment complementation assay. (In preparation)

Abd-Rabbo D and Michnick SW

Delineating functional principles of a kinase-phosphatase network in the budding yeast
BMC Systems Biology, 2017. 11(1):38-52.

2015

Ear PH, Kowarzyk J, Booth MJ, **Abd-Rabbo D**, Shulist K, Hall C, Vogel J and Michnick SW
Combining the optimized yeast cytosine deaminase protein fragment complementation assay and
an *in Vitro* Cdk1 targeting assay to study the regulation of the γ -tubulin complex.
Methods Mol Biol, 2015. 1342:237-57.

2013

Landry CR, Levy ED, **Abd-Rabbo D**, Tarassov K and Michnick SW
Extracting insight from noisy cellular networks.
Cell, 2013. 155(5):983-9.

Ear PH, Booth MJ, **Abd-Rabbo D**, Kowarzyk Moreno J, Hall C, Chen D, Vogel J and Michnick SW
Dissection of Cdk1–cyclin complexes *in vivo*.
Proc Natl Acad Sci USA, 2013. 110 (39):15716–15721.

N.B. * indicates that authors followed by a star contributed equally in the article.

Conferences and presentations

2014

Abd-Rabbo D and Michnick SW

Functional principles of the Kinase-phosphatase network

Conference Simon-Pierre Noël, Université de Montréal, Montréal, Québec, Canada, February 2014.

2012

Abd-Rabbo D and Michnick SW

Elucidating the hierarchical structure of the kinases and phosphatases regulatory network

- The 13th International Conference on Systems Biology, Toronto, Canada, August 2012.
- Rencontre Louis-Philippe Bouthillier. Mont-Gabriel, Québec, Canada, May 2012.

2011

Abd-Rabbo D and Michnick SW

Elucidating the hierarchical structure of the kinases and phosphatases regulatory network

MonBUG Bioinformatics Symposium, Institut de recherches cliniques de Montréal, Montréal, Québec, Canada, September 2011.