TYROSINE PHOSPHORYLATION IN *TETRAHYMENA THERMOPHILA* DURING INTERCELLULAR COMMUNICATION

ALBANA KUME

A THESIS SUBMITTED TO THE FACULTY OF GRADUATE
STUDIES IN PARTIAL FULFILMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

GRADUATE PROGRAM IN BIOLOGY
YORK UNIVERSITY,
TORONTO, ONTARIO
DECEMBER 2012

TYROSINE PHOSPHORYLATION IN *TETRAHYMENA THERMOPHILA* DURING INTERCELLULAR COMMUNICATION

By **ALBANA KUME**

A thesis submitted to the Faculty of Graduate Studies of York University in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

ABSTRACT

Emerging evidence suggests that tyrosine phosphorylation signaling is present in unicellular organisms. There is also emerging evidence suggesting the presence of tyrosine phosphorylation in the unicellular ciliate protozoan, *Tetrahymena thermophila*. *T. thermophila* shares a high number of orthologs with humans. Understanding tyrosine phosphorylation in unicells could help us understand how this pathway works in multicells. In this investigation, tyrosine phosphorylation signaling is investigated in *T.thermophila* during starvation and conjugation as it could play a role in mating in this organism. This was done by applying Western blot, indirect immunofluorescence microscopy, bioinformatics, and gene knockout. The immunodetection methods combined with anti-phosphotyrosine antibody suggest the presence of tyrosine phosphorylation in starved and conjugating cells. The bioinformatic data also show proteins that could be tyrosine phosphorylated during starvation and conjugation. In addition, for one of these proteins, its gene was prepared for investigation by knocking out its coding region and replacing it with a neo4 cassette to investigate its role and possible relevance to tyrosine phosphorylation signaling. These data combined with previous work suggest that there is tyrosine phosphorylation signaling in *T. thermophila* that occurs during mating and preparation for mating.

ACKNOWLEDGEMENTS

I would like to thank Dr. Ronald Pearlman for his support and the opportunity to study in his laboratory. His constant guidance and patience is greatly appreciated. I would also like to thank Dr. Michael Scheid for his guidance and advice with this research project. This extends to Dr. Jyoti Garg and Dr. Takahiko Akematsu for their instructions and advice throughout.

I also wish to thank Dr. Yasuhiro Fukuda, Rizwan Attiq, Matt Cadorin, Sussana Marquez, Kanwal Ashraf, members of the the Pearlman lab.

Lastly, I would like to thank my family for their help and encouragements. This would have not been possible without them.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVATIONS

ASI2	anlagen stage induced 2
CHAPS	3-((3-Cholamidopropyl)dimethylammonium)-1-propanesulfonate
DABCO	1,4-diazabicyclo (2,2,2) octane
DNA	deoxyribonucleic acid
ECL	electrochemiluminescent
EGF	endothelial growth factor
FGF	fibroblast growth factor
JAK	janus protein tyrosine kinase
MAC	macronucleus
MIC	micronucleus
neo	neomycin resistance gene
Nck1	non-catalytic region of tyrosine kinase 1
NRTK	non receptor tyrosine kinase
PDGF	platelet derived growth factor
PBS	phosphate buffered saline
PK	protein kinase
PSK	protein serine/threonine kinase
PTK	protein tyrosine kinase
PTP	protein tyrosine phosphatases
pTyr	tyrosine phosphorylation
RNA	ribonucleic acid
RNase	ribonuclease
RTK	receptor tyrosine kinase
SDS	sodium dodecyl sulphate
SH2	src homology - 2
SH3	src homology - 3
STAT	signal transducer and activator of transcription
TBS	tris buffered saline
TKL	tyrosine-kinase like
TrkA	neurotrophic tyrosine kinase receptor type 1
UTR	untranslated region
VEGF	vascular endothelial growth factor

1. INTRODUCTION

1.1. *Tetrahymena thermophila* as a model organism

T. thermophila is a unicellular eukaryotic organism. It is a member of the kingdom Protist (animal-like characteristics and a protozoan), superphylum Alveolata and the phylum Ciliophora (fresh water ciliate) (Eisen et al., 2006). This thesis focuses on studying tyrosine phosphorylation signaling in *Tetrahymena* to understand signaling in *Tetrahymena* and the evolution of this pathway which is linked to various cellular events in mammals (e.g. embryonic development and immune system development and function).

Tetrahymena is an excellent animal model system for diverse studies in cell and molecular biology. Large quantities of *T. thermophila* can be easily cultured axenically in either complex or defined medium which allows for in depth biochemical and molecular studies. It can readily grow to 3 to $5\text{x}10^5$ cells/ml in exponential growth and it has a doubling time of 2.5 hours in complex medium. Also, *Tetrahymena* can grow at a wide range of temperatures between 18° C to 41° C. Furthermore, the ciliate takes in nutrition by active transport of small particles and phagocytosis (Orias, 2000). Starvation induces pairing and then conjugation when two different mating types are present together. A successful conjugation is indicated when 85 – 90 % of the cells are paired within 2-3 hours of mixing the cells (Karrer, 2000) and the complete process requires an additional approximately 15 hours for completion. These features, among others, have allowed research to advance in various areas.

Studies with *T. thermophila* have lead to Canada Gairdner International Awards and Nobel prizes awarded for research on catalytic RNA and ribozymes as well as telomeres and telomerase (Kruger et al., 1982; Greider and Blackburn, 1985). Initial studies on identifying a nuclear histone acetyltransferase as a transcription factor leading to ideas about a histone code, were carried out in *T. thermophila* and led to recognition with a Canada Gairdner International Award to Dr. David Allis (Ohba et al., 1999). Other fundamental studies such as the discovery of an RNAi-like pathway and heterochromatin formation in gene silencing and developmentally programmed genome rearrangement, roles for histone H1 function, tubulin modifications, and much more have been developed using *T. thermophila*. Most of these fundamental studies have become important in the development of biomedical and other applications of this fresh water ciliate (Coyne et al., 2012).

The somatic and germinal genomes of *Tetrahymena* have been fully sequenced and annotated, including kinase orthologs shared among selected eukaryotic genomes and *Tetrahymena* (Eisen et al., 2006; Coyne et al., 2008). Furthermore, analysis of the fully sequenced genome of *T. thermophila* has shown that this organism shares more orthologs with humans (2280) than between the yeast *Saccharomyces cerevisiae* and humans (2097) or *Plasmodium falciparum* and *T. thermophila* (1325). However, this doesn't mean that *T. thermophila* and humans are more alike than the species in the sister phyla. *S. cerevisiae* and humans have processes that originate from the common ancestor of animals and fungi. This high number of orthologs between *T. thermophila* and humans is a reflection of gene retention of various ancestral eukaryotic functions in *T. thermophila* with a loss of function in other eukaryotic lineages. Some of the genes that are orthologs with those in *T. thermophila* but not in *S. cerevisiae*, correspond to genes that are linked to human diseases. The fully sequenced genome and other characteristics of *T. thermophila* make this organism an exceptional unicellular animal model which provides great insights into understanding the evolutionary processes between unicellular and multicellular organisms. Furthermore, this could help understand how these processes work in multicellular organisms (Eisen et al., 2006; Liu et al., 2011a).

1.2. *Tetrahymena* conjugation

T. thermophila exhibits nuclear dimorphism. It has two nuclei in a single cell, the macronucleus (MAC) and the micronucleus (MIC). The MIC is diploid and has five pairs of metacentric chromosomes. The MIC is the germline nucleus which stores the genetic information that is passed on to the progeny created by conjugation of cells of two of the seven different mating types. The MAC is the somatic nucleus and its DNA is derived from the MIC. It is also the site of gene expression during vegetative growth. *Tetrahymena* has both sexual reproduction which is induced when cells of different mating type are starved and mixed, and asexual reproduction which occurs when the cells grow vegetatively (Cole et al., 1997; Eisen et al., 2006).

Conjugation involves meiosis and starts when starved cells of two different mating types pair. A temporary junction forms between the two mating types which will allow for exchange of genetic material (Orias et al., 1983). The MIC undergoes meiosis producing four haploid nuclei. Three of these nuclei degrade and the fourth duplicates by mitosis. One of the MICs from each member of the pair will then cross the junction to migrate to the other parent. This migratory MIC fuses with the stationary MIC to produce a diploid zygotic nucleus, containing genetic material from both parents. This is followed by mitosis of the newly formed MIC, generation of a new MAC from one of the new MIC, and degradation of the old MAC by programmed nuclear death (Akematsu and Endoh, 2010). Lastly, when nutrients become available, the cells resume vegetative growth and they grow asexually and divide amitotically to produce identical daughter cells at each division, until the next conjugation (Karrer, 2000; Yao and Chao, 2005).

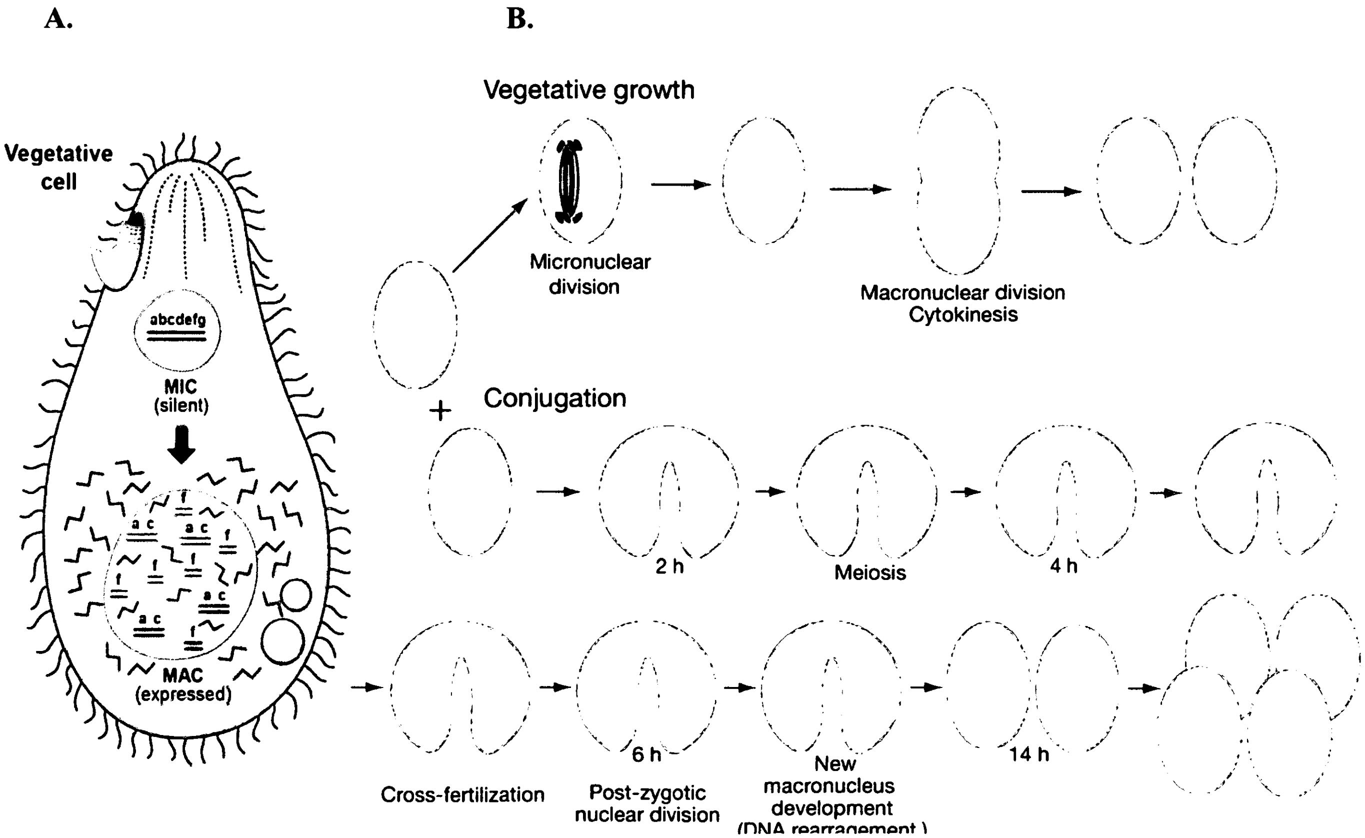


Figure 1. Life cycle of *Tetrahymena*.

A) Vegetative *Tetrahymena* showing MIC as silent (genetic information storage) and expression of genes in MAC which are derived from MIC. B) The top panel shows *Tetrahymena* during vegetative growth when nutrients are plentiful. The panel at the bottom shows conjugation between two *Tetrahymena* mating types after being starved. Diagram from Yao and Chao (2005).

1.3. Protein kinase classification

Development of intercellular communication was an important stage in the evolution of multicellular organisms. Embryonic development, the immune system, and metabolism functions are controlled by complicated signaling networks. These signaling networks are governed by hormones, growth factors and cytokines. The extracellular ligands interacting with the corresponding receptors initiate a cascade of intracellular biochemical events leading to cellular responses (Darnell, 1997; King and Carroll, 2001; Pawson et al., 2001; King et al., 2003; Shiu and Li, 2004). Protein phosphorylation at serine, threonine and/or tyrosine sites is a post-translational modification and occurs as a result of intercellular communication. The proteins that are responsible for these modifications are enzymes known as protein kinases (PKs) (Hubbard and Till, 2000). There are three types of protein kinases: protein serine/threonine kinases (PSKs) and protein tyrosine kinases (PTKs). The third type of kinase phosphorylates at both the serine/threonine and tyrosine sites and are known as dual specificity kinases. PTKs are further separated into non-receptor (cytosolic) tyrosine kinases (NRTKs) and receptor tyrosine kinases (RTKs) (Gu and Gu, 2003). PTKs also facilitate the effects of adhesion molecules, hormones, cytokines and antigens (Pawson, 2004).

1.3.1. Receptor Tyrosine kinases

The RTKs are composed of an extracellular domain which binds polypeptide ligands, transmembrane helix and the cytoplasmic domain which has tyrosine kinase catalytic activity (Figure 2). Most RTKs exist as monomers with a single polypeptide chain when the ligand is absent. The tyrosine kinase catalytic domain is located in the cytoplasm and the extracellular region contains a wide array of globular domains such as EGF-like domains and Ig-like domains (Li et al., 2003).

Ligand binding to RTKs induces receptor oligomerization which initiates tyrosine autophosphorylation of the catalytic domains, leading to the recruitment of downstream signaling proteins (Blume-Jensen and Hunter, 2001). All RTKs thus far identified have between one to three tyrosines in the kinase activation loop (located in the kinase catalytic domain). This autophosphorylation activity is important for the biological function and catalytic activity of VEGF, insulin, FGF, PDGF, TrkA and Met receptors (RTKs). Autophosphorylation of the receptor could occur in *trans* (between receptors) or in *cis* (within a receptor) (Hubbard and Till, 2000). Furthermore, no conformational change of the receptor is required upon ligand-induced dimerization for *trans*-autophosphorylation, as close proximity is sufficient for this event to occur. However to facilitate *cis*-autophosphorylation a conformational change occurs in the receptor (Hubbard and Till, 2000). Structural studies using insulin receptor kinase domain as a model have indicated that the activation loop of PTKs are *trans*-autophosphorylated and the other sites like carboxy-terminal tail and juxtamembrane region might be *cis*-autophosphorylated (Hubbard, 1997). RTKs can be downregulated in different ways such as by protein tyrosine phosphatases (PTPs), ubiquitin-directed proteolysis and receptor mediated endocytosis (Hubbard and Till, 2000).

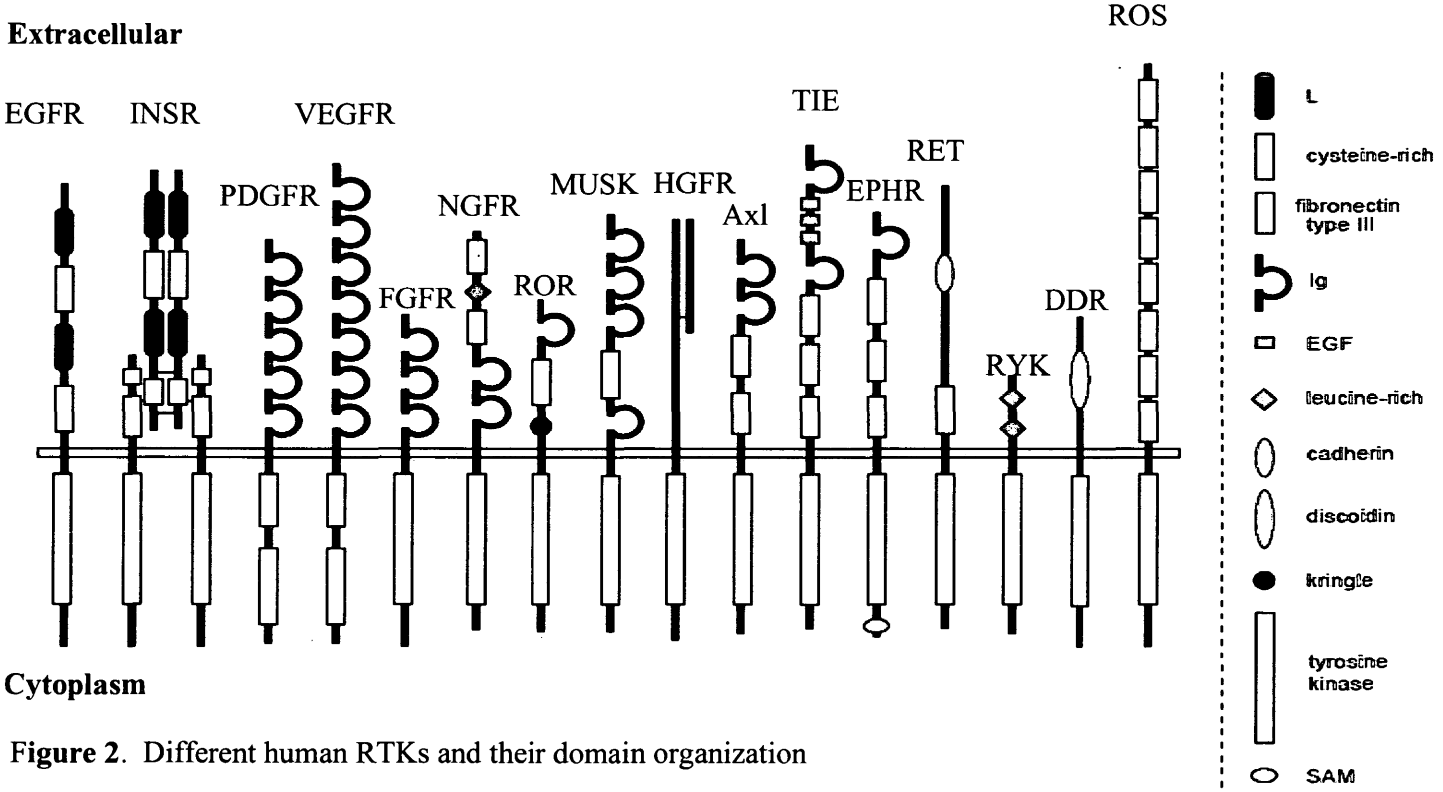


Figure 2. Different human RTKs and their domain organization

At the top it is the extracellular portion of the receptors and at the bottom it is the cytoplasmic portion. The break in the rectangular symbol for some RTKs (i.e. PDGF receptor) represents a large insert in the tyrosine kinase domain. The receptors starting from left are epidermal growth factor receptor (EGFR), insulin receptor (InsR), platelet-derived growth factor receptor (PDGFR), vascular endothelial growth factor receptor (VEGFR), fibroblast growth factor receptor (FGFR), nerve growth factor receptor (NGFR), receptor orphan (ROR), muscle-specific kinase (MUSK), hepatocyte growth factor receptor (HGFR), Axl, a Tyro3 PTK, tyrosine kinase receptor in endothelial cells (TIE), ephrin receptor (EPHR), rearranged during transfection (RET), receptor related to tyrosine kinases (RYK), discoidin domain receptor (DDR), and reactive oxygen species (ROS). Diagram from Hubbard and Till (2000).

1.3.2. Src homology-2 (SH2) domain

Autophosphorylation of tyrosine sites recruits cytoplasmic signaling proteins that contain SH2-domains (Blume-Jensen and Hunter, 2001). SH2 domains are approximately 100 amino acids in length and specifically bind to autophosphorylated tyrosine kinases by selectively recognizing phosphotyrosine sites. At the phosphotyrosine site, there are between 3-6 residues C-terminal that are specific to that protein and are discriminately recognized by SH2 domain containing proteins. This ability of SH2 domains to differentiate between various phosphorylated motifs provides the element of specificity. In other words the sequence at the receptor's autophosphorylation site regulates which SH2-containing proteins to recruit therefore activating a specific biochemical pathway (Blume-Jensen and Hunter, 2001; Pawson et al., 2001; Lim and Pawson, 2010). SH2 domains are found in different proteins such as enzymes (e.g. phospholipase C, protein/lipid kinase, phosphatase domains), docking proteins (e.g. mammalian ShcA), adaptors (e.g. Grb2), transcription factors (e.g. Stat), oncogenic proteins (e.g. Bcr-Abl) and regulators (e.g. SOCS) (Pawson et al., 2001). Mutations of SH2 domains, C-terminal and phosphotyrosine proteins have a broad medical relevance since they are principal players in various human malignancies and disorders such as immune deficiencies, cancer and diabetes (Liu et al., 2011a).

The evolution of SH2 domains in premetazoans concomitantly expanded with the evolution of PTPs and was then followed by the appearance of tyrosine kinases (Liu et al., 2011a). In the study by Liu et al. (2011a) the genome of 21 organisms from two eukaryotic divisions, the Unikonta ('single flagella') and Bikonta ('two flagella') were used to study the evolution of SH2 domains by applying bioinformatic analysis. The Bikonta included organisms such as *T. thermophila, Arabidosis thaliana* and others; and the Unikonta included organisms such as *Homo sapiens, Monosiga brevicollis, Dictyostelium discodeium*, and others. The evolution of the complete set of tyrosine phosphorylation signaling components (PTP, PTK, and SH2) may have facilitated in the evolution of metazoans. SH2 domain families grow larger in numbers the more complex the organism. This diversification of tyrosine phosphorylation signaling by domain duplication and shuffling could have evolved from a common ancestor as proposed by Liu et al. (2011a). Three major events that seem to have driven the diversification of SH2-encoding genes are domain gain, whole gene duplication and domain loss. A small number of proteins in various metazoans could not be classified in the SH2 families, indicating that these proteins could be lineage specific, lost proteins or don't have a present common ancestor (Liu et al., 2011a).

The gene TTHERM_01093550 (zinc knuckle family protein) was identified as encoding an SH2 containing-protein in *T. thermophila* by Liu et al. (2011a). This protein is an ortholog to the conserved gene encoding Spt6 (transcriptional regulator). Spt6 has the most recognizable SH2 domain across all Eukaryota (except two Unikonta organisms identified by Liu et al. (2011a): *Dictyostelium purpureum* and *Entamoeba histolytica*). Another study (Li et al., 2006) investigated *ASI2* (anlagen stage induced 2) gene in *T. thermophila* which is not essential for vegetative growth but it is up-regulated during the development of new macronuclei (sexual cycle). The protein product of this gene, Asi2p, resembles a signal transduction receptor according to sequence analysis and it is proposed that it is a transmembrane protein. Also an SH2 domain binding site was identified in the Asip2 amino acid sequence using bioinformatic analysis. When phosphorylated this protein is proposed to bind to the SH2 domain of an adaptor protein of the Nck1 (non-catalytic region of tyrosine kinase 1) class. These recent findings provide evidence of a possible tyrosine phosphorylation signaling pathway in the ciliate *T. thermophila*.

1.3.3 Non-Receptor Tyrosine Kinases

NRTKs (Figure 3) don't have receptor-like features such as the transmembrane helix or the extracellular ligand domain. The majority of NRTKs are localized within the cytoplasm but some NRTKs can be localized to the cell membrane by amino-terminal modifications. Aside from the tyrosine kinase domain that NRTKs contain, they also have domains that mediate protein-lipid, protein-protein and protein-DNA interactions. NRTK, like RTK is regulated by using tyrosine phosphorylation modification. NRTKs are therefore activated by tyrosine phosphorylation. Tyrosine phosphorylation at the activation loop (a-loop) in these kinases occurs via *trans*-autophosphorylation or phosphorylation by a different NRTK. Common NRTK protein-protein interaction domains are SH2 and SH3 (Src homology 3) domains (Hubbard and Till, 2000). The SH3 domain is composed of about 60 residues and interacts with Pro-X-X-Pro (proline rich) motifs of downstream targets (Lim and Pawson, 2010). NRTKs are crucial components to the signaling cascades triggered by RTKs and other cell surface receptors like G protein-coupled receptors and receptors of the immune system (Hubbard and Till, 2000).

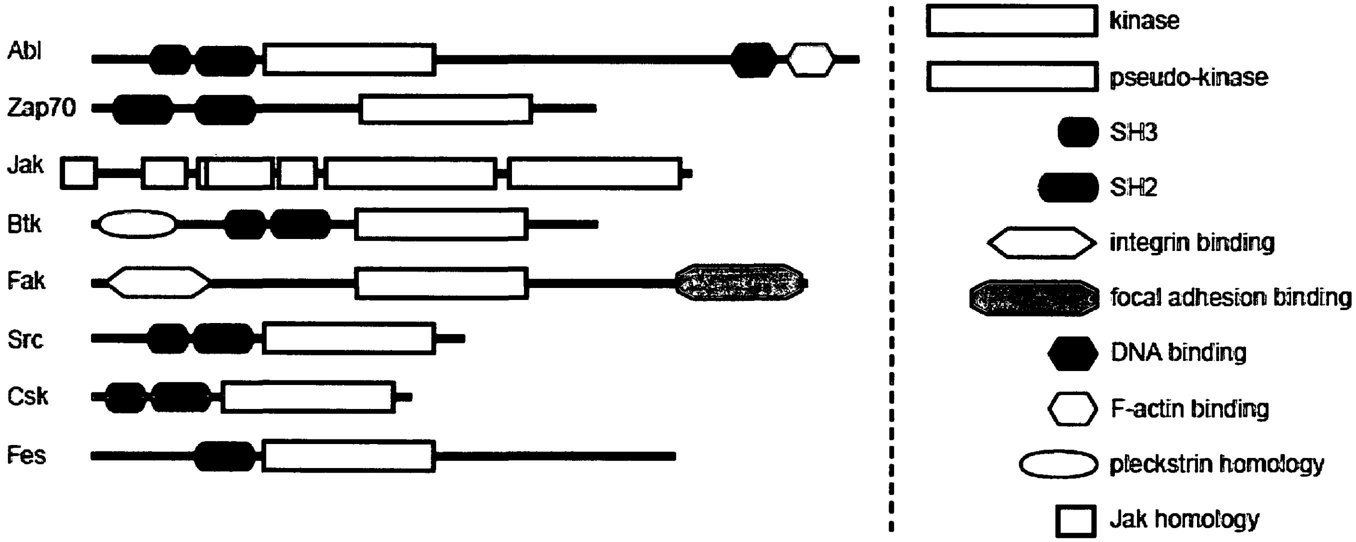


Figure 3. Domain organization for major subfamilies of NRTKs.

On the left is the amino terminus and to the right is the carboxy terminus. Diagram from Hubbard and Till (2003).

1.3.4 Src subfamily

NRTKs' largest subfamily is the Src subfamily which contains nine members. Src subfamily members are involved in various signaling processes such as cytoskeleton restructuring, mitogenesis and T-and B-cell activation. Members of this family contain both SH2 and SH3 domains and tyrosine kinase domains. The c-Src protein (a homologue to a viral oncoprotein which was the first tyrosine kinase discovered, v-Src) becomes deactivated or autoinhibited when it is tyrosine phosphorylated at Tyr527 at the tyrosine kinase domain in humans. The SH2 domain binds to Tyr527 and the SH3 domain binds to a proline rich region which is located between the SH2 and kinase domain (Figure 4). Activation of this protein occurs when it is dephosphorylated at the tyrosine kinase domain. Here, the SH2 domain interacts with the tyrosine kinase domain of a ligand (RTK) and the SH3 domain with the proline rich region of the ligand (Blume-Jensen and Hunter, 2001).

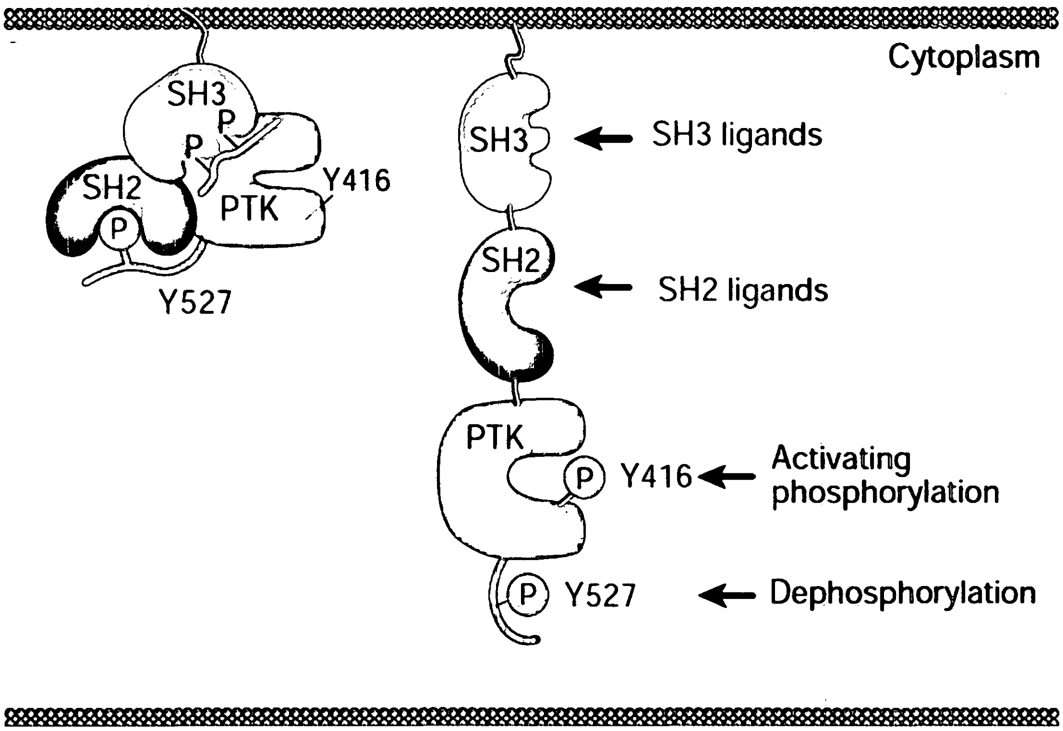


Figure 4. c-Src activation

Left: Unstimulated state leads to tightly repressing c-Src kinase activity. Phospho-Tyr527 interacts with SH2 domain and SH3 domain binds to a proline rich region which is located between the SH2 and kinase domain. Right: Dephosphorylation of protein leads to its activation. SH2 domain interacts with ligand that contains tyrosine kinase domain and SH3 domain with the proline rich region of the ligand. Diagram from Blume-Jensen and Hunter (2001).

1.3.5. Jak/Stat subfamily

The Janus PTKs (JAKs) are cytoplasmic PTKs and lack SH3 and SH2 domains. They mediate signaling downstream of cytokine receptors and RTKs. STATs (signal transducer and activator of transcription) are latent transcriptional factors that have a C-terminal SH2 domain and a central DNA-binding region. Ligand-induced oligomerization of the receptors (i.e. cytokines) cause the receptor to be tyrosine phosphorylated by JAKs followed by STATs (through its SH2 domains) binding to the phosphotyrosine residues. The STATs also become tyrosine phosphorylated by JAKs at the C-terminal tyrosine residue. Next, dimeric STATs are released from the receptors to travel to the nucleus to activate transcription of certain genes (Figure 5) (Blume-Jensen and Hunter, 2001).

1.4. Tyrosine phosphorylation in unicellular organisms

It had for some time been thought that tyrosine phosphorylation signaling was not present in unicellular eukaryotes. Recent studies with a number of unicellular eukaryotes have provided evidence for tyrosine phosphorylation and have provided insights into the origins and evolution of tyrosine phosphorylation signaling. Some organisms used in these studies are *M. brevicollis*, *S. cerevisiae* and *D. discoideum*. *M. brevicollis* contains a true tyrosine phosphorylation system (tyrosine kinases, SH2 domain, and PTP) (Lim and Pawson, 2010) whereas *S. cerevisiae* has PTP and tyrosine kinases (Zhu et al., 2000; Pincus et al., 2008) and *D. discoideum* has SH2 domains and PTP (Plowman et al., 1999; Kawata, 2011). *M. brevicollis* is the closest known relative to metazoans followed by *S. cerevisiae* and *D. discoideum* (Figure 6) (Lim and Pawson, 2010). Also, a paper on *Tetrahymena* phosphoproteome has been submitted and it provides evidence for tyrosine phosphorylated proteins (W. Miao, personal communication). Looking at various other organisms at different points along an evolutionary path can help in understanding the development of this signaling pathway and its functions (Pincus et al., 2008).

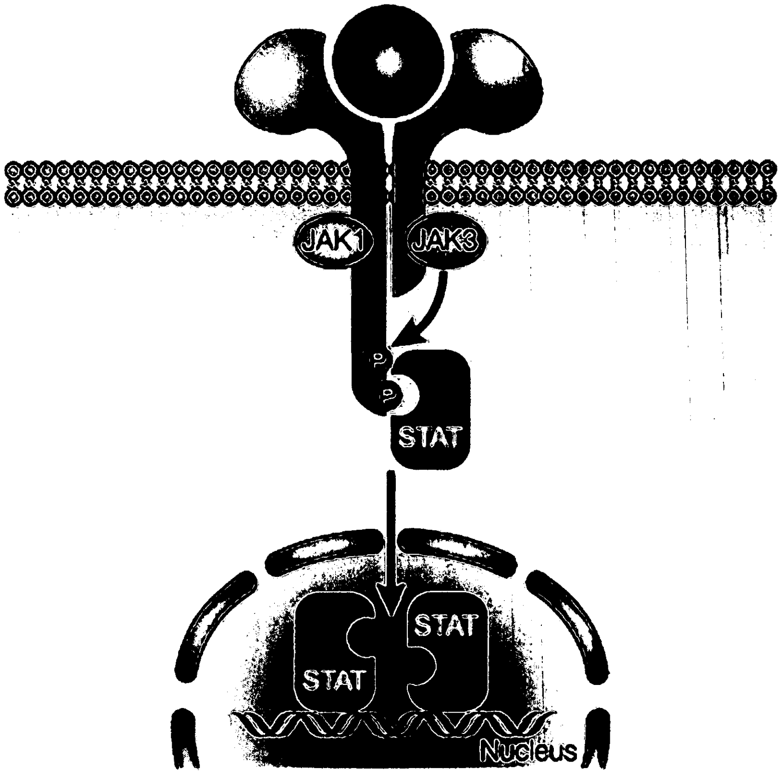


Figure 5. JAK/STAT signaling

Receptor is tyrosine phosphorylated by JAK. STATs then bind to the phosphotyrosine residues becoming phosphorylated as well. Dimeric STATs travel to the nucleus to activate transcription of specific genes. Diagram from O'Shea et al. (2004).

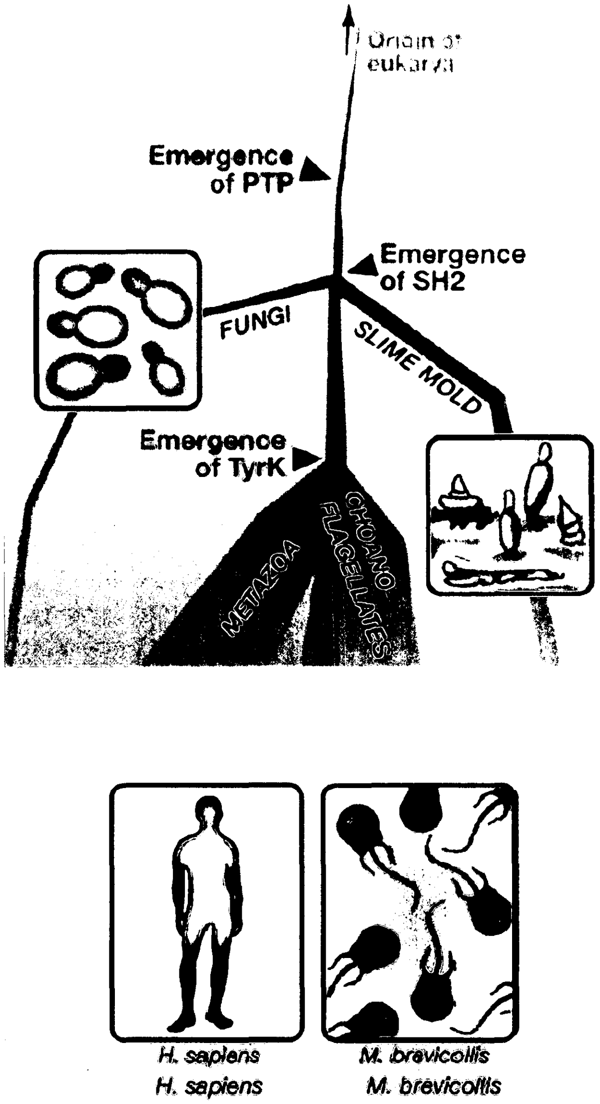


Figure 6. Evolution of pTyr signaling

It is postulated that PTP is first to emerge in the evolution of a true tyrosine phosphorylation system. This is followed by the emergence of SH2 and tyrosine kinases. PTP goes as far back to fungi, SH2 to slime mold and tyrosine kinases to metazoans and choanoflagellates. Diagram from Lim and Pawson (2010).

1.4.1. Tyrosine phosphorylation in *M. brevicollis*

The choanoflagellates are a group of colonial and unicellular flagellates. These cells resemble those found in metazoans (or Animalia). *M. brevicollis* belongs to this group of organisms and is the closest known relative to Metazoa. Phylogenetic analysis of actin, α-tubulin, β-tubulin and elongation factor 2 sequences showed strong evidence that choanoflagellates and Metazoa are closer relatives to each other than to Fungi (King and Carroll, 2001). This was further supported by phylogenetic analysis (Hedges et al., 2004).

Recent analysis of the genome and kinome of this organism has led to intriguing insights into understanding tyrosine phosphorylation evolutionary origins. Analysis of the expressed sequences in this organism by King and Carroll (2001) led to the discovery of the first protein tyrosine kinase (MBRTK1) found outside multicellular organisms. This protein has multiple extracellular ligand-binding domains which resembles that of RTKs in Metazoa.

Later studies (Manning et al., 2008) revealed that this organism has a complex tyrosine phosphorylation network. The same level of complexity is also seen in humans. This was determined by analysis of the genome which led to the discovery of 128 tyrosine kinases (seven orthologous to those in humans), 123 phosphotyrosine (pTyr)-binding SH2 proteins (26 orthologous to those in humans) and 39 tyrosine phosphatases (four orthologous to those in humans).

This unicellular organism contains a three part system, pTyr (tyrosine phosphorylation) toolkit (Figure 7), as described by Lim and Pawson (2010) like that of the metazoans (i.e. humans). The pTyr toolkit is composed of a writer (i.e. tyrosine kinase), reader (i.e. SH2) and eraser (i.e. PTP) (Lim and Pawson, 2010).

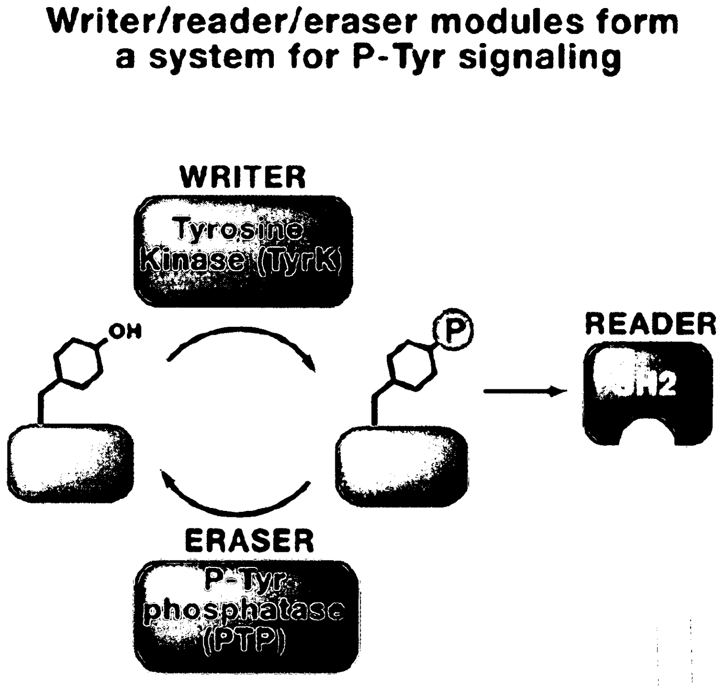


Figure 7. Writer, reader and eraser pTyr toolkit

The tyrosine kinase, SH2 and P-Tyr phosphatase serve as the writer, reader and eraser system, respectively, in tyrosine phosphorylation signaling. Diagram from Lim and Pawson (2010).

1.4.2. Tyrosine phosphorylation in *S. cerevisiae*

The budding yeast *S. cerevisiae* genome has been sequenced and has provided insight into determining the total number of genes and the specific families they belong to. Sequencing the genome allowed investigators to search for possible genes encoding protein kinases. 113 conventional protein kinase genes have been determined which make up 2% of the total genes. From these protein kinases, 60% have unknown or known suspected functions (Hunter and Plowman, 1997). Furthermore, analysis by Hunter and Plowman (1997) has shown seven dual specificity kinases (i.e. MAPKs) in this organism.

In vitro analysis using a novel protein chip technology to identify phosphorylated kinases in yeast helped identify 32 kinases that preferentially phosphorylate one or two substrates and 27 of these are potential tyrosine kinases (Zhu et al., 2000).

However, no members of non-receptor PTK or true receptor family has been detected using the genome of *S. cerevisiae* (Hunter and Plowman, 1997). Furthermore, PTPs and SH2 domains in Spt6 protein have been detected in this organism (Lim and Pawson, 2010; Liu et al., 2011b). Therefore, *S. cerevisiae* has potential tyrosine kinase (writer) proteins and multiple PTP (eraser) but no true tyrosine phosphorylation-recognizing SH2 (reader) domain (Lim and Pawson, 2010). This SH2 domain containing protein, Spt6, doesn't just recognize tyrosine phosphorylation but also serine/threonine phosphorylation (Lim and Pawson, 2010; Liu et al., 2011b) making it a dual specificity SH2 domain protein. Spt6 is found in all eukaryotes including all fungi and metazoans but not in prokaryotes. It is considered to likely represent an early ancestor or relative that eventually gave rise to modern SH2 domains.

1.4.3. Tyrosine phosphorylation in *D. discoideum*

D. discoideum is a Mycetozoa (slime mold) protozoan and a soil living amoeba which in the presence of bacterial food has a unicellular lifestyle. Otherwise, when there is food shortage, the individual cells are triggered by the chemoattractant cAMP to turn into a multicellular structure. This is then followed by the stalk and spore cells differentiating and developing into a fruiting body (Lim and Pawson, 2010). The process of differentiation is a series of highly coordinated morphological, physiological and cellular changes. The process of multicellular development simultaneously occurs with a large transition in the transcription pattern and it takes place throughout 6-8 hours of starvation (Van Driessche et al., 2002).

The differentiation process in this organism is important in understanding the evolution of tyrosine phosphorylation. Studies in *D.discoideum* have shown the presence of phosphotyrosine phosphatases and dual specificity kinases. The *D.discoideum* genome shows 13 proteins that have SH2 domains (including an Spt6 homolog). Two of these proteins (Cbl and STAT) are homologous to SH2 proteins in metazoans and three others (Shk, LrrB and FbxB proteins) are unique to this organism. There are four STAT (homologous to metazoan and have SH2 domain) but no JAK in *D.discoideum* such as the JAK/STAT system found in the metazoan (Plowman et al., 1999; Lim and Pawson, 2010). The fact that this organism has STAT which is not present in yeast indicates that proteins of this family must have originated early in evolution regardless of the fact that *Dictyostelium* is not regarded as a direct ancestor to metazoans (Darnell, 1997; Kawata, 2011). The first STAT discovered is known as Dd-STATa (*D.discoideum*) and it is phosphorylated at the tyrosine residue near the C-terminus at position 702. However, the kinase responsible for tyrosine phosphorylation of STAT still remains a mystery.

Therefore, *D.discoideum* has SH2 domains (reader) and PTPs (eraser). There is a TKL (tyrosine-kinase-like) kinase that could possibly phosphorylate Dd-STATa but it is not required. This indicates that STAT signaling in *D.discoideum* might not be as simple as that of the metazoan JAK/STAT system (Kawata, 2011).

1.5. CDC2

Analysis of specific genes and their role in *Tetrahymena* could help us understand their function and provide insight into understanding their possible involvement in tyrosine kinase signaling. To study tyrosine phosphorylation in a protozoan, *Trypanosoma brucei,* Nett et al., (2009) used a gel-free, phosphopeptide enrichment-based proteomics. They used a PhosphoScan PTyr- 100 kit (Cell Signaling) followed by mass spectrometry to identify phosphorylated tyrosine peptides in *T. brucei.* Bioinformatic analysis was done on these peptides (Table 1 and 2 from (Nett et al., 2009). Bioinformatic analysis of these peptides were compared to *Tetrahymena* and revealed that the protein product of the gene TTHERM_00483640, a protein kinase of the CDC2 subfamily, potentially has tyrosine phosphorylation sites and we will refer to the protein as TtCdc2.

CDC2 and its protein product (Cdc2) is a cell division cycle protein kinase and a homolog to the catalytic subunit of cdk (cell cycle cyclin-dependent kinase) (O'Connell et al., 2003; Dou et al., 2005). Cdc2 controls the entry of cells into mitosis and studies on this kinase have shown that it is present in all eukaryotic cells suggesting that it is a universal mitotic trigger (Lee and Nurse, 1987). Activation of Cdc2 leads to downstream events and the regulation of NIMA-related (Nrk) kinases which are also highly conserved in evolution (Figure 8). The transition from G2 to mitosis includes the activation of Cdc2 by dephosphorylation at tyrosine 15 (Y15) (fungi). This is catalyzed by a Cdc2 tyrosine phosphatase leading to the phosphorylation of NIMA. This is followed by a rise in kinase levels and NIMA proteins. NIMA becomes further phosphorylated but it is unclear if this is because of Cdc2 again. As the cells transfer from metaphase to anaphase, ubiquitin-dependent proteolysis of the B-type cyclins (dependent on anaphase-promoting complex) inactivates Cdc2. Another ubiquitin-dependent proteolysis that affects NIMA is dependent on C-terminal PEST sequences (fungi) (O'Connell et al., 2003).

1.5.1. *Tetrahymena* Cdc2

In *Tetrahymena*, histone H1 is highly phosphorylated and plays an important role in gene expression regulation during the starvation response, while its role in the regulation of transcription can be negative or positive (Dou et al., 2005). Starved *Tetrahymena* cells treated with phosphatase inhibitor (phosphatase is an enzyme that removes the phosphate group from the substrate) causes overexpression of the CDC2 gene. Also, it has been shown that Cdc2 kinase phosphorylates *Tetrahymena* H1.These observations indicate that Cdc2 is a kinase which phosphorylates histone H1 in *Tetrahymena* (Dou et al., 2005). Further studying Cdc2 could help determine the role of tyrosine phosphorylation signaling in *Tetrahymena* and possible involvement in cell division.

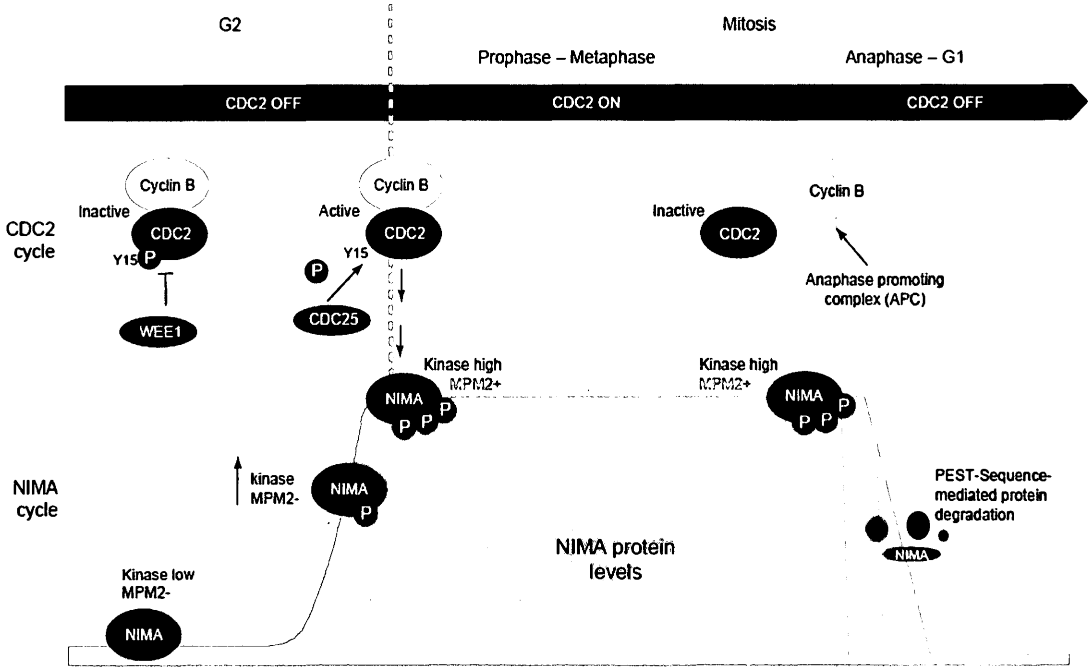


Figure 8. Regulation of Cdc2 and NIMA

Activation of Cdc2 by dephosphorylation at tyrosine 15 (Y15) (fungi) leads to the transition from G2 to mitosis. Cdc2 tyrosine phosphatase catalyzes this leading to the phosphorylation of NIMA. Kinase levels and NIMA protein levels rise. NIMA becomes further phosphorylated. During the transfer from metaphase to anaphase, Cdc2 is inactivated by ubiquitin-dependent proteolysis of the B-type cyclins. Diagram from O'Connell et al. (2003).

1.6 Previous work and objective

Previous studies have indicated that there are no typical receptor tyrosine kinases (RTK) and growth receptors in unicellular eukaryotes (Schemarova, 2006). Therefore it would have been logical to assume that eukaryotic unicells shouldn't become stimulated by mammalian growth hormones. However some data suggest that this is not the case (Schemarova, 2006). Although these studies are not all convincing, some reports (Christopher and Sundermann, 1996; Christensen et al., 2003; Schemarova, 2006) suggest that induction by insulin alters the growth rate and metabolism level in *Tetrahymena.*

Furthermore, stimulation by mammalian insulin induces a high level of a putative tyrosine kinase (in fact a NRK/NIMA kinase not definitively shown to be a tyrosine kinase) correlated with an increase in tyrosine phosphorylation activity in *Tetrahymena* (Christensen et al., 2003). NRK family of protein kinases are often found associated with microtubular organelles such as flagella, basal bodies, centrioles and others. Macronuclear genome sequencing in *T. thermophila* led to the identification of 39 NRKs. This number is about 3 times higher than the number of such loci in humans (Eisen et al., 2006). The kinase domains of these NRKs have between 36 to 59% amino acid sequence similarity to those in humans. Overproduction of NRKs in *Tetrahymena* have been shown to cause rapid shortening of cilia. On the contrary, overexpression of the kinase-inactive variants led to lengthening of the cilia (Wloga et al., 2006).

Bioinformatic analysis from the Pearlman laboratory (Anafi, M., Glowacka, W., Golding, B., and Pearlman, R.E., unpublished observations), suggest that there are no true tyrosine kinases in *Tetrahymena* but there are least 12 dual specificity kinases designated as *Tetrahymena* Serine/Threonine/Tyrosine Kinases (tSTYK). At least 10 of these kinases have a higher e-value for tyrosine than for serine/threonine kinase based on their structural motifs and sequence. The fully sequenced genome of *Tetrahymena* has also allowed the identification of kinase orthologs (Eisen et al., 2006). Also, the Pearlman lab has identified three tyrosine specific phosphatases designated as *Tetrahymena* Protein Tyrosine Phosphatase (tPTP) (Anafi, M., Glowacka, W., Golding, B., and Pearlman, R.E., unpublished observations). These bioinformatic analyses support indirect immunofluorescence data (Figure 9) which indicate that there likely is tyrosine phosphorylation in *Tetrahymena.* Treatment of *Tetrahymena* cells with 100μM genistein, a tyrosine specific protein kinase inhibitor, showed a decrease in the percentage of cells paired and a decrease in phosphorylation signal as compared to cells not treated with genistein. A paper submitted on *Tetrahymena* phosphoproteome analysis has presented direct evidence for tyrosine phosphorylated proteins in this organism (W. Miao, personal communication).

My research focused on determining the presence of tyrosine kinase phosphorylation in the ciliated protozoan, *T. thermophila.* The reputation of this organism as an excellent animal model for much biomedical research and the availability of its fully sequenced genome (including kinase orthologs) allowed us to carry on the necessary analysis. To study tyrosine phosphorylation in a protozoan organism, *Trypanosoma brucei,* Nett et al., (2009) used a gel-free, phosphopeptide enrichment-based proteomics. They used a PhosphoScan PTyr-100 kit (Cell Signaling) coupled to mass spectrometry to identify phosphorylated tyrosine peptides in *Trypanosoma brucei.* Then bioinformatic analysis was done on these peptides (Table 1 and 2 from (Nett et al., 2009). I used the peptides from Nett et al. (2009) and performed bioinformatic analysis by comparing them to a database of *Tetrahymena* predicted protein sequences. SDS PAGE, Western blotting and immunofluorescence microscopy analysis were done with *Tetrahymena* after starving and pairing (conjugation) to assay for the presence of tyrosine phosphorylated proteins. *Tetrahymena* was transformed to knockout the TtCDC2 gene and based on bioinformatic analysis it has potential tyrosine phosphorylation sites. Producing knockout strains for this gene will allow us to study its function in *Tetrahymena* and determine its relevance to tyrosine phosphorylation signaling.

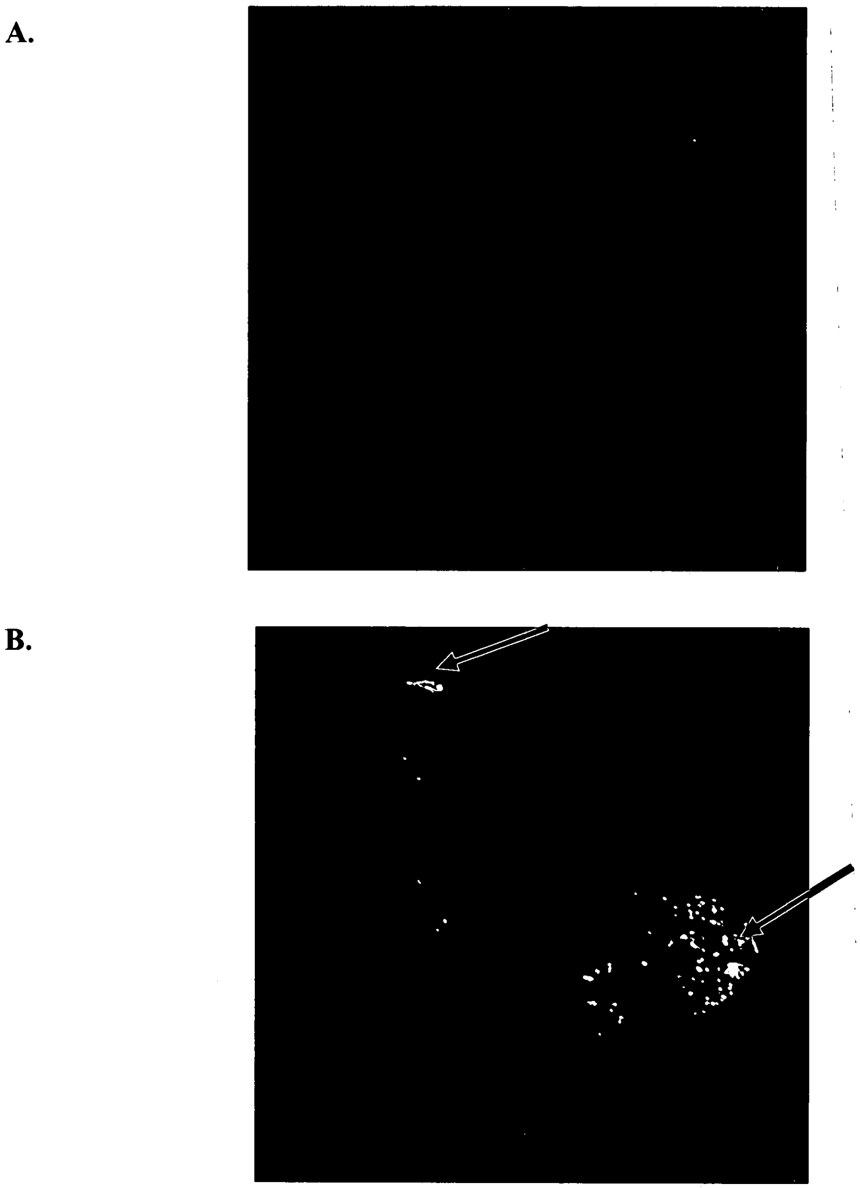


Figure 9. Indirect immunofluorescence results from Pearlman lab

Previous indirect immunofluorescence results using anti-phosphotyrosine antibodies to detect tyrosine phosphorylated proteins in *T. thermophila,* done in the Pearlman lab (Anafi,M., Glowacka, W., Golding, B., and Pearlman, R.E., unpublished observations). A) Starved strain B2086. B) Two mating types conjugating at 3hrs. The arrows represent a concentrated region of tyrosine phosphorylation at the posterior region of the cell as well as general cytoplasmic labelling in some cells.

2. MATERIALS AND METHODS

2.1 Cell strains

Detection of tyrosine phosphorylated proteins was investigated in *T. thermophila* using strains B2086 (Mpr+/Mpr+ ; mating type II; mp-s) and CU428 (Mpr/Mpr; mating type VII; mp-s). Both these strains are of inbreeding line B and are obtained from the *Tetrahymena* Stock Center, Cornell University, http://tetrahymena.vet.cornell.edu/.

2.2 Cell growth

Bench top cultures of *T. thermophila* strains B2086 and CU428 were grown separately in 5.0 ml of SPP medium (Appendix A) at room temperature and transferred regularly. Analysis was done using three day old cultures. A sample from the bench cultures was inoculated to fresh SPP medium and grown to a density of 3.0×10^5 cells/ml at 30°C with mild agitation for 18 hrs. The concentration of the cells was determined using a hemocytometer .

2.3 Washes, starvation and conjugation

The overnight cultures of each strain were starved separately. This was done by taking each strain and transferring to polypropylene tubes. For large volumes (i.e. 100ml) and small volumes (i.e. 20ml), the volumes were split equally and transferred into polypropylene tubes of 50ml and 10ml respectively. Once the cultures of each strain were transferred to polypropylene tubes, they were centrifuged for 5 min at 1, 000 rpm at room temperature using an IEC PR-600 centrifuge (Model 49695M-P; Thermon Electron Corp.). The supernatant was removed by aspiration leaving the pellets in the tube. The pellets were resuspended in 10mM Tris-HCl pH 7.4 using the same volume as that of the initial cultures. Centrifugation and aspiration were repeated. The pellets were again resuspended in 10mM Tris-HCl pH 7.4 using the same volume as that of the initial cultures (to give the same cell concentration as the overnight cultures) followed by their transfer to Erlenmeyer flasks. Nominal flask size was 10X the volume of culture; e.g. 1L flask for 100 ml of culture. These flasks were placed at 30°C without shaking for 18hrs and processed for analysis or mixed with cells of a different mating type to initiate conjugation.

2.4 Western blotting

Some of the experiments involve analysis of conjugating cells which includes pairing and mating of B2086 and CU428 cells. Both strains that were separately starved were counted to determine cell concentration. If cell concentrations were unequal between the two strains, they were adjusted to equal cell concentrations. The cell concentrations for both strains were kept at 3.0×10^5 cells/ml. Equal volumes of B2086 and CU428 were mixed. The cells were left to pair for a specific amount of time up to about 90% pairing and then transferred to a polypropylene tube and centrifuged (5min at 1,000rpm; IEC PR-600 centrifuge Model 49695M-P (Thermon Electron Corp)) to collect the pellet containing the cells for protein analysis. The supernatant was discarded.

2.4.1 Preparation of blotto

The 3% blotto was prepared fresh the day of performing Western blot experiments. 3% blotto was made by dissolving powder skimmed milk (Carnation Nestle) in 1X TBS pH 7.5 solution in a flask with a magnet inside, stirring for 15 min. The dissolved milk was transferred to Sorvall bottles and centrifuged in Avanti J-301(JA-30.50 Ti Rotor; Beckman Coutler) for 20 min at 6,000 rpm at 5°C. The supernatant was collected and filtered using a 3mm Whatman paper. The filtrate (3% blotto) was kept at 4°C for the time that it wasn't being used and placed on ice when being used.

2.4.2 SDS PAGE and sample preparation

Tris-HCl pH 7.4 buffer was added to the cell pellet until there was buffer above the pellet, suspended and mixed with an equal amount of Laemmli buffer (see Appendix A). Bradford analysis was done for all samples used in protein gel electrophoresis sample to determine protein concentration using Bradford Protein Assay Kit I (Bio-Rad). Equivalent loading was checked by Coomassie staining of the gel.

2.4.3 SDS PAGE, transfer and analysis

Mini 10% polyacrylamide gels were prepared according to the standard method (see Appendix A). Samples of 100 μg of total cell lysate were loaded next to a molecular weight marker. A Mini-PROTEAN Tetra Cell apparatus (Bio-Rad) was used to separate proteins in the gel. The buffer system contained 25mM Tris base, 190mM glycine, and 0.1% SDS and the gel electrophoresis ran for about 1.5hr at 100 Volts. The gel was assembled into the transferring sandwich (sponge-filter-membrane-gel-filter-sponge; Mini Trans-Blot Cell – Bio-Rad) followed by electrotransferring at 300mA for 2 hrs. The PVDF membrane was washed twice with TBST and blocked with 3% blotto for 1hr at room temperature with constant agitation. The membrane was blocked with a primary antibody 4G10 Platinum (mouse; anti-phosphotyrosine; Millipore) overnight at 4°C and shaking. The primary antibody dilution used was 1:1000 in 3% blotto with TBST. This was followed by washing the membrane four times with TBST. The membrane was then stained with 1:3000 secondary antibody (goat anti -mouse IgG tagged HRP; Thermo Scientific) in 3% blotto with TBS for 1hr at room temperature with constant agitation. The membrane was washed in TBST and rinsed in ddH_2O. The membrane was processed for ECL detection (Amersham Biosciences) according to the supplier's instructions. In the dark room the signal was detected by using HyBlot CL Autoradiography films (Denville Scientific) for different lengths of exposure such as starting at 30 seconds.

2.5 Indirect immunofluorescence microscopy

Indirect immunofluorescence was used to complement Western blot data and to confirm previous indirect immunofluorescence results from our lab (Anafi, M., Glowacka, W., Golding, B., and Pearlman, R.E., unpublished observations).

Cells that had been pairing for 3hrs were collected at $3x10^5$ cells/ml. A volume of 6mL of cells was treated with Schaudinn's reagent (Appendix A) at 20µl for 5 min and centrifuged in the IEC clinical centrifuge (Model 49695M-P; Thermon Electron Corp) for 90s at setting 4. The supernatant was removed by aspiration and the pellet was resuspended in 95% methanol for 2 min. The sample was again centrifuged and resuspended in 2ml of 95% methanol. The fixed cells (slurry) were placed on a microscope coverslip which was left to air dry and then incubated in PBS for 30 min at room temperature. The cells on the coverslip were blocked with 5% blotto (5% instant skimmed milk - Carnation Inc. in TBS). The blocking solution was placed on the cover slip for 30 minutes.

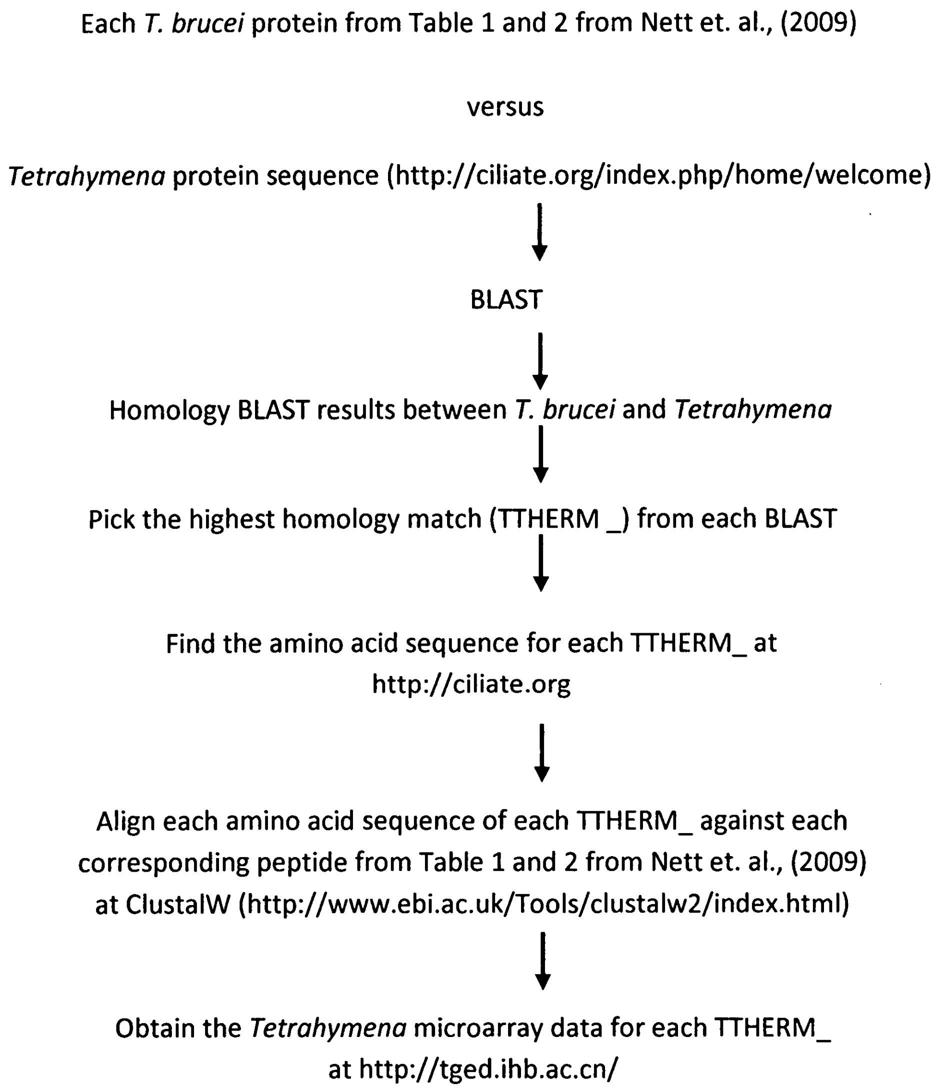
The cells were incubated with 4G10 Platinum antibody at a 1:200 dilution at 4°C overnight, followed by a washing with the blocking solution and incubation with secondary antibody (rabbit anti mouse FITC antibody – Cedarlane) at a dilution of 1:500 for 2hrs. The cells were again washed with blocking solution, dried and mounted on the slide with glycerol buffer mixed with DABCO (50mg/ml 1,4-diazabicyclo (2,2,2) octane) and 1µg/ml of propidium iodide for nuclei staining and staging of conjugating cells. The cells where observed with an Olympus Fluoview 300 laser scanning confocal microscope. A 60X oil immersion objective lens was used to take pictures of the cells.

2.6 Bioinformatic analysis

A flow chart of the bioinformatic analysis performed is presented in Figure 10. The list (Table 1 and 2 from Nett et al. (2009)) of *T. brucei* proteins (amino acid sequence) that these phosphorylated peptides corresponded to was analyzed by BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) against a database of all predicted *Tetrahymena* protein sequences (http://www.cilate.org) to look for putative *Tetrahymena* orthologues of the *Trypanosoma* tyrosine phosphorylated proteins. The highest homology match (TTHERM_) from each BLAST was selected and the amino acid sequence for each TTHERM_ was obtained from http://ciliate.org.

Further alignment analysis between each peptide (Table 1 and 2 from Nett et al. (2009)) and the corresponding *Tetrahymena* protein amino acid sequence (TTHERM_) was done using ClustalW (www.ebi.ac.uk/clustalw).

Also NetPhos 2.0 (http://www.cbs.dtu.dk/services/NetPhos/) was used to detect putative high probability tyrosine phosphorylation sites for each *Tetrahymena* protein sequence. (Appendix B-4). Results that had a score higher than 0.500 were considered to have a significant probability of having a tyrosine phosphorylation site in that specific protein amino acid sequence. This was followed by analysis of the *Tetrahymena* expression microarray data (Miao et al., 2009; http://tged.ihb.ac.cn/) for each putative *Tetrahymena* phosphorylated protein (TTHERM_) from the BLAST analysis.



Each *T. brucei* protein from Table 1 and 2 from Nett et. al., (2009)

versus

Tetrahymena protein sequence (http://ciliate.org/index.php/home/welcome)

BLAST

Homology BLAST results between *T. brucei* and *Tetrahymena*

Pick the highest homology match (TTHERM _) from each BLAST

Find the amino acid sequence for each TTHERM_ at http://ciliate.org

Align each amino acid sequence of each TTHERM_ against each corresponding peptide from Table 1 and 2 from Nett et. al., (2009) at ClustalW (http://www.ebi.ac.uk/Tools/clustalw2/index.html)

Obtain the *Tetrahymena* microarray data for each TTHERM_ at http://tged.ihb.ac.cn/

Figure 10. Bioinformatic flow chart

A flow chart of the bioinformatic analysis using programs such as BLAST, ClustalW, *Tetrahymena* microarray database and others, to compare the peptides and proteins from Table 1 and 2 of Nett et al., (2009) to the *Tetrahymena* protein amino acid sequences to determine putative tyrosine phosphorylation sites.

2.7 TtCDC2 knockout

The following methods were applied to make TtCDC2 gene knockout clones of *T. thermophila* to investigate the function of this gene and its possible link to tyrosine phosphorylation signaling.

2.7.1 *Tetrahymena* DNA extraction

DNA from *Tetrahymena* was extracted as follows. The cells were grown overnight at 30°C with shaking to a concentration of 3.0x10^5 cells/ml at a total volume of 20ml. The cells were centrifuged at 3000 rpm (IEC PR-600 centrifuge Model 49695M-P; Thermon Electron Corp.) for 2 min and the supernatant was aspirated, the pellet (300µl) was resuspended in lysis solution (700µl) (Appendix A), mixed and split into two equal volumes in two Eppendorf tubes. The solution was treated with phenol•$CHCl_3$(1:1 volume/volume) at half the volume of the supernatant, mixed, and centrifuged at 13,000 rpm (Sorvall Legend Micro 21R Centrifuge; 24x 1.5/2.0ml Rotor; Thermo Scientific) for 1 minute. The supernatant was transferred to a fresh tube and the process repeated. The next repeat was with $CHCl_3$. The supernatant was treated with 5M NaCl at a volume of 1/3 of the supernatant and with isopropanol at an equal volume as the supernatant. This was followed by centrifuging the mixture at 13,000 rpm for 2min, remove supernatant, dislodge the pellet with 70% ethanol at 1/3 volume as above and centrifuge for 2min at 13,000 rpm. The pellet was washed two more times with ethanol and centrifuged for 5min at 13,000 rpm.The supernatant was discarded and the pellet was dried under vacuum and suspended in 150µl of 10mM Tris-HCl pH 7.4. RNase was added at a final concentration of 10mg/ml in the sample and was incubated at 37°C for 1hr to digest RNA. This was left at 4°C overnight and used the following day or transferred to -20°C and stored until use.

2.7.2 TtCDC2 knockout primers

Custom primers were ordered from Invitrogen, shown in Figure 11a to make the plasmid construct required to knock out the gene of interest. A combination of primers # 4 & 5, and 4 & 6, were used to confirm for the replacement of TtCDC2 with the Neo4 cassette in *T.thermophila* (Figure 11a-b).

2.7.3 Amplification of TtCDC2

Polymerase chain reactions (PCR) were done using Phusion Hot Start II High-Fidelity DNA polymerase (New England Biolabs) in a 50µl reaction volume. The reaction included *Tetrahymena* DNA as the template and the primers outlined in Figure 11a. One reaction was prepared using the primers # 1 & 2 and another using 3 & 4. Duplicates were also made for each reaction. The reactions were prepared according to the manufacturer's protocol. The PCR settings were selected as in Table 1a (right) for the reactions containing primers 1 & 2, and Table 1a (left) for 3 & 4.

Following electrophoresis through agarose gels (Appendix A) at 100 V for 45min, the samples were purified using Spin Column PCR Products Purification Kit (Bio Basic Inc.). A sample from the purified PCR product was eletrophoresed through a 0.8% agarose gel with TBE buffer at 100 V, to confirm that the product contained the expected amplified regions. Gel data were documented using (AlphaImager HP System; Protein Simple).

2.7.4 Enzymatic digestion

To attach the purified PCR products which contain the DNA fragments of interest mentioned above, into the plasmid (pNeo4), corresponding compatible ends were generated. This was first done by using enzymatic digestion. Enzymes ApaI and XhoI were used to digest TtCDC2 5' UTR fragments and SacI and SacII for TtCDC2 3' fragments.

The plasmid used to create the knockout strains was pNeo4 and that too was digested with the same four enzymes (Figure 12a). The protocol and settings used were those provided from the manufacturer, New England Biolabs. Large scale digestions were done in a 400µl reaction.

The final products were again purified using the same kit as above and confirmed by eletcrophoresis through a 0.8% agarose gel and photographed as above.

2.7.5 Ligation, transformation and miniprep

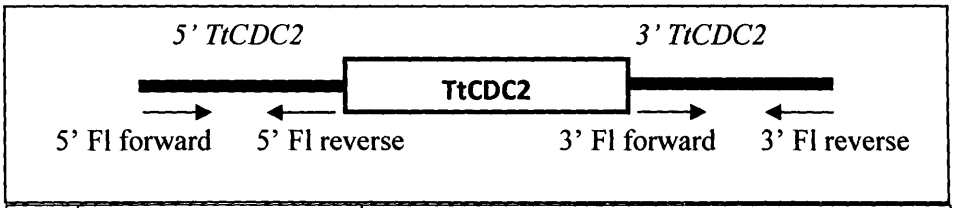
The next step, digested 5' and 3' DNA fragments were ligated to the plasmid as follows. First, 5' fragment was ligated to the plasmid then the other one. T4 DNA Ligase (400-600 units; New England Biolabs) was used for each ligation following the protocol provided by the manufacturer.

The ligation was left overnight at 16°C, this product was then transformed into *Escherichia coli* (DH5-α competent *E. coli*, New England Biolabs) following the manufacturer's protocol. The final mixture containing transformed cells was plated on YT with ampicillin agar plates (Appendix A) and incubated overnight at 37°C to select for transformants.

12 colonies that were ampicillin resistant were picked the following day. Each picked colony was incubated in a test tube with YT+ampicillin and placed at 37°C with shaking overnight. The next day, to isolate the plasmid with the inserts, a miniprep was performed for all 12 cultures generated from the 12 colonies. This was done according to manufacturer's guidelines (Qiagen Miniprep Kit).

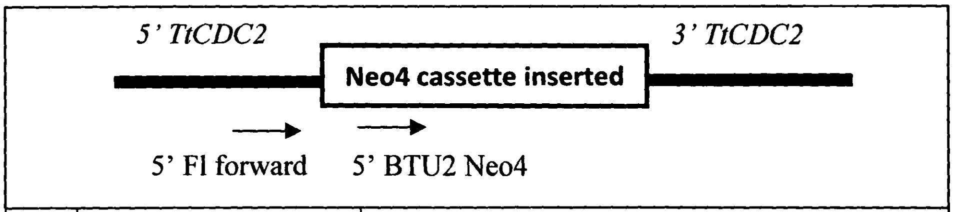
To confirm which cultures had the plasmid with the inserts, the DNA from the 12 cultures were digested with ApaI and SacI and checked by electrophoresis through a 0.8% agarose gel. Colonies putatively containing inserts were further confirmed by sending a DNA sample to the York University Core Molecular Biology DNA Sequencing Facility.

A.



	DNA to amplify	Primer sequence
1.	***TtCDC2 5' Flank*** **Forward**	5'GCGGGGCCCATTTATATTTCTCCCTTATAA AATTTAGAA 3'
2.	***TtCDC2 5' Flank*** **Reverse**	5'GCGCTCGAGGTCTTCAATAGTTTCCAAAT TTTCT 3'
3.	***TtCDC2 3' Flank*** **Forward**	5'GCGCTCGAGGTCTTCAATAGTTTCCAAAT TTTCT 3'
4.	***TtCDC2 3' Flank*** **Reverse**	5'GCGGAGCTCAATTTAAATTTATTTTATCA AAAAAATTTC 3'

B.



	DNA to amplify	Primer sequence
5.	***BTU2 5'*** **Neo4 Forward**	5' GAGCTAACATGTATGTGAAGAGG 3'
6.	***TtCDC2 5' Flank*** **Forward** (different position than #1)	5'GCGCGCAGAAAATTTGGAAACTATTGAAG AC 3'

Figure 11. Primers

(A) Top: The direction of the forward and reverse of the 5' and 3' region amplified in *T. thermophila*. Bottom: The sequence of the primers. (B) Top: The direction of the 5' region on Neo4 cassette and 5' region on TtCDC2 amplified using primers #5 and 6, respectively. Bottom: The sequence of the primers. The red arrows indicate the direction that the primers amplify and location. Not to scale.

Table 1. Conditions for PCR reactions

Temperature, time and cycle conditions used for PCR reaction using Phusion Hot Start II High-Fidelity DNA polymerase kit with the primers from Figure 11. A) The settings here are for 5' primers (left) and 3' primers (right) that correspond to Figure 11a. B) The settings here are for the primers used in the PCR reactions using Prime Star Max 2x, to detect transformants (primers #4, 5-6 from Figure 11a-b).

A.

Steps	**Temp (°C)**	**Time (sec)**	**Cycles**
Initial Denaturation	98	30	1
Denaturation	98	10	5
Annealing	55	30	
Elongation	72	45	
Denaturation	98	10	5
Elongation	58	30	
Extension	72	45	
Denaturation	98	10	13
Annealing	50	30	
Elongation	72	45	
Final Extension	72	10min	1
Storage	4	∞	-

Steps	**Temp (°C)**	**Time (sec)**	**Cycles**
Initial Denaturation	98	30	1
Denaturation	98	10	30
Annealing	55	30	
Elongation	72	30	
Final Extension	72	10min	1
Storage	4	∞	-

B.

Steps	**Temp (°C)**	**Time (sec)**	**Cycles**
Initial Denaturation	98	3min	1
Denaturation	98	10	30
Annealing	55	30	
Elongation	72	1min	
Storage	4	∞	-

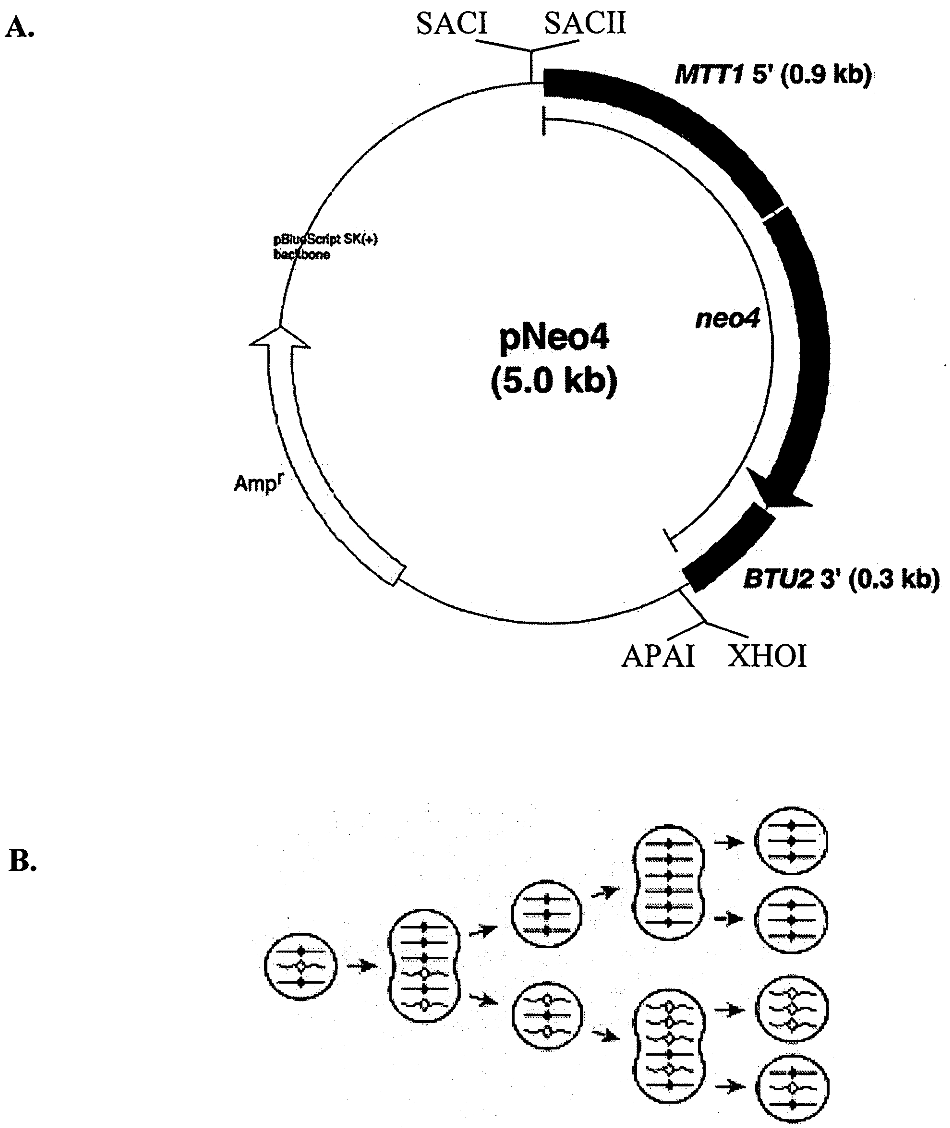


Figure 12. pNeo4 vector and phenotypic assortment

A) pNeo4 vector was used to create macronuclear knockouts. 5' flank of TtCDC2 was digested with ApaI and XhoI and 3' flank was digested with SacI and SacII. Both flanks were inserted as shown. The neo4 cassette then replaces the coding region of TtCDC2. pNeo4 was provided by Dr. K. Mochizuki, Institute of Molecular Biotechnology of the Austrian Academy of Sciences, Vienna, Austria. Amp^r shows resistance to ampicillin and *BTU2* shows the β tubulin gene sequence. B) A diagram showing how a mutant allele accomplishes full expression in the MAC by phenotypic assortment. The open circles with the wavy lines represent genomic contribution from one parent and the dark circles with the straight lines that of the other parent (Eisen et al., 2006).

2.7.6 *Tetrahymena* transformation with TtCDC2 inserts+plasmid

Transformation of *Tetrahymena* with the TtCDC2 inserts+plasmid leads to the replacement of the TtCDC2 gene with the neo4 cassette producing a knockout. The ligated plasmid with TtCDC2 5' and 3' flanks was inserted into the two strains of *Tetrahymena*, starved B2086 and CU428 cells. This was done using the BioRad PDS-1000/HE Particle Delivery system as outlined by Bruns and Cassidy-Hanley (2000). Cultures at 2.0×10^5cells/ml were used with at least 4μg of DNA coated onto 0.6μm gold particles. Cells were shot with about 900 psi and allowed to recover at 30°C.

Tetrahymena has multiple copies of a gene and to make sure that these copies are shut down to let the transformed copy be expressed, a drug selection was applied. Both strains were treated with cadmium chloride at a concentration of 1μg/ml to induce the *MTT1* promoter of the knockout construct. At the same time the cells were treated with increasing concentrations of anti-microbial drug paromomycin sulfate. B2086 was treated with a concentration up to 0.1mg/ml and CU428 up to 7.0mg/ml. All cells were plated on a 96-well microtitre plates and stored in humid boxes at 30°C. The transfers were done every 3 days, to allow for phenotypic assortment to allow for complete replacement of the many copies of the wild type in the MAC (Figure 12b) (Eisen et al., 2006). After the drug treatments the transformants were confirmed for having the knockout construct by performing 20μl PCR reaction volume (settings in Table 1b) using Prime Star Max 2x (Takara) with the primers # 4 and 5 for one reaction and 4 and 6 for another (Figure 11). The DNA templates were from the transformants and the wild type to compare.

3. RESULTS

3.1 Western blot analysis

Initial results with Western blot analysis to assess tyrosine phosphorylation and to possibly initiate approaches to identify tyrosine phosphorylated proteins gave variable and inconsistent results. This as well as the likely low levels of tyrosine phosphorylated proteins in the cell indicated that it would be necessary to optimize the protocol for indicating the presence of tyrosine phosphorylated proteins in *Tetrahymena*. These results also serve to provide further support to the indirect immunofluorescence results previously done in our lab, indicating the presence of tyrosine phosphorylated proteins.

Western blotting was used to detect tyrosine phosphorylated proteins in *T. thermophila* (materials and methods section 2.4.3). Total cell lysate of *T. thermophila* was collected at different time points in their life cycle (10min, 2hrs and 3hrs into conjugation) and the % of paired cells was determined after the two starved strains (B2086 and CU428) were mixed and collected at those time points (up to 90% pairing occurred after 3hrs of pairing). Total cell lysates of starved cells (B2086 and CU428) and an anti-phosphotyrosine antibody positive control (EGF-stimulated A431 Cell Lysate – EMD Millipore) were also analyzed. Signal was not detected for any of the *T. thermophila* total cell lysates but signal was detected for the positive control indicating that the antiphosphotyrosine antibody (4G10 Platinum-EMD Millipore) was capable of detecting tyrosine phosphorylated proteins (blot not shown). A control experiment using only secondary antibody (anti-mouse antibody HRP tagged) and no primary antibody, didn't show non-specific binding of the secondary antibody. This Western blot method was not capable of detecting tyrosine phosphorylated proteins in *T.thermophila*.

3.1.1 Western blot analysis: SDS-PAGE-MES-Urea

The Western blot protocol was modified to address this issue. 10% SDS-PAGE with an MES buffer system (Tris/2-(*N*-morpholino) ethanesulfonic acid) containing 7M urea (Appendix A) was used. The preparation of the lysate was also modified to be compatible with this type of gel. The pellet of cells was fixed with 10% trichloroacetic acid (TCA) on ice for 30 minutes and then centrifuged two times at 3 minutes each (Sorvall Legend Micro 21R Centrifuge; 24x 1.5/2.0mL Rotor; Thermo Scientific) to remove the supernatant after each spin. The lysis buffer designated as urea lysis solution, containing 7M urea, 2M thiourea, 3% CHAPS and 1% Triton X-100 was added to each pellet (90μL for a pellet from a culture of $1\text{x}10^{5\text{-}6}$cells) to produce a lysate. This lysate could be stored (at 4°C) for up to 24 hours before use. One hour prior to electrophoresis, 20μg (as determined by Bradford analysis) of protein lysate was treated with a final concentration of 2% dithiothreitol (DTT). Samples used were total cell lysate of cells paired for 2 and 3hrs, and individually starved B2086 (see Figure 13). The gel and blot were then processed as in section 2.4.3. Tyrosine phosphorylated proteins are detected (Figure 13). Background signal was however still high using this method. Non-specific interaction from the secondary antibody was not present. This was determined by incubating the membrane only with secondary antibody, no primary antibody, and there were no bands detected (Figure 13c).

3.1.2 Western blot analysis: antibody concentrations

The next step was to further optimize the signal on the blot and that was done by optimizing primary and secondary antibody dilutions using the above method, 10% SDS-PAGE-MES-Urea. The following combination of primary and secondary antibody dilutions were tested respectively: 1:1000 and 1:5000 (Figure 13a); 1:1000 and 1:3000 (Figure 13b); none and 1:3000 (Figure 13c); and 1:500 and 1:3000 (Figure 13d). All lanes contained samples from the same total cell lysate.

The blot treated with primary and secondary antibody dilution of 1:1000 and 1:3000 (Figure 13b) showed one band detected at 16kDa for the *Tetrahymena* total cell lysates (2hrs, 3hrs, B2086). Lowering the secondary antibody dilution to 1:5000 but keeping the primary antibody the same (1:1000), more bands are detected between 16-155 kDa and the background is less noisy (see Figure 13a compared to Figure 13b). However, decreasing the primary antibody dilution to 1:500 but keeping the secondary antibody dilution the same (1:3000), improved signal detection. Various bands were detected between 16 kDa to 155kDa and these bands were stronger than those in Figure 13a. Although the signal improved, this blot also had a higher background than Figure 13a. These experiments were repeated.

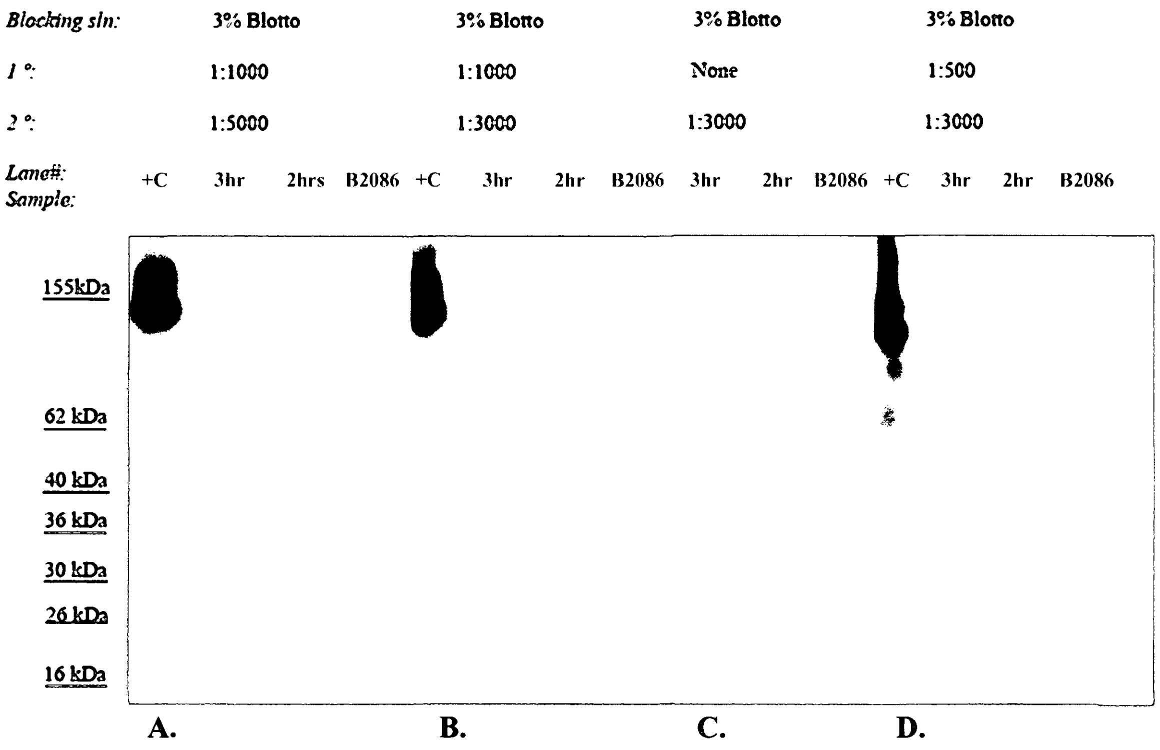


Figure 13. SDS PAGE and Western blot – gel, lysate and antibody dilution

The gel and lysates were prepared using the SDS-PAGE-MES-Urea system and urea lysis solution which showed improvement in signal detection compared to the standard SDS PAGE system and Laemmli lysis solution (see section 3.1). Total cell lysates were collected from starved B2086 and conjugating cells at 2 and 3hours after mixing. +C represents a positive control for the 4G10 Platinum antibody (Millipore) – anti-phosphotyrosine antibody. 3% blotto was used for blocking and HyBlot CL Autoradiography (Denville Scientific) to visualize the signal. Different combinations of primary and secondary antibody dilutions were used to compare as follows (A) 1:1000 and 1:5000, (B) 1:1000 and 1:3000, (C) none and 1:3000, and (D) 1:500 and 1:3000, respectively. The combination used in (D) was picked as best in comparison to the other combinations.

3.1.3 Western blot analysis: antibody choice

To both minimize the background and improve signal detection, this Western blot method (SDS PAGE-MES-Urea) was then combined with a 'direct' detection method by using 4G10 Platinum-HRP tagged (Millipore) at 1:500 dilution. Since this antibody is directly tagged to HRP, there was no need to use a secondary antibody, therefore there were fewer steps in the protocol. This method does show an improvement in background reduction but the signal detection is still weak (see Figure 14a).

3.1.4 Western blot analysis: autoradiography film choice

This was followed by testing a different autoradiographic film (Amersham Hyperfilm ECL-GE Healthcare Life Sciences), instead of the previous one (HyBlot CL Autoradiography-Denville Scientific) to investigate the possibility of signal improvement. In the same experiment as above, the blot was first exposed to HyBlot CL Autoradiography (see Figure 14a) for a minute and this was followed by another exposure using Amersham Hyperfilm ECL for a minute (see Figure 14b). The signal detected using the Amersham Hyperfilm ECL (Figure 14b) such as the ones in the blue box are stronger than those using HyBlot CL film (Figure 14a-red box). This experiment indicates that the Amersham Hyperfilm ECL is better than the HyBlot Cl at detecting signals indicating tyrosine phosphorylated proteins/peptides in *T. thermophila*.

3.1.5 Western blot analysis: amount of sample

To investigate if increasing the amount of total cell lysate would make a difference in improving the signal, the same total cell lysate of B2086 from Figure 14 was treated again using the Western blot protocol with the new updates so far (SDS PAGE-MES-Urea system, 4G10 Platinum-HRP antibody at a 1:500 dilution and Amersham Hyperfilm ECL). This time the amount of sample loaded per well of total cell lysate was 80μg/well (Figure 15) instead of 20μg/well (Figure 14) which generated an improvement in the intensity of the bands especially at about 31kDa, as compared to Figure 14b. There was little effect on the observed intensity of higher molecular weight bands.

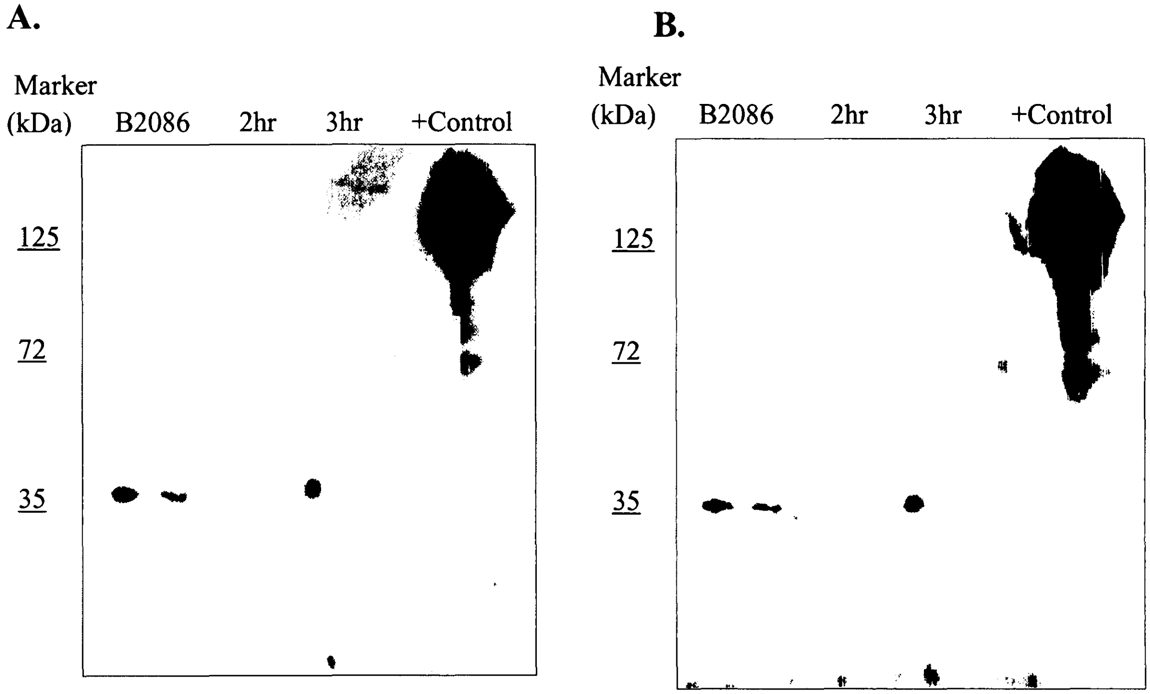


Figure 14. Autoradiography film choice and 4G10 Platinum-HRP

Two autoradiography films were compared, (A) HyBlot CL Autoradiography (Denville Scientific) versus (B) Amersham Hyperfilm ECL (GE Healthcare Life Sciences) to improve signal detection of bands indicating tyrosine phosphorylated proteins. The HyBlot CL Autoradiography has been used regularly in our lab. However the Amersham Hyperfilm ECL was a better choice compared to HyBlot CL Autoradiography as there was less background and stronger signal (blue box) (B) than that in the red box (A). Red box highlights the signal for HyBlot CL Autoradiography and blue box that of the Amersham Hyperfilm ECL. The membrane was first exposed to the film in (A) for 1min and then it was exposed to (B) for 1min. Starved B2086 and pairing of the two mating types (at 2 and 3hrs) total cell lysate were tested at 20μg/well. The anti-phosphotyrosine antibody used here was directly linked to HRP (4G10 Platinum-HRP; Millipore).

3.1.6 Western blot analysis: different lysis solutions

Finally, following the above method a different approach was taken by comparing different lysis solutions. The following modifications were made to the standard urea lysis solution described above: 1) included 20 % glycerol; 2) thiourea was excluded and 20 % glycerol was included; 3) thiourea was excluded; 4) thiourea was excluded, urea concentration was reduced to 2M and 20% glycerol was included; 5) urea lysis solution, no changes (control); 6) lysate was incubated at 37°C when treated with final concentration of 2% DTT; 7) contained cells lysed with Laemmli buffer including 7M urea; 8) mass spectrometry grade urea (Cell Signaling Technology) was used. In each lane, 80μg of lysate was loaded.

Lanes 1-4 and 6-8 are compared to lane 5 (urea lysis solution - control) (Figure16) except when stated otherwise. The bands on lane 1 show that the addition of glycerol to the urea lysis solution further preserves bands seen in lane 5. The exclusion of thiourea (lane 3) in the urea lysis buffer leads to the loss of bands at 42 kDa and higher. The same occurs when thiourea is excluded and glycerol is included (lane 2). This shows the importance of both these ingredients in the stability of the signal. In lane 4, excluding thiourea, including glycerol and reducing urea concentration to 2M leads to loss of all bands detected in lane 5. Everything for lane 6 was kept the same as in lane 5 except being placed at 37°C when treated with DTT, which made no difference in preserving the bands. In lane 8 everything was the same as in lane 5 except it contained mass spectrometry grade urea (Urea Mass Spectrometry Grade - Cell Signaling Technology), this lead to loss of all bands except two at 57 and 72 kDa. Finally, in lane 7 a different approach was taken, the lysis solution was Laemmli buffer with 7M urea where there was a stronger preservation of the bands compared to those seen in lane 5. The lysis buffer used in lane 7 (Laemmli buffer + 7M urea) was chosen as the optimum lysis solution among those tested.

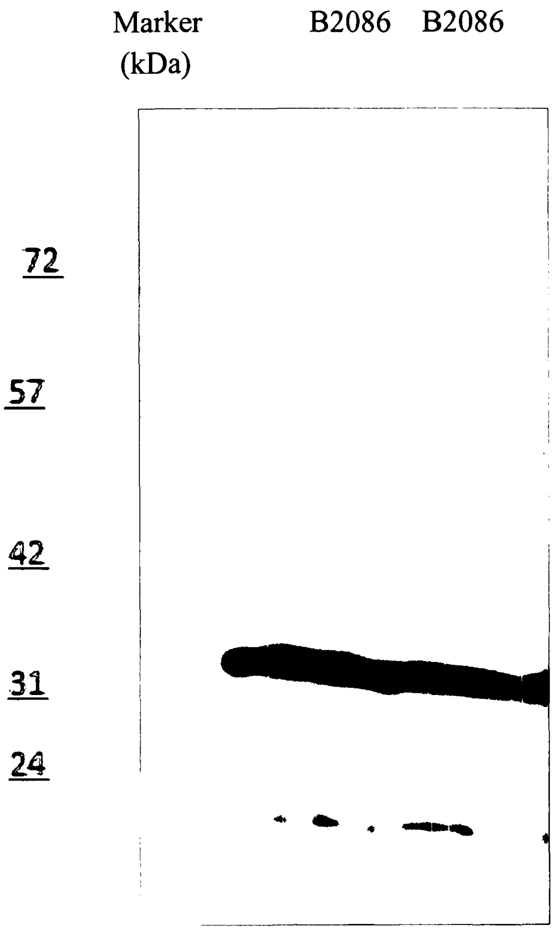
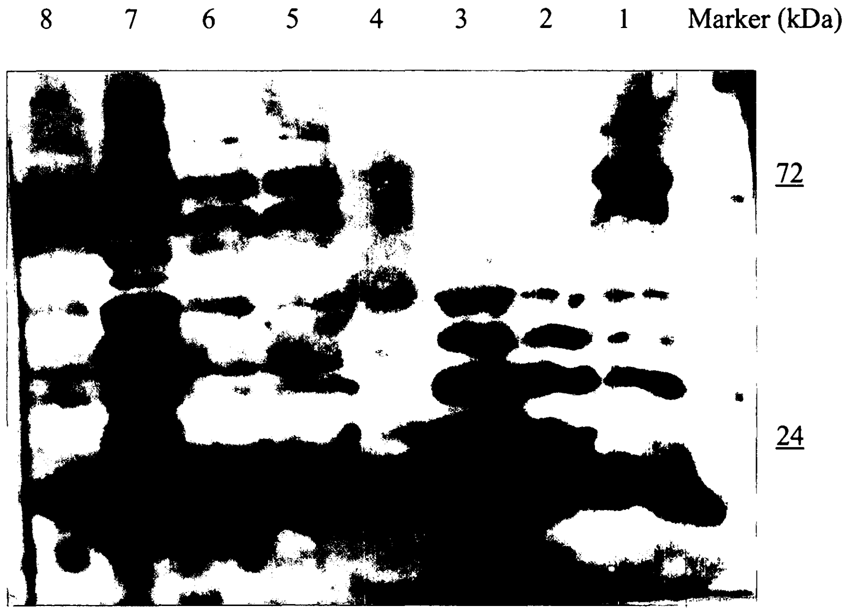


Figure 15. Amount of sample

Both lanes contain total cell lysate from starved B2086 at 80μg/well. This was from the same total cell lysate as in Figure 14. Western blot method to detect tyrosine phosphorylated proteins was done using an anti-phosphotyrosine antibody (4G10 Platinum-HRP; Millipore).



LEGEND	
Lane #	**Lysis solution changes from the control**
1	Included 20% glycerol
2	Thiourea excluded; included 20% glycerol
3	Thiourea excluded
4	Thiourea excluded, urea concentration reduced to 2.0M; included 20% glycerol
5	Urea lysis solution (control)
6	As in lane #5 but lysate incubated at 37°C
7	Only Laemmli buffer with 7M urea
8	Mass spectrometry grade urea

Figure 16. Lysis solution modification

Modifications were done to the lysis solution to lyse the total cell lysate for the Western blot method. The control is the lysis solution in lane #5 and the changes are as listed in the legend. Tyrosine phosphorylated proteins were detected using an anti-phosphotyrosine antibody.

3.2 Tyrosine phosphorylation signaling during pairing

All these optimizations led to the development of a specific Western blot protocol to analyze tyrosine phosphorylation in *T. thermophila*. Figure 17a shows a clear blot and sharp preserved bands by using the optimized protocol. This allowed for a clear analysis of tyrosine phosphorylation signaling throughout conjugation from 10min to 4hrs (90% pairing) and of single starved cells. Analysis using ImageJ 1.44P was done to compare which bands have a stronger signal between the bands at that particular molecular weight. In other words, this determines the time point (10min, 2hrs, 3hrs or 4hrs) the signal is stronger for these proteins. The higher the relative density the stronger the signal is detected by 4G10-HRP Platinum (antiphosphotyrosine antibody; Millipore) in the total cell lysate (Table 2 and Figures 17-18). The positive control (+control – Figure 17a) indicates the specificity for 4G10-HRP Platinum to detect tyrosine phosphorylated proteins. At 10min of cell pairing, there is an increase in tyrosine phosphorylation signal detected compared to single and starved cells (B2086 and/or CU428), as seen in most of the bands at this time (#1, 4, 6-9 – Figure 17a; Figure 18). By 3hrs the signal drops as seen in most of the bands at this time (#1, 3-5, 8, 9 – Figure 17; Figure 18). By 4hrs we see an increase in signal as compared to 3hrs. This is shown by most of the bands at 4hrs (#1-3, 6, 7 – Figure 17; Figure 18). The same trend is also seen when observing the average relative density (three Western blot trials including the one in Figure 17). The average relative shows that most of the bands also indicate an increase at 10min as compared to starved cells, decrease by 3hrs as compared to 10min, and increase after 4hrs into conjugation.

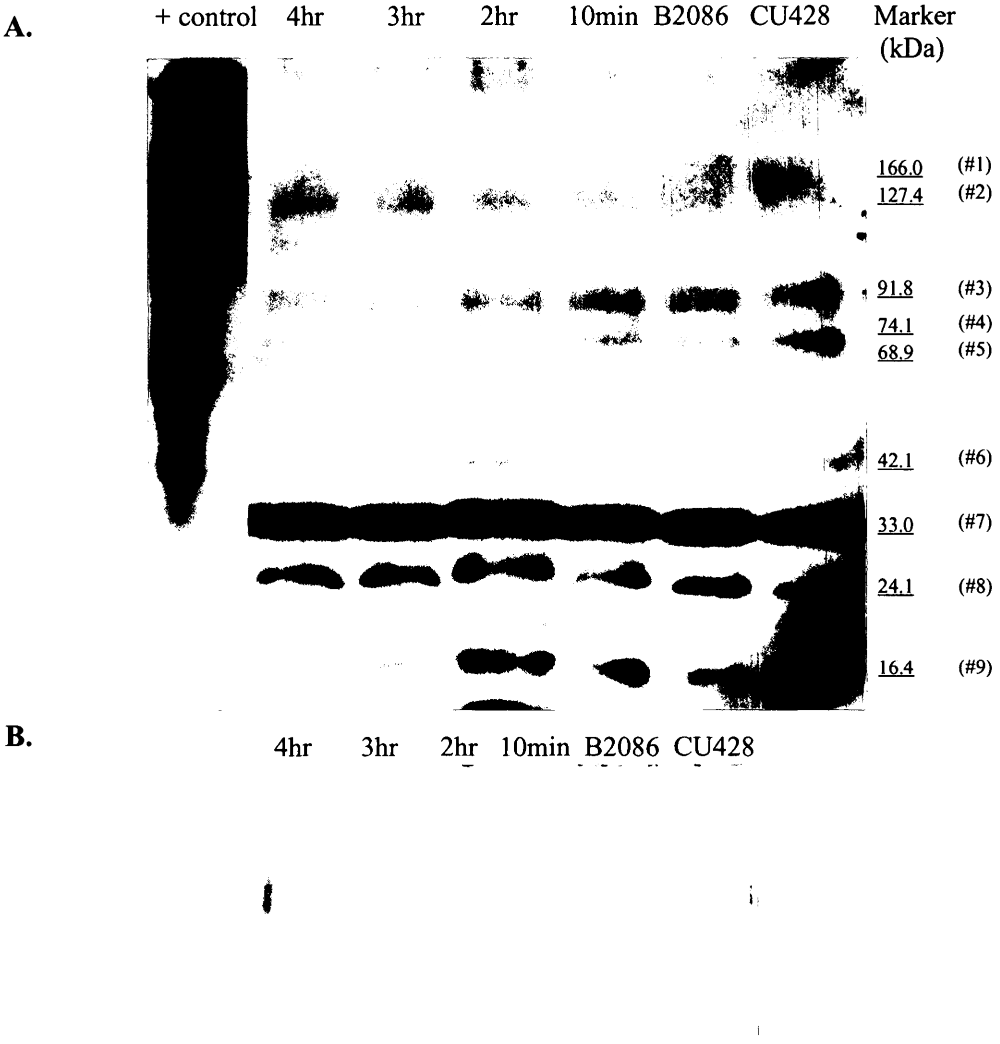


Figure 17. Tyrosine phosphorylation detection during pairing

A) The lanes contain total cell lysate collected from starved B2086 and CU428, and of the mating types after pairing for 10min, 2hrs, 3hrs and 4hrs. This is the optimized Western blot method (SDS-PAGE-MES-Urea; Amersham Hyperfilm ECL; Laemmli buffer +7M urea) to detect tyrosine phosphorylated proteins using an anti-phosphotyrosine antibody (4G10 Platinum HRP; Millipore). The +control represents EGF-treated human A431 carcinoma cells (Millipore) used as a positive control for the specificity of the antibody. B) This is a commassie stained gel of a SDS-PAGE-MES-Urea gel containing total cell lysate as in (A). Equivalent amount of protein was loaded per well as calculated by Bradford analysis and visualized here.

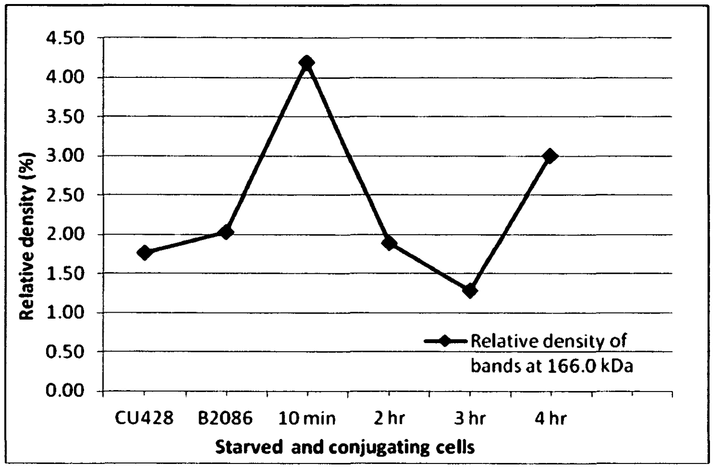
Table 2. Relative density of signals in Figure 17a and average relative density.

ImageJ 1.44P was applied to Figure 17a to determine relative density of each band compared to each other at that specific molecular weight. Relative density was calculated by dividing each percent with the lowest percent (%) for the bands at that particular size. This determines the time point (starved, 10min, 2hrs, 3hrs or 4hrs into conjugation) the signal is stronger for the proteins at that specific molecular weight. The average relative density column contains data calculated from three experimental trials. The relative density, area and percent column contain data calculated using the bands from Figure 17a.

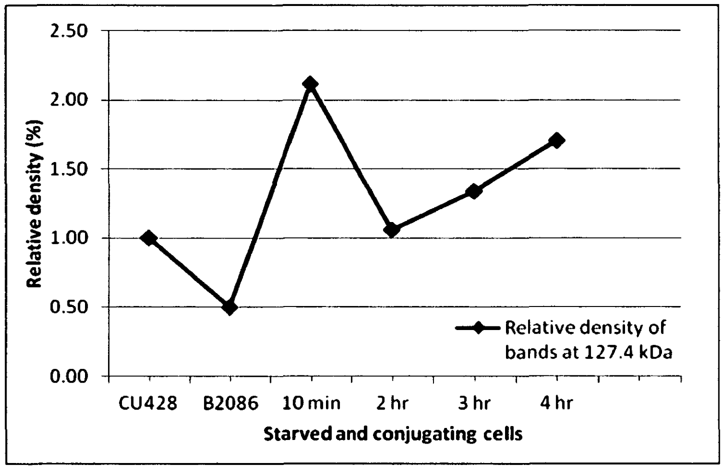
Total cell lysate	**Area**	**Percent (%)**	**Relative density (%)**	**Average relative density (%)**
Bands at 166.0 kDa				
CU428	-	-	0	1.76
B2086	1778.32	9.49	1.00	2.03
10 min	2149.10	11.47	1.21	2.19
2 hr	3557.56	18.98	2.00	1.89
3 hr	3296.82	17.59	1.85	1.28
4 hr	4202.22	22.42	2.36	3.00
Bands at 127.4 kDa				
CU428	-	-	0	1.00
B2086	-	-	0	0.50
10 min	-	-	0	2.12
2 hr	-	-	0	1.06
3 hr	1478.61	20.88	1.00	1.34
4 hr	2414.17	34.10	1.63	1.70
Bands at 91.8 kDa				
CU428	10497.39	19.66	2.14	3.54
B2086	10402.97	19.49	2.12	3.30
10 min	10019.21	18.77	2.04	2.42
2 hr	5771.89	10.81	1.18	2.12
3 hr	4903.73	9.19	1.00	1.00
4 hr	5202.92	9.75	1.06	3.11

Total cell lysate	**Area**	**Percent (%)**	**Relative density (%)**	**Average relative density (%)**
Bands at 74.1 kD				
CU428	6604.66	56.30	8.78	2.76
B2086	1471.89	12.55	1.96	2.72
10 min	2902.88	24.74	3.86	2.90
2 hr	-	-	0	1.46
3 hr	-	-	0	0.78
4 hr	-	-	0	1.40
Bands at 68.9 kDa				
CU428	-	-	0	1.34
B2086	1201.23	33.66	1.00	1.62
10 min	-	-	0	0.33
2 hr	-	-	0	0.88
3 hr	-	-	0	2.42
4 hr	-	-	0	0.89
Bands at 42.1 kDa				
CU428	3020.81	6.93	1.00	1.08
B2086	6028.80	13.83	2.00	1.48
10 min	4863.25	11.16	1.61	1.46
2 hr	6550.37	15.03	2.17	1.75
3 hr	7267.32	16.68	2.41	1.52
4 hr	8005.02	18.37	2.65	1.71

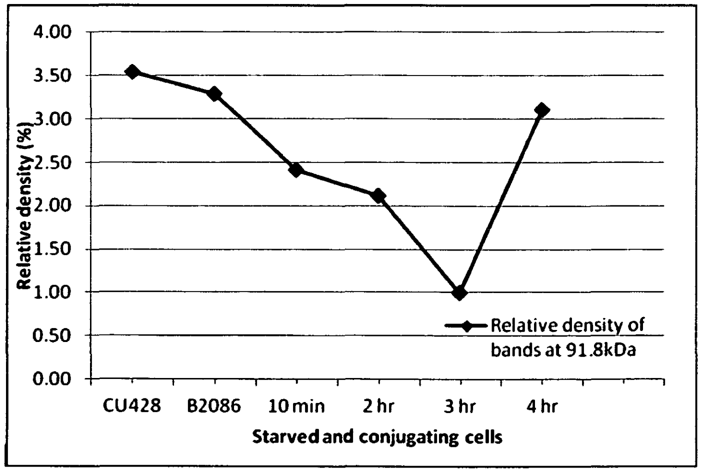
Total cell lysate	**Area**	**Percent (%)**	**Relative density (%)**	**Average relative density (%)**
Bands at 33.0 kDa				
CU428	13757.97	10.86	1.00	1.24
B2086	18370.95	14.50	1.34	1.55
10 min	16434.10	12.97	1.19	1.33
2 hr	18058.73	14.25	1.31	1.25
3 hr	19063.95	15.04	1.39	1.15
4 hr	19649.48	15.51	1.43	1.44
Bands at 24.1 kDa				
CU428	4841.55	9.27	1.00	2.10
B2086	11374.92	21.78	2.35	3.75
10 min	6258.99	11.99	1.29	1.76
2 hr	8495.97	16.27	1.75	1.94
3 hr	8115.02	15.54	1.68	0.89
4 hr	7501.37	14.36	1.55	1.84
Bands at 16.4 kDa				
CU428	2105.34	4.42	1.00	2.02
B2086	5657.41	11.87	2.69	1.03
10 min	10132.55	21.27	4.81	2.27
2 hr	13706.89	28.77	6.51	0.57
3 hr	6047.63	12.69	2.87	2.26
4 hr	4829.68	10.14	2.29	0.84


4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
Relative density (%)
Relative density of bands at 166.0 kDa
CU428
B2086
10 min
2 hr
3 hr
4 hr
Starved and conjugating cells

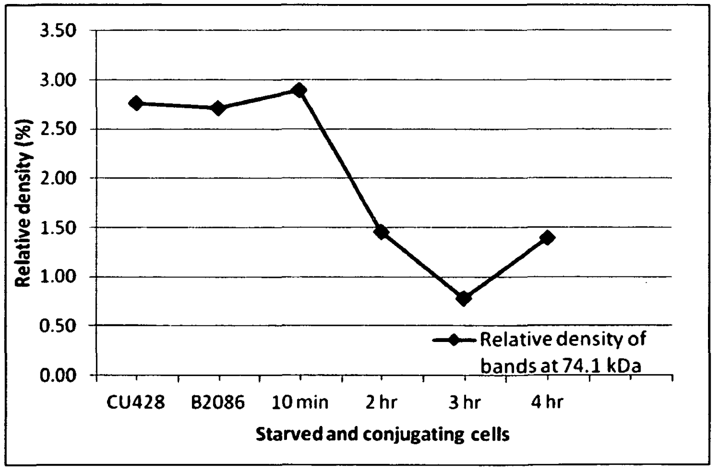
A.


2.50
2.00
1.50
1.00
0.50
0.00
Relative density (%)
Relative density of bands at 127.4 kDa
CU428
B2086
10 min
2 hr
3 hr
4 hr
Starved and conjugating cells

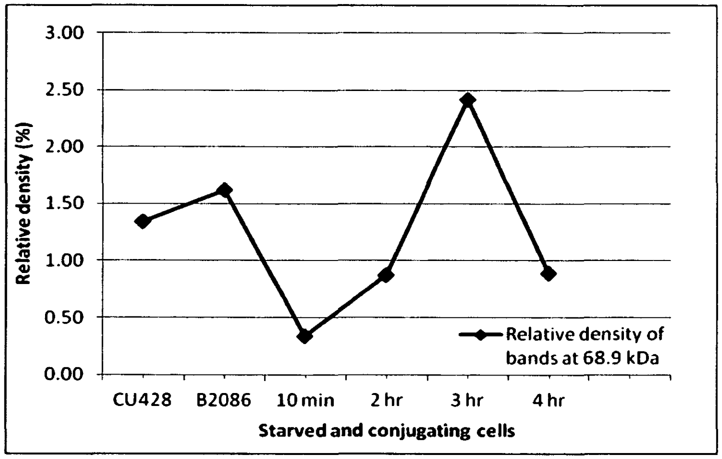
B.


4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
Relative density (%)
Relative density of bands at 91.8kDa
CU428
B2086
10 min
2 hr
3 hr
4 hr
Starved and conjugating cells

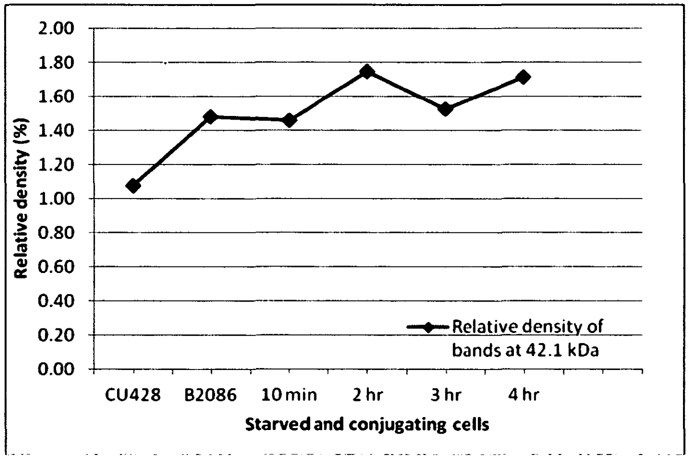
C.


3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
Relative density (%)
Relative density of bands at 74.1 kDa
CU428
B2086
10 min
2 hr
3 hr
4 hr
Starved and conjugating cells

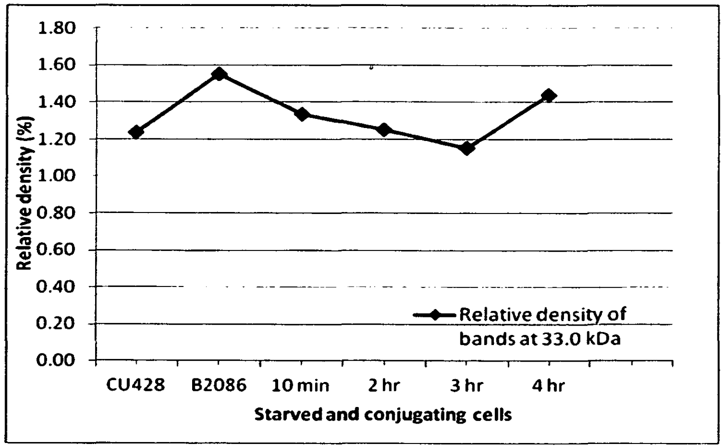
D.


3.00
2.50
2.00
1.50
1.00
0.50
0.00
Relative density (%)
CU428
B2086
10 min
2 hr
3 hr
4 hr
Starved and conjugating cells
Relative density of bands at 68.9 kDa

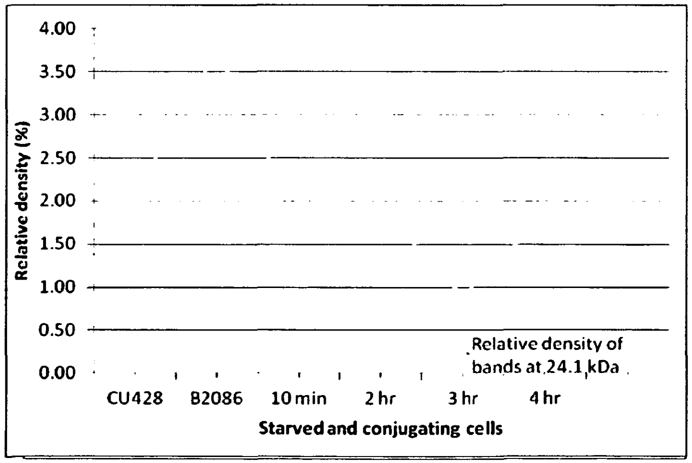
E.


2.00
1.80
1.60
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
Relative density (%)
CU428
B2086
10 min
2 hr
3 hr
4 hr
Starved and conjugating cells
Relative density of bands at 42.1 kDa

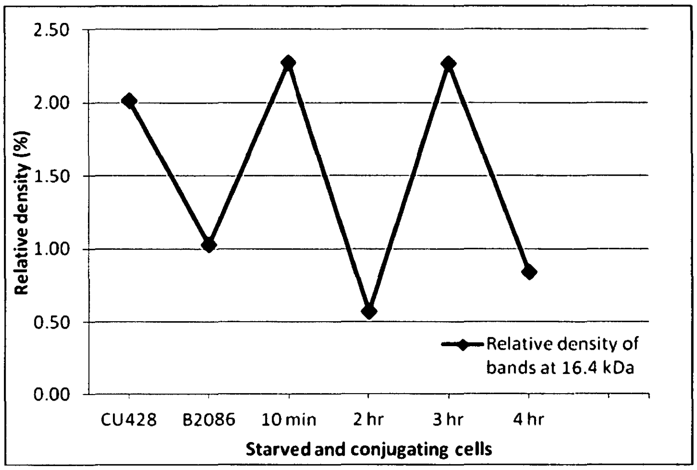
F.


1.80
1.60
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
Relative density (%)
CU428
B2086
10 min
2 hr
3 hr
4 hr
Starved and conjugating cells
Relative density of bands at 33.0 kDa

G.


4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
Relative density (%)
CU428
B2086
10 min
2 hr
3 hr
4 hr
Starved and conjugating cells
Relative density of bands at 24.1 kDa

H.



I.

Figure 18. Relative density represented in graphs

The average relative density from Table 2 is represented in graphs. This is to determine the time point (starved, 10min, 2hrs, 3hrs or 4hrs into conjugation) the signal is stronger for the proteins at that specific molecular weight. Samples used were starved CU428 and B2086 and pairing. Pairing samples were collected at 10 min, 2hrs, 3hrs and 4hrs.

3.3 Immunofluorescence microscopy

Indirect immunofluorescence microscopy was carried out to confirm previous results done in our lab using the same protocol and to support the Western blot results. Figure 19 shows a tyrosine phosphorylation signal concentrated at the posterior end of a cell in a conjugating pair collected at 3hrs (90% of the cells were paired). The negative control didn't show non-specific binding of the secondary antibody. This confirms the results observed in our lab previously and further supports the Western blot results.

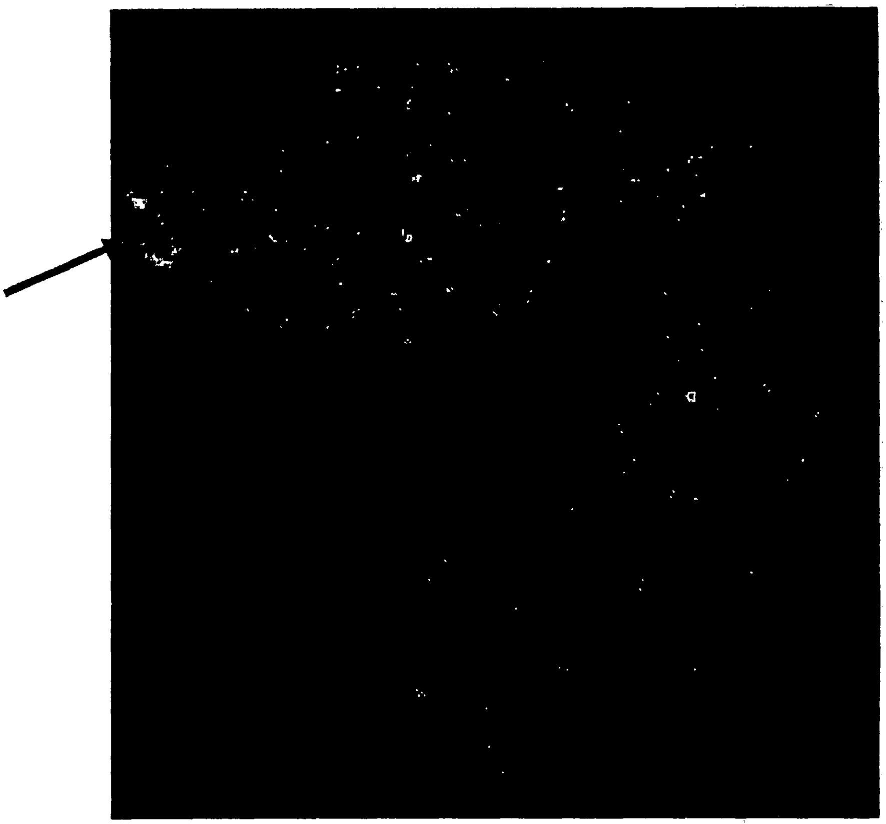


Figure 19. Indirect immunofluorescence microscopy results

Indirect immunofluorescence microscopy used to detect tyrosine phosphorylated proteins in the two mating types after 3hrs of pairing. The primary antibody was an anti-phosphotyrosine antibody (4G10 Platinum; Millipore). The arrow represents a concentrated region of tyrosine phosphorylation at the posterior region of the cell.

3. 4 Bioinformatic analysis

To also determine the presence of tyrosine phosphorylation signaling in *T. thermophila*, bioinformatic analysis was performed. The BLAST results obtained from comparing the *T. brucei* proteins (amino acid sequence) from Table 1 and 2 from Nett et al. (2009) to a database of *Tetrahymena* protein sequences (http://www.cilate.org), showed that there was similarity between *T. brucei* and *T. thermophila* proteins (Appendix B-1). Many of the matches in *Tetrahymena* had similar annotations/description to the *Trypanosoma* protein kinase domains, phosphofructo kinase, phosphoenolpyruvate carboxykinase, MPK's (mitogen activated protein kinase), GSK (glycogen synthase kinase), and ERK (extracellular signal regulated kinases).

Further alignment analysis between each peptide and the corresponding *Tetrahymena* protein amino acid sequence (TTHERM_) (TTHERM_'s from Appendix B-2) using ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/) (all alignments in Appendix B-3) were done. These alignments indicated that the tyrosine residues that were detected as being phosphorylated in *T. brucei* were conserved in the *Tetrahymena* match, as the examples shown in Figure 20 (c and g) (yellow highlight). The only comparison that didn't show conservation of the tyrosine sites was between Hypothetical protein (Tb927.4.3130) and TTHERM_00161270 (Appendix B-3).

Analysis using Netphos 2.0 (http://www.cbs.dtu.dk/services/NetPhos/) was used to detect putative high probability tyrosine phosphorylation sites for each *Tetrahymena* protein sequence from Appendix B-2 (Figure 21). All the NetPhos scores are in Appendix B-4. The system considers those with scores of 0.500 and higher as likely being putative tyrosine phosphorylation sites. For example, the tyrosine phosphorylation prediction for TTHERM_0483640 had a score of 0.569 for the *IGEGSYGVVF* which corresponds to the peptide *IGEGSYGVVFR* (CRK1) of *T. brucei* (Figure 20c and 21a). NetPhos results of the other TTHERM_'s for the corresponding 'Y' sites to *T. brucei* peptides that had scores of 0.500 or higher ranged between 0.647 - 0.963 (Appendix B-3, NetPhos score in brackets; or Appendix B-4). These peptides are likely to become tyrosine phosphorylated at that specific tyrosine site. This range of scores includes some of the TTHERM_'s that correspond to CRK1, 2 and 3; MAPK1, 9 and 11; GSK3; PK's, ECK1 and TbPFK ATP-dependent phosphofructokinase of *T. brucei* (Appendix B-3; NetPhos score in brackets).

The alignment results (ClustalW2) that showed matches between the two organisms but have low NetPhos scores were between MAPK2 (Tb10.70.2070) and TTHERM_00537240; PK (Tb11.02.0640) and TTHERM_00348940; ribosomal protein S27 and TTHERM_01043090; glycosomal phosphoenolpyruvate carboxykinase and TTHERM_00420030; ATP-dep DEAD/H RNA helicase and TTHERM_00190830; and hypothetical protein (Tb09.160.1160) and TTHERM_00827160. Therefore, 6 of the 25 alignments that had matching tyrosine sites, had low NetPhos scores for that specific *Tetrahymena* protein tyrosine site.

Expression microarray data (obtained from http://tfgd.ihb.ac.cn/ and shown in Appendix B-5) showed the expression profile of the *Tetrahymena* genes that corresponded to each *Tetrahymena* protein. For example, the data for TTHERM_00469230 (MPK1) shows that there is a high rise in expression (6048.04 to 28610.54 microarray score) of this gene during conjugation between 2 to 4hrs with a lower expression at other times in the life cycle (Figure 20h), indicating the possibility of this gene being expressed specifically during conjugation. Microarray results for TTHERM_00483640 (CDC2 PK which we will refer to TtCDC2 here) shows this gene as also having an expression peak during conjugation at 2hrs with a microarray score of 39656.63 and later at a lower expression. The *Tetrahymena* genes that show expression during conjugation between 0 to 4hrs includes the genes listed in Appendix B (page 104). The rest of the genes do show expression peaks during conjugation but at later time points (Appendix B-5). Therefore the bioinformatic analysis further supports the possibility of putative tyrosine phosphorylation signaling occurring in *T. thermophila* during conjugation.

A.

T. brucei	*Tetrahymena*
	Smallest Sum High Probability Sequences producing High-scoring Segment Pairs: Score P(N) N
Crk1 (TB10.70.7 040) Corresponding peptide: IGEGSYG VVFR	TTHERM_00483640 Protein+kinase+domain; alias=58.m00270,T0... 691 3.3e-69 1 TTHERM_00286770 Protein+kinase+domain; alias=26.m00259,T0... 670 5.5e-67 1 TTHERM_00784290 Protein+kinase+domain; alias=131.m00106,T... 624 4.1e-62 1 TTHERM_00318700 cell+division+control+protein; note=simil... 481 7.3e-55 2 TTHERM_00066860 Protein+kinase+domain; alias=4.m00575,T00... 532 2.3e-52 1 TTHERM_00185770 Protein+kinase+domain; alias=14.m00466,T0... 514 4.1e-50 1 : :

B.

TTHERM_00483640

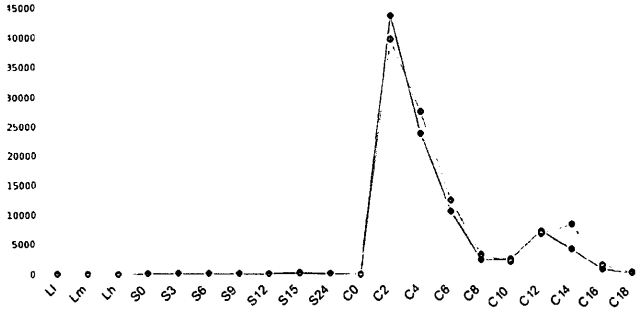
MENLENLETIEDRYQKIKKIGEGTYGKVFQAKDLKTDQIVALKKVKNDYEEIGDEGIPSTALREISCLKALDHPNVVKLVDVVYIMKKNKLYLVFEYIDYDLKA
YQKKIGKIPEQTVKSYMHQILKGIEHCHSRRIFHRDLKPQNILVNNKGDIKIADFGLGKIFGIPLNTITHEVETLWYRAPEILLGDKLYSLPVDVWSIGCIFAELII
GHPLFHGDSEIDQIFKIFQFFGTPKSSKLEGTYNLKYWSNLFPRFKSQKDEILKPIIETDPQAADLLVKLIDIEPAKRISVSQALKHPYFDNIQSPLNSQIFNF

C.

TTHERM_00483640 and *IGEGSYGVVFR* peptide

```
pep      -------------------IGEGSYGVVFR------------------------------ 11
pro      MENLENLETIEDRYQKIKKIGEGTYGKVFQAKDLKTDQIVALKKVKNDYEEIGDEGIPST 60
                            ****:** **:
```

D.



Legend:

- Vegetative cells (L) collected at ~1X10^5 cells/ml, and ~3.5X10^{5}cells/ml and ~1*10^6 cells/ml.
- Starved cells (S) collected between 0-18 hrs.
- Conjugating cells (C) collected between 0-18 hrs.

E.

T. brucei	*Tetrahymena*			
	Sequences producing High-scoring Segment Pairs:	High Score	Smallest Sum Probability P(N)	N
Mapk11-Tb10.389.1730 Peptide: EESDQGEH MTDYVTMR	TTHERM_00469230 MPK1%3A%20Putative%20mitogen%2Dactivated%...	800	9.2e-81	1
	TTHERM_00195990 Protein+kinase+domain; alias=16.m00440,T0...	799	1.2e-80	1
	TTHERM_00760270 Protein+kinase+domain; alias=125.m00076,T...	767	2.9e-77	1
	TTHERM_00689900 Protein+kinase+domain; alias=105.m00153,T...	728	3.9e-73	1
	TTHERM_00760190 MPK2%3A%20Putative%20mitogen%2Dactivated%...	620	8.2e-72	2
	TTHERM_00660130 MPK3%3A%20Putative%20mitogen%2Dactivated%...	383	3.1e-64	2
	: :			

F.

TTHERM_00469230

MQQQQNQQQQAQQQQLQQQQQQQQQHQQAQQHQEKPQKPIQVPPGKHKFDCGQTIFVVDEKYEYIKQIGYGAYGVVCSAVNKENGSKVAIKKVQNAYDDLIDAKRIVR
EIKLLKFFEHDNIISLLDVIKPNAPVGDEDIYMVFDLMETDLHRVIYSRQELSDEHIQYFVYQILRGLLYIHSANVIHRDLKPSNILLNKNCDLKICDFGLARGFEEPGESLTEYVITRWY
RAPEVILNASEYNQAVDIWSVGCITAELLGRTPLFPGENYLDQVQRVIAVLGTPTQEDMAYIGNQDAIKYIKSQPKRTKQSWQSLYPKANPLALDLLSKMLTFNPEKRYTVKECIEH
PYFEGLHYPDEEPECEQVFDWSWDKFTPTKEILQKMIYQEAKSFHPIH

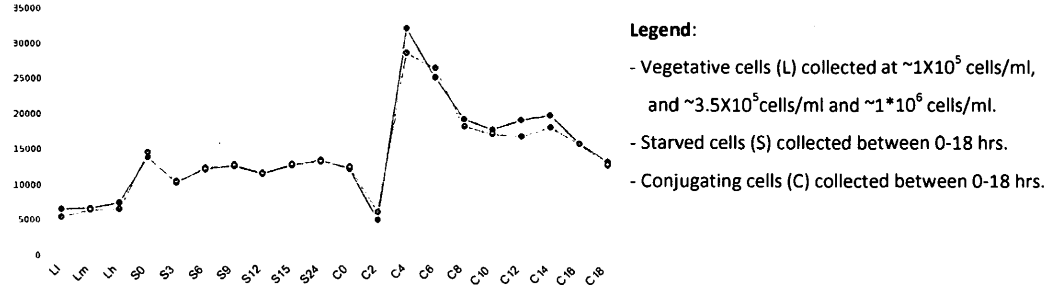
G.

TTHERM_00469230 and *EESDQGEHMTDYVTMR* peptide

```
pep              --------------------------------------EESDQ----------------- 5
pro              IISLLDVIKPNAPVGDEDIYMVFDLMETDLHRVIYSRQELSDEHIQYFVYQILRGLLYIH 180
                                                       *  **.

pep              ------------------------------------GEHMTDYVTMR------------- 16
pro              SANVIHRDLKPSNILLNKNCDLKICDFGLARGFEEPGESLTEYVITRWYRAPEVILNASE 240
                                                     **  :*.**   *
```

H.



Legend:

- Vegetative cells (L) collected at ~1X10^5 cells/ml, and ~3.5X10^{5}cells/ml and ~1*10^6 cells/ml.
- Starved cells (S) collected between 0-18 hrs.
- Conjugating cells (C) collected between 0-18 hrs.

Figure 20. Bioinformatic results summary (part I)

A sample of the results obtained by following the flow chart from Figure 10, to compare the *T. brucei* peptides from Table 1 and 2 of Nett et al., (2009) to *Tetrahymena*. A and E) The BLAST results obtained by submitting a *T. brucei* protein amino acid sequence (CRK1 and MAPK11) from Table 1 and 2 of Nett et al., (2009) to *Tetrahymena* Genome Data (http://ciliate.org/), respectively. B and F) The amino acid sequence of TTHERM_0483640 and TTHERM_00469230, respectively. C and G) Alignment of the TTHERM_00483640 and TTHERM_00469230 amino acid sequences to the corresponding peptides using ClustalW (www.ebi.ac.uk/clustalw), respectively. The yellow highlight on Y shows the conserved tyrosine sites between the two organisms. The red oval shape shows a conserved tyrosine site between the two organisms and a NetPhos 2.0 score higher than 0.500 for that specific tyrosine site. D and H) The microarray expression data represented in a graph for TTHERM_00483640 and TTHERM_00469230 from http://tged.ihb.ac.cn/ , respectively.

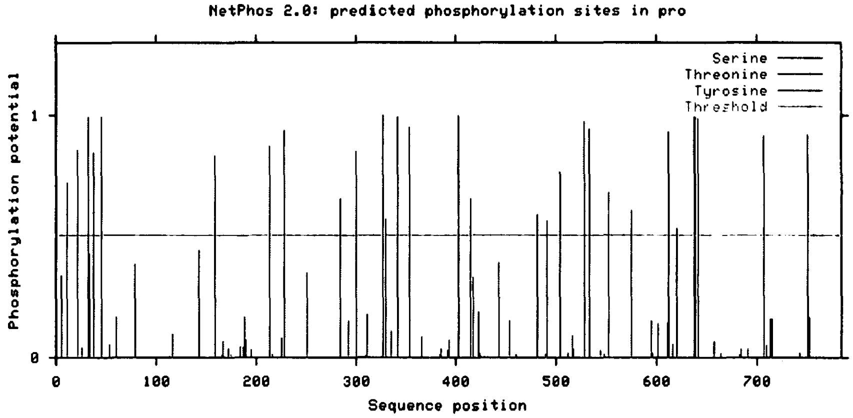
A.

TTHERM_00483640

Pos	Context	Score	Pred
14	IEDRYQKIK	0.953	*Y*
25	GEGTYGKVF	0.569	*Y*
49	VKNDYEEIG	0.980	*Y*
84	VDVVYIMKK	0.315	.
92	KNKLYLVFE	0.140	.
97	LVFEYIDYD	0.254	.
100	EYIDYDLKA	0.383	.
105	DLKAYQKKI	0.021	.
120	TVKSYMHQI	0.010	.
181	ETLWYRAPE	0.048	.
193	GDKLYSLPV	0.052	.
244	LEGTYNLKY	0.217	.
248	YNLKYWSNL	0.913	*Y*
300	LKHPYFDNI	0.256	.

Matches the corresponding peptide sequence in Figure 20c.

B.



C.

TTHERM_00469230

Pos	Context	Score	Pred
62	VDEKYEYIK	0.974	*Y*
64	EKYEYIKQI	0.358	.
70	KQIGYGAYG	0.213	.
73	GYGAYGVVC	0.056	.
97	VQNAYDDLI	0.198	.
140	DEDIYMVFD	0.983	*Y*
155	HRVIYSRQE	0.038	.
167	EHIQYFVYQ	0.057	.
170	QYFVYQILR	0.026	.
178	RGLLYIHSA	0.441	.
223	SLTEYVITR	0.923	*Y*
229	ITRWYRAPE	0.007	.
241	NASEYNQAV	0.288	.
269	PGENYLDQV	0.522	*Y*
290	EDMAYIGNQ	0.581	*Y*
299	DAIKYIKSQ	0.631	*Y*
315	WQSLYPKAN	0.153	.
338	PEKRYTVKE	0.665	*Y*
348	IEHPYFEGL	0.256	.
354	EGLHYPDEE	0.454	.
384	QKMIYQEAK	0.244	.

Matches the corresponding peptide sequence in Figure 20g.

D.

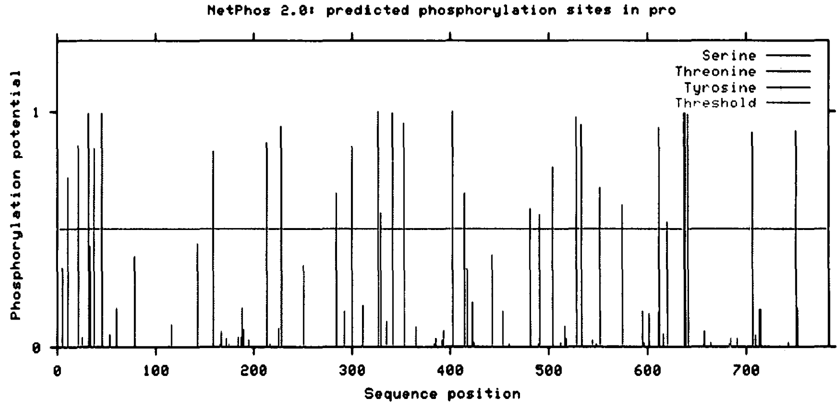


Figure 21. Bioinformatic results summary (part II)

A sample of the results obtained by following the flow chart from Figure 10. A and C) The amino acid sequence of TTHERM_0483640 and TTHERM_00469230 , respectively submitted to NetPhos 2.0 (http://www.cbs.dtu.dk/services/NetPhos/) to predict phosphorylated tyrosine sites. Scores above 0.500 show that there is a significant probability that this is a tyrosine site that could become phosphorylated. The higher the score the higher the chance for this to occur. The red box shows the context that corresponds to the peptide *IGEGSYGVVFR* (CRK1) and *EESDQGEHMTDYVTMR* (MAPK11) of *T. brucei* from Table 1 and 2 of Nett et. al., (2009), respectively. B and D) The graphs show the predicted tyrosine, serine and threonine phosphorylation sites (sequence position versus phosphorylation potential) for TTHERM_0483640 and TTHERM_00469230, respectively.

3.5 TtCDC2

The protein product, TtCdc2, of the gene TtCDC2 in *Tetrahymena*, was determined by bioinformatic analysis as possibly having tyrosine phosphorylation sites. To investigate the role of this gene in *Tetrahymena* and its relevance to tyrosine phosphorylation signaling, the first step was to create null mutants. This was done by making a knockout construct in which the coding region of the gene was replaced.

3.5.1 TtCDC2 gene replacement

The DNA from *Tetrahymena* cells was used as the basis to create 5' and 3' flanks of TtCDC2. The *Tetrahymena* genome is about 104Mb, therefore larger than the ladder as confirmed in Figure 22a. The primers #1-4 (Figure 11) that corresponded to 5' and 3' UTR of TtCDC2 from the *Tetrahymena* genome were used to amplify the corresponding regions (flanks) (Figure 22b). Figure 22b (left) shows the presence of 3' flank (600bp) and Figure 22b (right) shows the presence of 5' flank (480bp) in the PCR product as expected. pNeo4 was digested with the enzymes as in Figure 23a. Also, the flanks were digested with the corresponding enzymes and then inserted into the plasmid pNeo4 at the sites indicated (enzyme sites) in Figure 23a.

The 3' flank was first inserted into pNeo4. Figure 23b (lane 1) confirms this by showing 3'flank (0.60kb) released from pNeo4 (5.00kb) when digested with SacI and SacII. Then the 5'flank was inserted forming a full construct (6.08kb) as shown in lane 2 (Figure 23b). The complete construct was also digested with SacI and ApaI to show the flanks ligated to the neo4 cassette (3.41kb) as shown in lane 3 (Figure 23b). The complete construct was also confirmed by sequence analysis (York University Core Molecular Biology DNA Sequencing Facility). The complete construct was then inserted into *Tetrahymena* to replace the TtCDC2 region with the neo4 cassette using biolistic transformation.

The construct was incorporated into the *Tetrahymena* genome through homologous recombination (Figure 23c). The transformants were selected using drug selection (paromomycin sulfate). The transformants were confirmed by amplifying the region as indicated in Figure 24a by PCR, using their DNA as template and primers (Figure 11-primers #4 and 5) that amplify a part of the neo4 cassette at the BTU2 region and the 3'TtCDC2. The results indicate a band at the expected size, 2039bp for the transfomants and no band for the wild type (Figure 24a-b). The primers # 4 and 6 (Figure 11), were also used together to confirm transformants. The results indicate a band at the expected size for the transformants, 2749bp, and 1765bp for the wild type (Figure 24c-d).

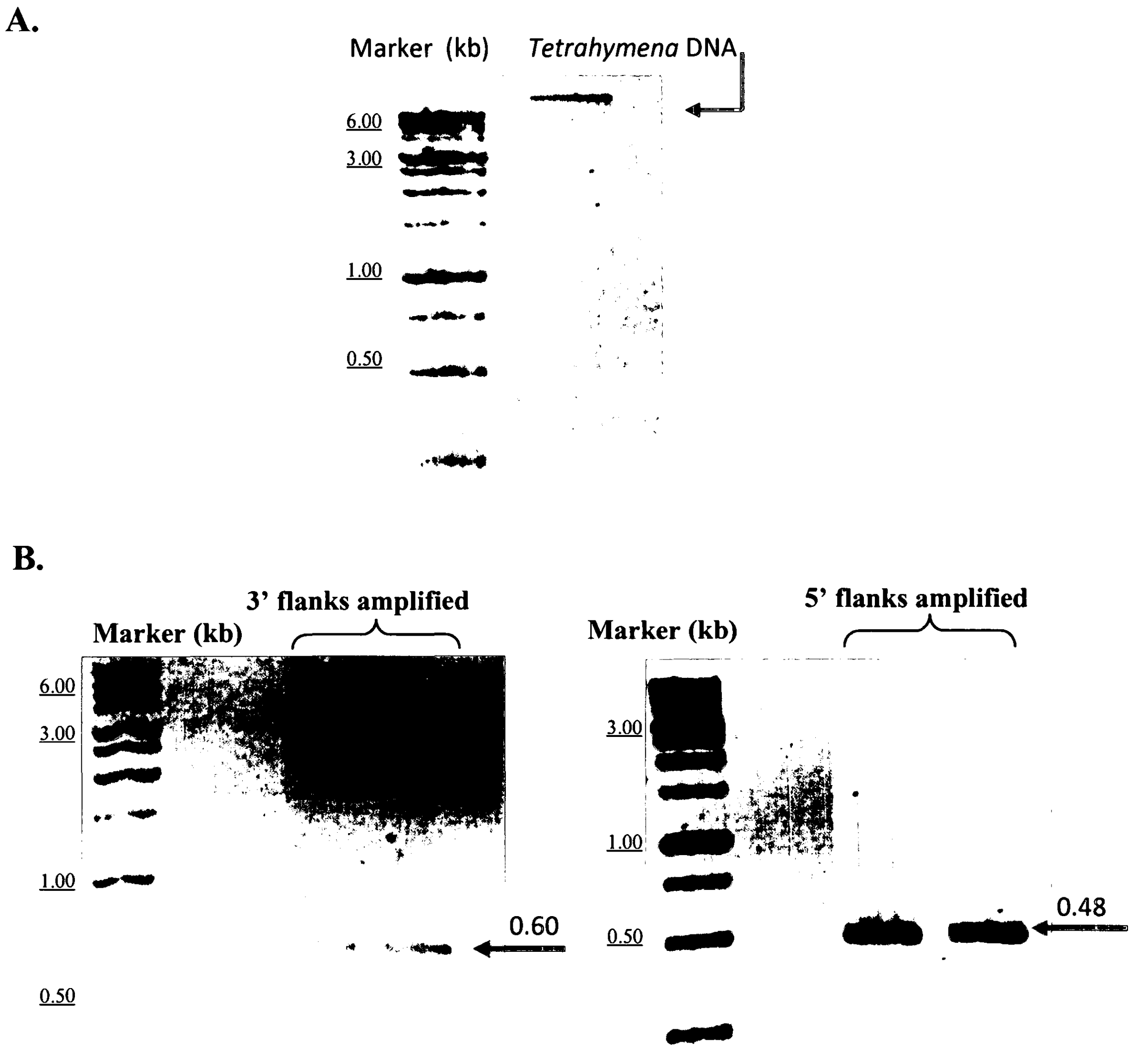
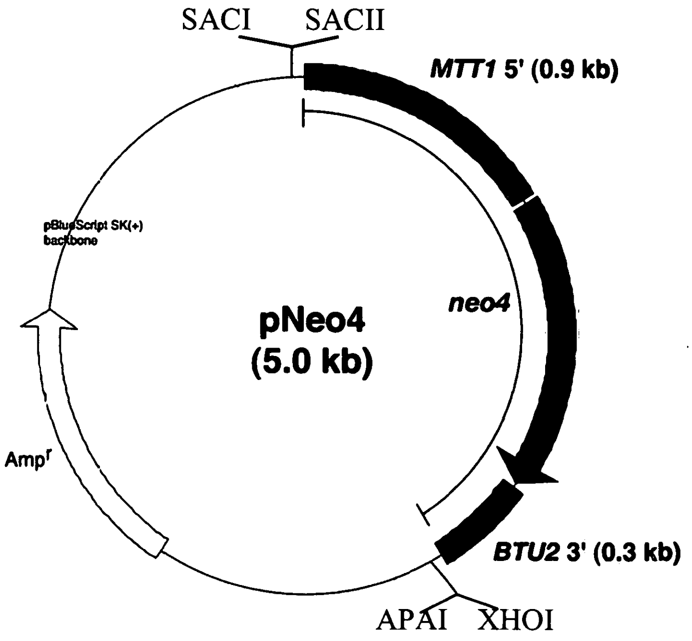


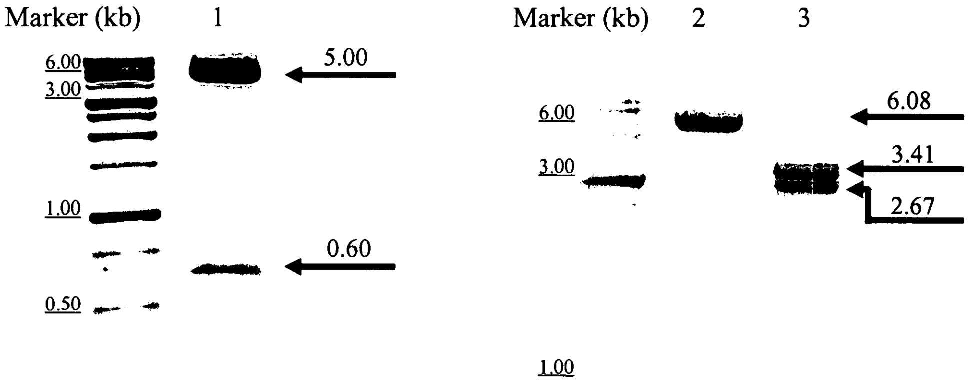
Figure 22. *Tetrahymena* DNA, 5' and 3' flanks

A) *Tetrahymena* DNA extracted and electrophoresed through a 0.8% agarose gel to confirm. Arrow indicates the band that represents *Tetrahymena* DNA. B) PCR was performed to amplify 3' flank (left) and 5' flank (right) using Tetrahymena DNA as a template and the primers from Figure 11. A sample of the product was used for electrophoresis to confirm their presence. The product is shown by the arrows pointing at the specific band (Figure 22b). Sizes of each band are as expected, indicated by the number on top of the arrows.

A.


SACI
SACII
MTT1 5' (0.9 kb)
pBluescript SK(+) backbone
neo4
pNeo4
(5.0 kb)
Amp^r
BTU2 3' (0.3 kb)
APAI
XHOI

B.


Marker (kb)
1
6.00
3.00
1.00
0.50
5.00
0.60
Marker (kb)
2
3
6.00
3.00
1.00
6.08
3.41
2.67

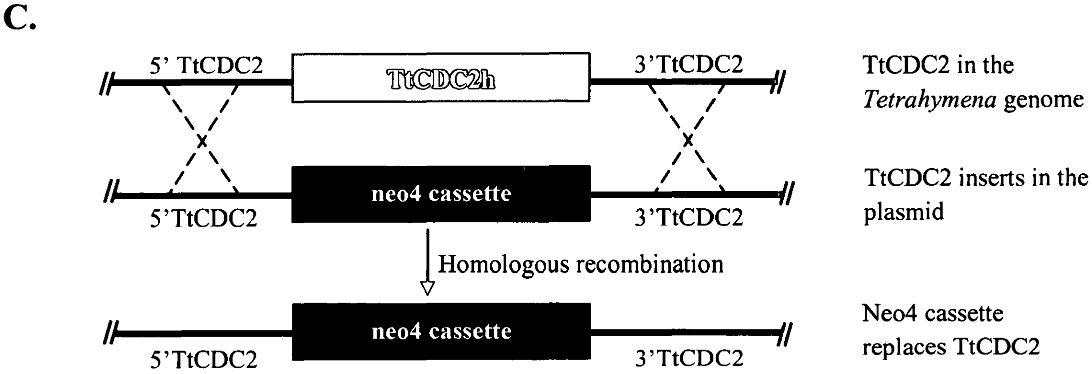
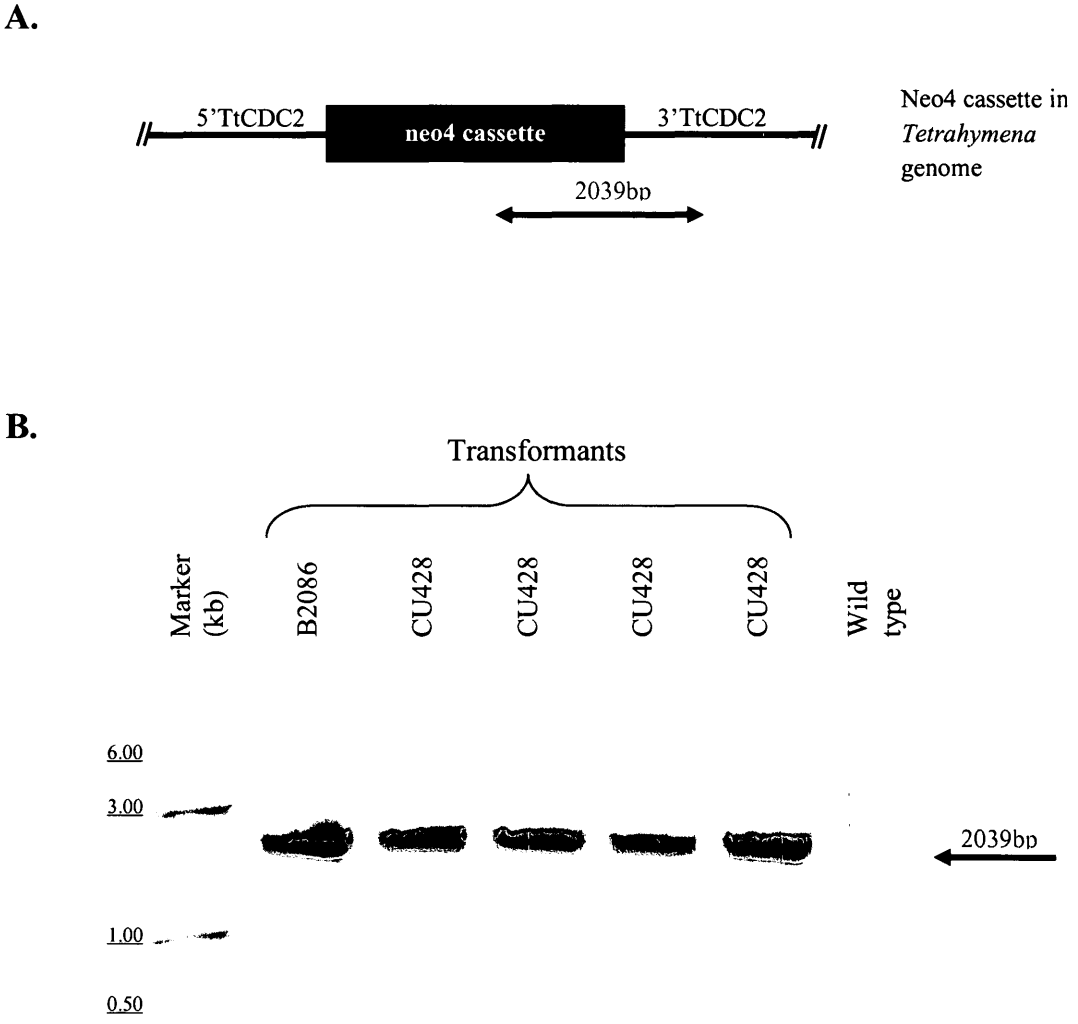
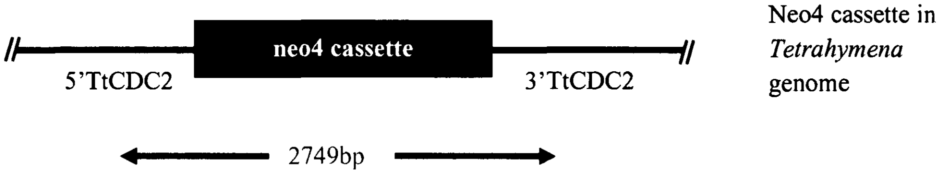


Figure 23. TtCDC2 Knockout Construct

A) Restriction map of pNeo4 plasmid which was used to create TtCDC2 knockout. Enzymes used for digests shown in part B are displayed. B) Lane 1 represents the construct digested with SACI and SACII to release the 3' flank (0.60kb). Lane 2 represents the complete construct (pNeo4 with the 5' and 3' flanks of TtCDC2) at 6.00kb. Lane 3 represents the completed pNeo4 construct digested with the enzymes SACI and APAI to release the neo4 cassette with the inserts (3.41kb). Sizes are as expected (red font) indicating the flanks inserted into the pNeo4 plasmid. This was further confirmed through sequence analysis. C) Successful ligation results in 5' and 3' TtCDC2 flanks surrounding the Neo cassette. *Tetrahymena* cells transformed with this construct created the null mutation for TtCDC2. "//" represents chromosomal DNA.


A.
5'TtCDC2
neo4 cassette
3'TtCDC2
2039bp
Neo4 cassette in *Tetrahymena* genome
B.
Transformants
Marker (kb)
B2086
CU428
CU428
CU428
CU428
Wild type
6.00
3.00
1.00
0.50
2039bp

C.



D.

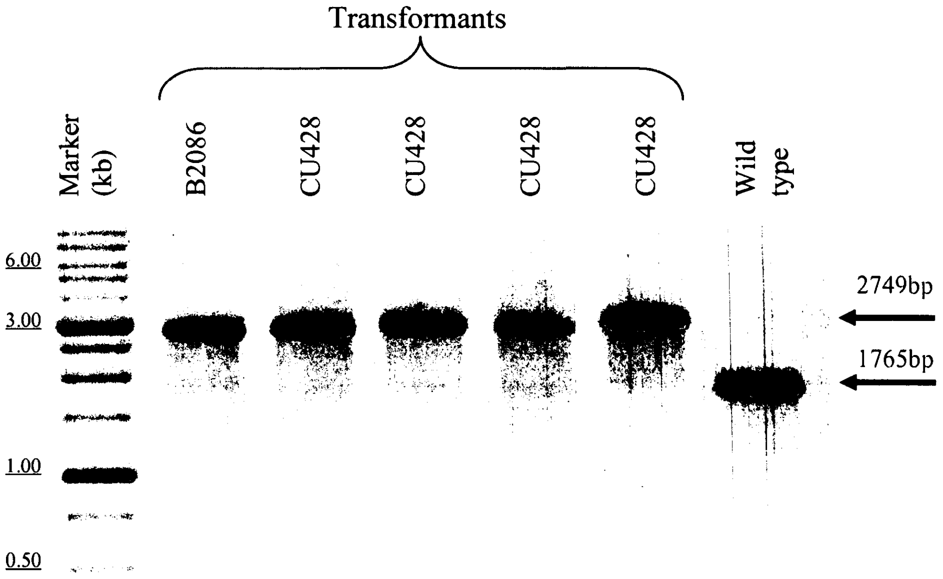


Figure 24. *Tetrahymena* transformants detected

A) A diagram showing that transformed *Tetrahymena* should have the neo4 cassette between the 5' and 3' regions. The red arrow and number indicate the size expected to be detected using specific primer (#4 and 5) and the region it amplifies. Not scaled. B) PCR was done using the primers and the DNA from each transformant as template. PCR was also done using *Tetrahymena* wildtype as a template for the specific primer. Electrophoresis through a 0.8% agarose gel was done running transformed *Tetrahymena* DNA (1 transformed B2086 and 4 different transformed CU428) and a wildtype (non-transformed). The red arrow and number indicates the band as expected. C) Same as (A) but with primers #4 and 6.

4. DISCUSSION

Protein tyrosine phosphorylation is an important process in multicellular eukaryotic organisms and it is well known for its crucial role in regulating signal transduction as stimulated by extracellular stimuli (Pawson et al., 2001). Ciliates could be seen as "swimming receptors" (Andersen et al., 1984) where molecular signal transduction could be a part of starting various physiological activities. In multicellular organisms, modifications of RTKs promote switch-like behaviour and ensure signal induction (Blume-Jensen and Hunter, 2001; Pawson et al., 2001; Pawson, 2002; Schlessinger, 2002). Until recently, tyrosine phosphorylation signaling was thought to be limited to multicellular organisms. Recent data however support tyrosine phosphorylation signaling occurring in unicellular eukaryotes as well (Darnell, 1997; King and Carroll, 2001; King et al., 2003; Shiu and Li, 2004). Some of these organisms are *S. cerevisiae* (PTP and tyrosine kinases), *D. discoideum* (SH2 domains and PTP), and *M. brevicollis* (tyrosine kinases, SH2 domain, and PTP - true tyrosine phosphorylation system) being the closest relative to metazoans (Plowman et al., 1999; Zhu et al., 2000; Pawson et al., 2001; Pincus et al., 2008; Manning et al., 2008; Lim and Pawson, 2010). It is suggested that PTP was first to emerge followed by SH2 and tyrosine kinases (Lim and Pawson, 2010).

Studies in *Tetrahymena* that show physiological activity induced by factors cross reacting between species indicate a molecular cross-recognition between *Tetrahymena* proteins and mammalian growth factors (Andersen et al., 1984; Kovacs and Csaba, 1992; Christopher and Sundermann, 1996; Csaba et al., 2004; Csaba, 2012). This suggests a possible conservation of certain aspects of physiological pathways in terms of evolution.

Exploring tyrosine phosphorylation in unicellular organisms could provide insights into the evolution of tyrosine phosphorylation signaling as well as addressing mechanistic and functional questions about tyrosine phosphorylation signaling. Tyrosine phosphorylation dynamics in *T. thermophila* in mating and the role this phosphorylation could play in this unicellular eukaryotic organism is presented. This research project focused on studying tyrosine phosphorylation signaling in this organism by applying Western blot, indirect immunofluorescence microscopy, bioinformatics, and gene knockout.

4.1 Immunodetection of tyrosine phosphorylation

Previous indirect immunofluorescence microscopy data from the Pearlman lab have shown the presence of tyrosine phosphorylation signaling in starved cells and during conjugation between 10min – 3hrs. This method was tested using different anti-phosphotyrosine antibodies (4G10 –Millipore; PT66 – Sigma Aldrich; PY99 – Santa Cruz Biotech; PY20 - Santa Cruz Biotech; P-Tyr-100 – Cell Signaling Technology). The data showed that there is a high level of tyrosine phosphorylation signaling at 10min after mixing starved cells for pairing, followed by a drop till 3hrs and increase after 3 hrs. The levels of signal detected in conjugating cells were usually above the level observed in starved cells. It is suggested that this signaling pathway plays a role in the activity of conjugation and in the preparation for conjugation (starved state) (Anafi, M., Glowacka, W., Golding, B., and Pearlman, R.E., unpublished observations). Recently a paper on the *Tetrahymena* phosphoproteome has been submitted for publication and provides direct evidence from mass spectrometry analysis for the presence of tyrosine phosphorylated proteins in *Tetrahymena* (W. Miao, personal communication). Also, the *Tetrahymena* genome has revealed kinase orthologs shared among selected eukaryotic genomes (e.g. humans and *S. cerevisiae*) (Eisen et al., 2006; Coyne et al., 2008). Therefore, understanding this signaling pathway in *Tetrahymena* could help us understand the evolution of this pathway between unicells and multicells.

To expand studies to identify tyrosine phosphorylated proteins in *Tetrahymena* and support previous findings in our lab, detection studies were used to determine the presence of tyrosine phosphorylation signaling in *T. thermophila* during conjugation between the two mating types (B2086 and CU428) and at their starved stage.

Western blot optimizations were done to confirm the presence of tyrosine phosphorylated proteins. Western blot analysis was applied to total cell lysate of *T. thermophila* collected at different times in their life cycle (10min, 2hrs and 3hrs (90% pairing) into conjugation and starved). Starvation of *Tetrahymena* cells (18hrs of starvation) serves as an induction for the cells to mate when in the presence of a different mating type. The antibody of choice here was the antiphosphotyrosine antibody, 4G10 Platinum (Millipore) combined with the standard protocol (section 2.4.3). This method was insufficient in providing data indicating the presence of tyrosine phosphorylation signal for conjugating and starved cell lysates. Repeat experiments were done for this and modified protocols were developed to increase signal detected.

Various modifications to the gel system, buffers, lysate preparation, antibody dilutions, and detection were incorporated to optimize signal to noise ratio and improve detection of putatively tyrosine phosphorylated proteins. For example, a 10% SDS-PAGE with a MES buffer system was used and the protocol for lysate preparation was also modified to be compatible with the MES gel buffer system by lysing the cells with urea lysis solution. This system is best for resolving low molecular mass proteins and membrane proteins (Kashino et al., 2007). In multicellular organisms, some membrane proteins become tyrosine phosphorylated and this then initiates specific signaling pathways such as the JAK-STAT pathway (Aaronson and Horvath, 2002). We also believe that if there is tyrosine phosphorylation in *Tetrahymena*, it is possible that it might occur at the membrane or near the membrane. Various bands were detected between 16 kDa to 155kDa (Figure 13d).

Moreover, for a more 'direct' approach at detecting tyrosine phosphorylation signaling, the 4G10 Platinum primary antibody was replaced with 4G10 Platinum-HRP tagged (Millipore) at a 1:500 dilution. This eliminates the requirement to use a secondary antibody therefore removing the background noise caused by this antibody (Figure 13c). This approach minimized the background. The signal detection was further improved by replacing the HyBlot CL film (previously used in our lab) with the Amersham Hyperfilm ECL and increasing the amount of sample loaded (protein) per well up to 80µg/well (Figure 14-15). Different films have different signal detection sensitivity therefore this was investigated between the two mentioned films and increasing the amount of sample loaded increases the amount of protein/peptides to be detected, therefore more signal to observe. These changes did improve signal detection but had little effect on the observed intensity of higher molecular weight bands.

A different approach was taken by modifying the urea lysis solution as outlined in Figure 16. Glycerol is used as a stability reagent for proteins (Gekko and Timasheff, 1981) and is found in the popular lysis solution Laemmli buffer at about 20% final concentration. The same concentration of glycerol was applied to the urea lysis buffer and it showed that compared to the control (just urea lysis buffer), the addition of glycerol to the urea lysis solution further preserved bands. Next, thiourea increases solubility of proteins such as membrane proteins (Rabilloud et al., 1998) so excluding thiourea from the urea lysis solutions led to loss of higher molecular mass bands at 42kDa and above. Urea is also used to solubilise proteins and decreasing urea concentration to 2M (lane 4 compared to lane 2 – Figure 16) led to loss of signal. Including urea and thiourea has also been shown to limit protease action (Rabilloud, 2009). Next, the temperature of the lysate treatment before loading on the gel was increased to 37°C from 25°C. This was done to determine the stability of the sample at a higher temperature since increasing temperatures lead to certain protein modifications. This made no difference in band preservation therefore the sample is stable between 25°C and 37°C. Lastly, previous experiments in our laboratory have shown that using Laemmli buffer alone has led to signal detection of tyrosine phosphorylation being unstable probably due to protease/phosphatase activity. Protease and phosphatase inhibitors were used but did not help in the stability of the signal when just Laemmli buffer was used for lysis. Therefore, 7M urea was combined with Laemmli buffer and the bands were more preserved than any of those in the other lanes and this was chosen as the optimum lysis solution.

4.1.1 *Tetrahymena* pairing and tyrosine phosphorylation signaling

All these findings to detect tyrosine phosphorylation signaling in *Tetrahymena* were applied to determine the level of tyrosine phosphorylation signaling after 18hrs of starvation for single starved cells and during conjugation (10min – 4hrs). Here, 90% of cells were paired after 4hrs of conjugation. After 10min of conjugation, there was an increase in tyrosine phosphorylation signal detected compared to single starved cells. This is followed by a drop in signal after 3hrs of conjugation compared to 10min. At 4hrs the signal increases compared to 3hrs (Figure 17a and 18, Table 3).

As well as Western blot analysis, indirect immunofluorescence microscopy was done to confirm previous results from our lab (Anafi, M., Glowacka, W., Golding, B., and Pearlman, R.E., unpublished observations) and to further support the Western blot observations. The indirect immunofluorescence microscopy observations of cells conjugating at 3hrs (90% pairing of cells) indicated the detection of tyrosine phosphorylation signal by using a 4G10 Platinum antibody (anti-phosphotyrosine antibody - Millipore) combined with a rabbit anti-mouse-FITC (Cedarlane) and controls. The signal was mostly concentrated at the posterior end of one of the mating types (Figure 19). Signal was also observed throughout the cytoplasm.

The previous indirect immunofluorescence results from our lab (Anafi, M., Glowacka, W., Golding, B., and Pearlman R.E., unpublished observations) are further supported by the Western blot data I obtained as stated above (Figure 17a and 18, Table 2). After 90% of cells were paired (after 3hrs of pairing) there is an increase in tyrosine phosphorylation and more concentration of this signal in the posterior region of one of the cells. Again, this is supported by the Western blot and indirect immunofluorescence results which I obtained when 90% of cells were paired (Figure 17a and 18, Table 3). Treatment of *Tetrahymena* mating types with genistein (100μM), a tyrosine specific protein kinase inhibitor, reduced the percentage of cells paired (0%) also reducing phosphorylation signal as compared to cells not treated with genistein (Anafi, M., Glowacka, W., Golding, B., and Pearlman R.E., unpublished observations). Analogous to activation of mammalian receptors, the concentrated region from the indirect immunofluorescence data could serve as a capping site for signaling molecules that collaborate to serve in the early signal transduction steps (Bourguignon and Singer, 1977; Kwiatkowska and Sobota, 1999). This could also suggest that such localization in one of the mating types could be a response to signal from the other one, possibly explaining that when cells are shaken during conjugation it abolishes the conjugation process, as it distributes the ligand (Bruns and Brussard, 1974; Finley and Bruns, 1980; McCoy and Madras, 2004).

Tyrosine phosphorylation signaling in *T. thermophila* in conjugated and starved cells was detected by 4G10 Platinum and 4G10 Platinum HRP (Millipore). This signal was also detected previously using 4G10 and other anti-phosphotyrosine antibodies (Anafi, M., Glowacka, W., Golding, B., and Pearlman, R.E., unpublished observations) Positive (EGF-stimulated A431 cell lysate – Millipore) control was used in this research project to further support the specificity of the 4G10 Platinum and 4G10 Platinum HRP (Millipore). Negative control (Western blot with only secondary antibody) was used to show that the secondary antibody doesn't bind non-specifically to proteins to give a false signal. Anafi, M., Glowacka, W., Golding, B., and Pearlman, R.E., (unpublished observations) also used negative control. Since these same antibodies, for example 4G10 Platinum HRP are also used to detect tyrosine phosphorylation signal in other organisms such as in humans (Burkitt lymphoma cell line (Kotani et al., 2012)), this could indicate the preservation of tyrosine phosphorylation protein domains through evolution between unicells and multicells.

Therefore all of these immunodetection data suggest the presence of tyrosine phosphorylation signaling in *T. thermophila* in conjugated and starved cells, as detected by anti-phosphotyrosine antibodies.

4.2 Bioinformatic observations in *T. thermophila*

Bioinformatic analysis was also applied to address the question of tyrosine phosphorylation signaling in *T. thermophila* by using a database of all predicted *Tetrahymena* protein sequences (http://www.cilate.org) to look for putative *Tetrahymena* orthologues of the *Trypanosoma* tyrosine phosphorylated proteins. The tyrosine phosphorylated proteins from *Trypanosoma* that were used here were those from Table 1 and 2 from Nett et al. (2009). These proteins were detected in *T. brucei* using a phospho-tyrosine-specific proteomics approach (e.g. mass spectrometry).

Tetrahymena's genome has been fully sequenced and annotated. This led to the identification of kinase orthologs that are shared among specific eukaryotic genomes and *Tetrahymena* (Eisen et al., 2006; Coyne et al., 2008). *Tetrahymena*'s fully sequenced and annotated genome in comparison to the *Trypanosoma* tyrosine phosphorylated proteins (Table 1 and 2 from Nett et al. (2009)) were compared through BLAST and revealed similarity to *Trypanosoma* proteins such as protein kinase domains, MPK's, and CDK. Alignment between the peptide sequences and the corresponding *Tetrahymena* protein amino acid sequences revealed homology for most of the comparisons between these sequences, more specifically between the tyrosine sites of the two sequences. Most of these *Tetrahymena* tyrosine sites that were homologous to their corresponding *Trypanosoma* peptides were identified by NetPhos as having a significant probability (scores of higher than 0.500) of them being tyrosine phosphorylation sites. Therefore these data indicate the possibility of these genes to have putative tyrosine phosphorylation sites. Since this research project focuses on the detection of tyrosine phosphorylation during conjugation and starvation because it might play a role in the mating of *T. thermophila*, the expression profiles (http://tfgd.ihb.ac.cn/ and Appendix B-5) of each *Tetrahymena* gene (Appendix B-1) were analyzed. All of the gene expression profiles analysed here showed expression of that gene during conjugation. Most of these genes had higher expression peaks during conjugation than starvation, and from the ones that did, most of them include peaks of increasing gene expression between 0-4hrs. Although these profiles just tell us when specific genes are expressed during the lifetime of *Tetrahymena*, but correlation with the NetPhos and BLAST data suggests that these genes could possibly be tyrosine phosphorylated.

Similarly, this was seen with the Western blot data where some of the bands had higher signal detection (higher relative density and average relative density) during conjugation (10min-4hrs) than starvation (Figure 18 and Table 2). Likewise, the presence of tyrosine phosphorylation signaling during conjugation and starvation is further supported by the indirect immunofluorescence microscopy from the Pearlman lab (I; Anafi, M., Glowacka, W., Golding, B., and Pearlman R.E., unpublished observations). The indirect immunofluorescence microscopy data indicated higher levels of signal detected during conjugation than starvation. Other bioinformatic analysis of the *Tetrahymena* genome from the Pearlman lab indicate the identification of 12 dual specificity kinases designated as *Tetrahymena* Serine/Threonine/Tyrosine Kinases (tSTYK). From these kinases, 10 of them have a higher probability of being tyrosine instead of serine/threonine kinase (Anafi, M., Glowacka, W., Golding, B., and Pearlman R.E., unpublished observations). Among this bioinformatic analysis, three tyrosine specific phosphatases designated as *Tetrahymena* Protein Tyrosine Phosphatase (tPTP) were identified. All these bioinformatic analyses from the Pearlman lab combined with the Western blot and indirect immunofluorescence data indicate that there likely is tyrosine phosphorylation signaling in *Tetrahymena,* possibly involved in the regulation of its lifecycle such as in conjugation.

Also as mentioned earlier, a paper on the *Tetrahymena* phopshoproteome has been submitted for publication and it provides direct evidence for tyrosine phosphorylated proteins (W. Miao, personal communication). From the proteins identified through *Tetrahymena* phosphoproteome analysis (W. Miao, personal communication), one of those proteins was TTHERM_00267860 which is a protein indentified through the Blast analysis described in this thesis. The same tyrosine site that was identified as being a tyrosine phosphorylation site for TTHERM_00267860, I identified as being tyrosine phosphorylated through homology comparison to *Trypanosoma* and NetPhos analysis.

The molecular weight of this protein also correlates to the Western blot bands detected at 74.1kDa (Figure 17a). In addition, the molecular weight of three other proteins that were identified as being tyrosine phosphorylated (W. Miao, personal communication) also corresponded to the molecular weight of the Western blot bands (Figure 17a) detected at 33.0, 68.9, and 74.1kDa. These proteins identified through Western blot analysis are targets for further analysis to confirm identity with proteins identified in the phosphoproteome analysis Miao. These data show that the bioinformatic method used demonstrates some efficiency in predicting putative tyrosine phosphorylated proteins and it is possible that there are more tyrosine phosphorylated proteins that weren't captured through phosphoproteome analysis.

Therefore the bioinformatic data also contribute to the above interpretation and suggest that there are putative tyrosine phosphorylated proteins in *T. thermophila* that may play a role during conjugation and starvation of the cells.

4.3 *Tetrahymena* CDC2

Among the *Tetrahymena* proteins identified by similarity with the *Trypanosoma* phospho peptides, TtCdc2 (a protein kinase from the cell division cycle family) was one of them. Based on the bioinformatic analysis for this protein (and its corresponding gene, TtCDC2) an interest was taken in this protein to show a more specific approach in studying tyrosine phosphorylation signaling in *Tetrahymena*. This was done by making knockout constructs in which the coding region of the gene was replaced with the Neo4 cassette from the pNeo4 plasmid by homologous recombination. Southern blot analysis in the future will follow to further confirm the construct. Investigation of knockout TtCDC2

Tetrahymena strains in the future would allow for investigation of the role of this gene in this organism and determine its relevance to tyrosine phosphorylation signaling in the cells' life cycle. For example, this could be done by repeating the Western blot and indirect immunofluorescence method with total cell lysate from these strains and assay for tyrosine phosphorylation signaling. The protein lysate for this protein collected after starvation and at specific times during conjugation could also be processed using these immunodetection methods to detect tyrosine phosphorylation signaling and compared to the wild type to determine differences in the level of tyrosine phosphorylation.

Bioinformatic analysis of TtCdc2 (Figure 20-21) predicted this protein as containing multiple tyrosine phosphorylation sites including one site that is homologous to the *Trypanosoma* peptide of CRK1 (Tb10.70.7040). The microarray gene expression profile for its gene, TtCDC2 indicates expression only during conjugation in *Tetrahymena,* with its highest expression peak at 2hrs.

The protein product of TtCDC2 is a protein kinase and a homolog to cdk's catalytic subunit (cell cycle-dependent kinase) (O'Connell et al., 2003; Dou et al., 2005). Cdc2 is found in eukaryotic cells where it controls the entry of cells into mitosis and it is a universal mitotic trigger (Lee and Nurse, 1987). In the fission yeast *Schizosaccharomyces pombe* CDC2 was shown to be essential for both premeiotic DNA replication and meiosis II (Iino et al., 1995). In fungi, Cdc2 is activated by dephosphorylation at tyrosine 15 leading to the regulation of NRK kinases (highly conserved in evolution) (O'Connell et al., 2003).

Macronuclear genome sequencing in *T. thermophila* led to the identification of 39

NRKs (Eisen et al., 2006). A NRK/NIMA kinase is a putative tyrosine kinase that is induced by mammalian insulin as determined in *Tetrahymena* by Christensen et al., (2003). Studies in *Tetrahymena* have shown that Cdc2 phosphorylates histone H1 and strongly resembles Cdc2 kinase from yeast and mammals (Dou et al., 2005). So when there is higher level of H1 phosphorylation there is also increased expression of CDC2 (Song and Gorovsky, 2007). Also, studies in *Tetrahymena* have shown that H1 is hyperphosphorylated during early stages of conjugation, dephosphorylated after prolonged starvation and at later stages of conjugation (during macronuclear elimination) (Allis and Gorovsky, 1981; Dou et al., 2005). Since we see increasing expression of tyrosine phosphorylation signaling during conjugation and TtCdc2 expression peaks during conjugation it might be possible that this protein is among those contributing to this signal in *Tetrahymena.* More remains to be understood about the function of this protein and its gene, so future studies using the TtCDC2 knockout *Tetrahymema* strains could provide further insights into this and potentially its relevance to tyrosine phosphorylation signaling during specific stages of the life cycle of this organism.

4.4 Conclusion

In this investigation, *T. thermophila* has a complex and dynamic tyrosine phosphorylation signaling in a manner correlated to cell cycle regulation during starvation and conjugation. Possible molecules involved in tyrosine phosphorylation signaling are presently encoded in the genome of *Tetrahymena.* However, more remains to be understood about the role of these signaling molecules. This investigation extends on the importance of protein tyrosine phosphorylation in unicellular eukaryotes and shows evidence of possible tyrosine phosphorylation signaling in *Tetrahymena* such as during starvation and conjugation between different mating types. The Western blot and indirect immunofluorescence data indicate the presence of tyrosine phosphorylated proteins. These proteins are present throughout the cytoplasm but more concentrated in the posterior region of the cell during mating. This is further supported from the bioinformatic data. Analysing the function of *Tetrahymena* genes such as by knocking out the TtCDC2 gene to produce a null mutant or knockdown strains could provide a more detailed understanding of its role including tyrosine phosphorylation in the organism. Future work in *Tetrahymena* could include producing more knockout strains for the genes from the BLAST data, tSTYK and tPTP to investigate their function and relevance to tyrosine phosphorylation.

Furthermore, investigating protein tyrosine phosphorylation in *Tetrahymena* could provide insights into understanding the evolution of tyrosine kinase signaling between unicellular and multicellular organisms.

5. REFERENCES

Aaronson, D.S., and C.M. Horvath. 2002. A road map for those who don't know JAK-STAT. *Science.* 296:1653-1655.

Akematsu, T., and H. Endoh. 2010. Role of apoptosis-inducing factor (AIF) in programmed nuclear death during conjugation in Tetrahymena thermophila. *BMC Cell Biol.* 11:13.

Allis, C.D., and M.A. Gorovsky. 1981. Histone phosphorylation in macro- and micronuclei of Tetrahymena thermophila. *Biochemistry.* 20:3828-3833.

Andersen, H.A., H. Flodgaard, H. Klenow, and V. Leick. 1984. Platelet-derived growth factor stimulates chemotaxis and nucleic acid synthesis in the protozoan Tetrahymena. *Biochim.Biophys.Acta.* 782:437-440.

Blume-Jensen, P., and T. Hunter. 2001. Oncogenic kinase signalling. *Nature.* 411:355-365.

Bourguignon, L.Y., and S.J. Singer. 1977. Transmembrane interactions and the mechanism of capping of surface receptors by their specific ligands. *Proc.Natl.Acad.Sci.U.S.A.* 74:5031-5035.

Bruns, P.J., and T.B. Brussard. 1974. Pair formation in tetrahymena pyriformis, an inducible developmental system. *J.Exp.Zool.* 188:337-344.

Bruns, P.J., and D. Cassidy-Hanley. 2000. Biolistic transformation of macro- and micronuclei. *Methods Cell Biol.* 62:501-512.

Christensen, S.T., C.F. Guerra, A. Awan, D.N. Wheatley, and P. Satir. 2003. Insulin receptor-like proteins in Tetrahymena thermophila ciliary membranes. *Curr.Biol.* 13:R50-2.

Christopher, G.K., and C.A. Sundermann. 1996. Intracellular insulin binding in Tetrahymena pyriformis. *Tissue Cell.* 28:427-437.

Cole, E.S., D. Cassidy-Hanley, J. Hemish, J. Tuan, and P.J. Bruns. 1997. A mutational analysis of conjugation in Tetrahymena thermophila. 1. Phenotypes affecting early development: meiosis to nuclear selection. *Dev.Biol.* 189:215-232.

Coyne, R.S., N.A. Stover, and W. Miao. 2012. Whole genome studies of Tetrahymena. *Methods Cell Biol.* 109:53-81.

Coyne, R.S., M. Thiagarajan, K.M. Jones, J.R. Wortman, L.J. Tallon, B.J. Haas, D.M. Cassidy-Hanley, E.A. Wiley, J.J. Smith, K. Collins, S.R. Lee, M.T. Couvillion, Y. Liu, J. Garg, R.E. Pearlman, E.P. Hamilton, E. Orias, J.A. Eisen, and B.A. Methe. 2008. Refined annotation and assembly of the Tetrahymena thermophila genome sequence through EST analysis, comparative genomic hybridization, and targeted gap closure. *BMC Genomics.* 9:562.

Csaba, G. 2012. Hormonal imprinting in the unicellular Tetrahymena: the proto-model of epigenetics. *Acta Microbiol.Immunol.Hung.* 59:291-310.

Csaba, G., P. Kovacs, and E. Pallinger. 2004. Presence and localization of epidermal growth factor (EGF)- and EGF-receptor-like immunoreactivity in Tetrahymena. *Cell Biol.Int.* 28:491-496.

Darnell, J.E. 1997. Phosphotyrosine signaling and the single cell: metazoan boundary. *Proceedings of the National Academy of Sciences.* 94:11767.

Dou, Y., X. Song, Y. Liu, and M.A. Gorovsky. 2005. The H1 phosphorylation state regulates expression of CDC2 and other genes in response to starvation in Tetrahymena thermophila. *Mol.Cell.Biol.* 25:3914-3922.

Eisen, J.A., R.S. Coyne, M. Wu, D. Wu, M. Thiagarajan, J.R. Wortman, J.H. Badger, Q. Ren, P. Amedeo, K.M. Jones, L.J. Tallon, A.L. Delcher, S.L. Salzberg, J.C. Silva, B.J. Haas, W.H. Majoros, M. Farzad, J.M. Carlton, R.K. Smith Jr, J. Garg, R.E. Pearlman, K.M. Karrer, L. Sun, G. Manning, N.C. Elde, A.P. Turkewitz, D.J. Asai, D.E. Wilkes, Y. Wang, H. Cai, K. Collins, B.A. Stewart, S.R. Lee, K. Wilamowska, Z. Weinberg, W.L. Ruzzo, D. Wloga, J. Gaertig, J. Frankel, C.C. Tsao, M.A. Gorovsky, P.J. Keeling, R.F. Waller, N.J. Patron, J.M. Cherry, N.A. Stover, C.J. Krieger, C. del Toro, H.F. Ryder, S.C. Williamson, R.A. Barbeau, E.P. Hamilton, and E. Orias. 2006. Macronuclear genome sequence of the ciliate Tetrahymena thermophila, a model eukaryote. *PLoS Biol.* 4:e286.

Finley, M.J., and P.J. Bruns. 1980. Costimulation in Tetrahymena. II. A nonspecific response to heterotypic cell--cell interactions. *Dev.Biol.* 79:81-94.

Gekko, K., and S.N. Timasheff. 1981. Mechanism of protein stabilization by glycerol: preferential hydration in glycerol-water mixtures. *Biochemistry.* 20:4667-4676.

Greider, C.W., and E.H. Blackburn. 1985. Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. *Cell.* 43:405-413.

Gu, J., and X. Gu. 2003. Natural history and functional divergence of protein tyrosine kinases. *Gene.* 317:49-57.

Hedges, S.B., J.E. Blair, M.L. Venturi, and J.L. Shoe. 2004. A molecular timescale of eukaryote evolution and the rise of complex multicellular life. *BMC Evol.Biol.* 4:2.

Hubbard, S.R. 1997. Crystal structure of the activated insulin receptor tyrosine kinase in complex with peptide substrate and ATP analog. *EMBO J.* 16:5572-5581.

Hubbard, S.R., and J.H. Till. 2000. Protein tyrosine kinase structure and function. *Annu.Rev.Biochem.* 69:373-398.

Hunter, T., and G.D. Plowman. 1997. The protein kinases of budding yeast: six score and more. *Trends Biochem.Sci.* 22:18-22.

Iino, Y., Y. Hiramine, and M. Yamamoto. 1995. The role of cdc2 and other genes in meiosis in Schizosaccharomyces pombe. *Genetics.* 140:1235-1245.

Karrer, K.M. 2000. Tetrahymena genetics: two nuclei are better than one. *Methods Cell Biol.* 62:127-186.

Kashino, Y., T. Harayama, H.B. Pakrasi, and K. Satoh. 2007. Preparation of membrane proteins for analysis by two-dimensional gel electrophoresis. *J.Chromatogr.B.Analyt Technol.Biomed.Life.Sci.* 849:282-292.

Kawata, T. 2011. STAT signaling in Dictyostelium development. *Dev.Growth Differ.* 53:548-557.

King, N., and S.B. Carroll. 2001. A receptor tyrosine kinase from choanoflagellates: molecular insights into early animal evolution. *Proc.Natl.Acad.Sci.U.S.A.* 98:15032-15037.

King, N., C.T. Hittinger, and S.B. Carroll. 2003. Evolution of key cell signaling and adhesion protein families predates animal origins. *Science.* 301:361-363.

Kotani, N., Y. Ishiura, R. Yamashita, T. Ohnishi, and K. Honke. 2012. Fibroblast Growth Factor Receptor 3 (FGFR3) Associated with the CD20 Antigen Regulates the Rituximab-induced Proliferation Inhibition in B-cell Lymphoma Cells. *J.Biol.Chem.* 287:37109-37118.

Kovacs, P., and G. Csaba. 1992. Effect of inhibitors and activators of tyrosine kinase on insulin imprinting in Tetrahymena. *Cell Biochem.Funct.* 10:267-271.

Kruger, K., P.J. Grabowski, A.J. Zaug, J. Sands, D.E. Gottschling, and T.R. Cech. 1982. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell.* 31:147-157.

Kwiatkowska, K., and A. Sobota. 1999. Tyrosine phosphorylation/dephosphorylation controls capping of Fcgamma receptor II in U937 cells. *Cell Motil.Cytoskeleton.* 42:298-314.

Lee, M.G., and P. Nurse. 1987. Complementation used to clone a human homologue of the fission yeast cell cycle control gene cdc2. *Nature.* 327:31-35.

Li, S., N.D. Covino, E.G. Stein, J.H. Till, and S.R. Hubbard. 2003. Structural and biochemical evidence for an autoinhibitory role for tyrosine 984 in the juxtamembrane region of the insulin receptor. *J.Biol.Chem.* 278:26007-26014.

Li, S., L. Yin, E.S. Cole, R.A. Udani, and K.M. Karrer. 2006. Progeny of germ line knockouts of ASI2, a gene encoding a putative signal transduction receptor in Tetrahymena thermophila, fail to make the transition from sexual reproduction to vegetative growth. *Dev.Biol.* 295:633-646.

Lim, W.A., and T. Pawson. 2010. Phosphotyrosine signaling: evolving a new cellular communication system. *Cell.* 142:661-667.

Liu, B.A., E. Shah, K. Jablonowski, A. Stergachis, B. Engelmann, and P.D. Nash. 2011a. The SH2 domain-containing proteins in 21 species establish the provenance and scope of phosphotyrosine signaling in eukaryotes. *Sci.Signal.* 4:ra83.

Liu, J., J. Zhang, Q. Gong, P. Xiong, H. Huang, B. Wu, G. Lu, J. Wu, and Y. Shi. 2011b. Solution structure of tandem SH2 domains from Spt6 protein and their binding to the phosphorylated RNA polymerase II C-terminal domain. *J.Biol.Chem.* 286:29218-29226.

Manning, G., S.L. Young, W.T. Miller, and Y. Zhai. 2008. The protist, Monosiga brevicollis, has a tyrosine kinase signaling network more elaborate and diverse than found in any known metazoan. *Proceedings of the National Academy of Sciences.* 105:9674.

McCoy, B.J., and G. Madras. 2004. Cluster kinetics of density relaxation in granular materials. *Phys.Rev.E.Stat.Nonlin Soft Matter Phys.* 70:051311.

Nett, I.R., L. Davidson, D. Lamont, and M.A. Ferguson. 2009. Identification and specific localization of tyrosine-phosphorylated proteins in Trypanosoma brucei. *Eukaryot.Cell.* 8:617-626.

O'Connell, M.J., M.J.E. Krien, and T. Hunter. 2003. Never say never. The NIMA-related protein kinases in mitotic control. *Trends Cell Biol.* 13:221-228.

Ohba, R., D.J. Steger, J.E. Brownell, C.A. Mizzen, R.G. Cook, J. Cote, J.L. Workman, and C.D. Allis. 1999. A novel H2A/H4 nucleosomal histone acetyltransferase in Tetrahymena thermophila. *Mol.Cell.Biol.* 19:2061-2068.

Orias, E. 2000. Toward sequencing the Tetrahymena genome: exploiting the gift of nuclear dimorphism. *J.Eukaryot.Microbiol.* 47:328-333.

Orias, J.D., E.P. Hamilton, and E. Orias. 1983. A microtubule meshwork associated with gametic pronucleus transfer across a cell-cell junction. *Science.* 222:181-184.

O'Shea, J.J., M. Pesu, D.C. Borie, and P.S. Changelian. 2004. A new modality for immunosuppression: targeting the JAK/STAT pathway. *Nat.Rev.Drug Discov.* 3:555-564.

Pawson, T. 2004. Specificity in signal transduction: from phosphotyrosine-SH2 domain interactions to complex cellular systems. *Cell.* 116:191-203.

Pawson, T. 2002. Regulation and targets of receptor tyrosine kinases. *Eur.J.Cancer.* 38 Suppl 5:S3-10.

Pawson, T., G.D. Gish, and P. Nash. 2001. SH2 domains, interaction modules and cellular wiring. *Trends Cell Biol.* 11:504-511.

Pincus, D., I. Letunic, P. Bork, and W.A. Lim. 2008. Evolution of the phospho-tyrosine signaling machinery in premetazoan lineages. *Proc.Natl.Acad.Sci.U.S.A.* 105:9680-9684.

Plowman, G.D., S. Sudarsanam, J. Bingham, D. Whyte, and T. Hunter. 1999. The protein kinases of Caenorhabditis elegans: a model for signal transduction in multicellular organisms. *Proc.Natl.Acad.Sci.U.S.A.* 96:13603-13610.

Rabilloud, T. 2009. Membrane proteins and proteomics: love is possible, but so difficult. *Electrophoresis.* 30 Suppl 1:S174-80.

Rabilloud, T., S. Kieffer, V. Procaccio, M. Louwagie, P.L. Courchesne, S.D. Patterson, P. Martinez, J. Garin, and J. Lunardi. 1998. Two-dimensional electrophoresis of human placental mitochondria and protein identification by mass spectrometry: toward a human mitochondrial proteome. *Electrophoresis.* 19:1006-1014.

Schemarova, I.V. 2006. The role of tyrosine phosphorylation in regulation of signal transduction pathways in unicellular eukaryotes. *Curr.Issues Mol.Biol.* 8:27-49.

Schlessinger, J. 2002. Ligand-induced, receptor-mediated dimerization and activation of EGF receptor. *Cell.* 110:669-672.

Shiu, S.H., and W.H. Li. 2004. Origins, lineage-specific expansions, and multiple losses of tyrosine kinases in eukaryotes. *Mol.Biol.Evol.* 21:828-840.

Song, X., and M.A. Gorovsky. 2007. Unphosphorylated H1 is enriched in a specific region of the promoter when CDC2 is down-regulated during starvation. *Mol.Cell.Biol.* 27:1925-1933.

Wloga, D., A. Camba, K. Rogowski, G. Manning, M. Jerka-Dziadosz, and J. Gaertig. 2006. Members of the NIMA-related kinase family promote disassembly of cilia by multiple mechanisms. *Mol.Biol.Cell.* 17:2799-2810.

Yao, M.C., and J.L. Chao. 2005. RNA-guided DNA deletion in Tetrahymena: an RNAi-based mechanism for programmed genome rearrangements. *Annu.Rev.Genet.* 39:537-559.

Zhu, H., J.F. Klemic, S. Chang, P. Bertone, A. Casamayor, K.G. Klemic, D. Smith, M. Gerstein, M.A. Reed, and M. Snyder. 2000. Analysis of yeast protein kinases using protein chips. *Nat.Genet.* 26:283-289.

APPENDIX A: Reagents and media

Solution Name	Ingredients of Solution
Agarose gel (0.8%)	0.4 g agarose 50 ml 1XTBE 5 µl ethidium bromide
Buffer for running	1.2 g MES 3.6 g Tris 500 µl of 10% SDS Up to 10 ml with ddH_2O
Buffer for stacking	6.25 ml of 1 M Tris-HCl pH 6.8 1 ml of 10% SDS Up to 10 ml with ddH_2O
Laemmli buffer (2x)	1.36 ml of 1 M Tris-HCl pH 6.8 4.4 ml of 10% SDS 2.2 ml of 100% glycerol 2.04 ml ddH_2O
Urea lysis solution	12 g urea 1.75 ml of 5 M NaCl 0.25 ml of 1 M Tris HCl pH 7.4 0.25 ml of 1 M EDTA 2.5 ml of 10% SDS Up to 25 ml with ddH_2O
Glycerol buffer	90 ml of 100% glycerol 10 ml of 1XPBS
PBS (10X)	2.64 g NaH_2PO_4 16 g Na_2HPO_4 82 g NaCl pH 7.3 Up to 1.0 L with ddH_2O
SDS gel stain	1.25 g Coomassie Blue R250 225 ml methanol 50 ml glacial acetic acid 225 ml ddH_2O
SDS gel destain	100 ml glacial acetic acid 300 ml ethanol 600 ml ddH_2O
SDS Polyacrylamide gel (standard) - *10% running gel*	9.24 ml of 1 M Tris pH 9.0 8.31 ml of 30% acrylamide 25 µl of 10% SDS 6.25 µl TEMED 187 µl of 10% APS 7.0 ml of ddH_2O

SDS Polyacrylamide gel (standard) - *5% stacking gel*	625µl of 1 M Tris pH 6.8 950µl of 30% acrylamide 50 µl of 10% SDS 3.75 µl TEMED 31.25 µl of 10% APS 3.5 ml of ddH_2O
SDS Polyacrylamide gel (MES system) – *10% running gel*	6 ml 30% acrylimide 4 ml buffer for running 7.2 g urea powder 4.8 ml water 20 µl 10% APS 9 µl TEMED
SDS Polyacrylamide gel (MES system) – *5% stacking gel*	2 ml 30% acrylamide 2 ml buffer for stacking 3.6 g urea powder 3.5 ml ddH_2O 40 µl of 10% APS 4.5 µl TEMED
Schaudinn's reagent	74 mg $HgCl_2$ 0.5 ml of 95% EtOH to 1.5 ml with ddH_2O
SPP medium (10X)	0.3 g ferric EDTA 10 g yeast extract 20 g glucose 100 g proteose peptone Up to 1 L with ddH_2O
TBS	8.76 g NaCl 6.05 g Tris 1 M HCl to adjust to pH 7.5 Up to 1 L with ddH_2O
TBE (10x)	108 g tris base 55 g boric acid 20 ml of 0.5 M EDTA Up to 1 L with ddH_2O
YT	10 g bacto-tryptone 5 g yeast extract 5 g NaCl 1 L ddH_2O

YT agar	10 g bacto-tryptone 5 g yeast extract 5 g NaCl 15 g agar 1 L ddH_2O
YT + Amp	400 ml YT 2 ml of 10 mg/ml ampicillin

APPENDIX B: Bioinformatic data

The following are the bioinformatic results. Appendix B-1 shows the BLAST results of comparing the *T. brucei* proteins (amino acid sequence) from Table 1 from Nett et. al., (2009) to a database of *Tetrahymena* protein sequences (http://www.cilate.org). Appendix B-2 shows the amino acid sequence of each *Tetrahymena* protein that was the top match for each BLAST comparison. Appendix B-3 shows the *Tetrahymena* protein sequences (Appendix B-2) aligned with the corresponding *T. brucei* peptide sequences from Appendix B-1. The yellow highlight shows a match between the tyrosine site from the *T. brucei* peptide sequences to the tyrosine site on the corresponding *Tetrahymena* protein sequence. The red oval shape over the tyrosine site indicates that specific *Tetrahymena* site is likely to become tyrosine phosphorylated as determined by NetPhos 2.0 (score of 0.500 or higher). Since the *T. brucei* peptide was small compared to the *Tetrahymena* protein sequence, the rest of the *Tetrahymena* amino acid sequence that didn't fall within the alignment between the two organisms was removed as the full sequence is found in Appendix B-2. Appendix B-4-1 shows the NetPhos results for each *Tetrahymena* amino acid sequence from Appendix B-2 (first table). The *Y* indicates the tyrosine in that part of the amino acid sequence is likely to become tyrosine phosphorylated. Appendix B-4-2 sames as Appendix B-4-1 except it corresponds to the second table (Appendix B-2). Appendix B-5 shows the microarray gene expression profile for each *Tetrahymena* gene that corresponds to the *Tetrahymena* proteins from Appendix B-2. These profiles were obtained from http://tfgd.ihb.ac.cn/. The blue line is data obtained from Gorovsky and Miao labs and the red line indicates in addition to the blue line, 10 conjugation samples completed from the Pearlman lab. The genes that include expression peaks during conjugation between 0-4hrs are TTHERM_00161270, TTHERM_00185770, TTHERM_00190830, TTHERM_00219410, TTHERM_00267860, TTHERM_00316060, TTHERM_00339690, TTHERM_00348940, TTHERM_00355780, TTHERM_00469230, TTHERM_00483640, TTHERM_00529780, TTHERM_00537240, TTHERM_00780700, TTHERM_00827160, TTHERM_01035490, TTHERM_01043090 and TTHERM_01347900.

Appendix B-1. BLAST results

From Table 1 of Nett et al. (2009)				
***T. brucei* GeneDB accession no.**	***T. brucei* GeneDB annotation**	***T.brucei* phosphopeptide**	***Tetrahymena* Gene accession no. (BLAST match)**	***Tetrahymena* Gene annotation or (Description)**
Tb10.70.7040	CRK1	IGEGS**Y**GVVFR	TTHERM_00483640	(CDC2 PK)
Tb10.389.1730	MAPK11	EESDQGEHMTD**Y**VTMR	TTHERM_00469230	MPK1
Tb10.6k15.2790	MAPK1	CFNTQGGDNDLTE**Y**IATR	TTHERM_00575520	MPK4
Tb927.7.2420	GSK3α	NVP**Y**IFSR	TTHERM_00532680	(GSK PK)
Tb927.3.690	PK	ARPPYTD**Y**VSTR	TTHERM_00267860	(RCK PK)
Tb927.7.3580	PK	FNGG**Y**NSSSVSGNGVVK	TTHERM_00529780	NRK19
Tb10.70.2070	MAPK2	EQVARPVLTD**Y**IMTR	TTHERM_00537240	(ERK7 PK)
Tb11.01.4230	PK	KVT**Y**ALPNQSR	TTHERM_00219410	(CLK PK)
Tb10.329.0030	PK	DAQASDTF**Y**VCTR	TTHERM_00557780	(GSK PK)
Tb927.8.3770	PK	EDTQDPNKTH**Y**VTHR	TTHERM_00195990	(ERK PK)
Tb09.211.0960	PK	GLHVSQPLTE**Y**VSTR	TTHERM_00058800	(RCK PK)
Tb10.70.2210	CRK3	MDILGEGT**Y**GVVYR	TTHERM_01035490	(CDC2 PK)
Tb11.02.0640	PK	LFT**Y**IQSR	TTHERM_00348940	(DYRK2 PK)
Tb11.01.8550	ECK1	GNYTE**Y**VATR	TTHERM_00185770	(CDKL PK)
Tb10.61.1850	MAPK9	SRPPFTE**Y**VSTR	TTHERM_01347900	(RCK PK)
Tb10.61.0250	MAPK2	DDQCTQTSALTE**Y**VVTR	TTHERM_00469230	MPK1
Tb927.6.1780	PK	GVGVNVTS**Y**VVTR	TTHERM_00195990	(ERK PK)
Tb927.7.7360	CRK2	VGEGS**Y**GIVYK	TTHERM_01035490	(CDC2 PK)
Tb10.61.3140	PK	LAADEPNVA**Y**ICSR	TTHERM_00532680	(GSK PK)

From Table 2 of Nett et al. (2009)				
***T. brucei* GeneDB accession no.**	***T. brucei* Protein Product Description**	***T.brucei* phosphopeptide**	***Tetrahymena* Gene accession no. (BLAST match)**	***Tetrahymena* Gene annotation or (Description)**
Tb11.01.1475	Ribosomal Protein S27	LVQGPNS**Y**FMDVK	TTHERM_01043090	(Hypothetical Protein)
Tb927.2.4210	Glycosomal Phosphoenolpyruvate carboxykinase	EQVILGTE**Y**AGEMK	TTHERM_00420030	(Phosphoenolpyruvate Carboxykinase)
Tb927.3.3270	TbPFK ATP-dependent phosphofructokinase; 6-phospho-1-fructokinase	DKTDYIM**Y**NPRPR	TTHERM_00170320	(Phosphofructokinase family protein)
Tb09.211.3510	ATP-dep DEAD/H RNA helicase	FDVDF**Y**DRPR	TTHERM_00190830	(P68-like protein, putative)
Tb09.160.1160	Hypothetical Protein	IGGD**Y**EWSNTLAR	TTHERM_00827160	(Hypothetical Protein)
Tb927.4.3130	Hypothetical Protein	VGVTSG**Y**ANTQGR	TTHERM_00161270	BBC31
Tb927.3.1400	Hypothetical Protein	YLQSQNCRPTGNAG**Y**GG GNGPASSAEVR	TTHERM_00355780	MEC17

Appendix B-2. The amino acid sequences of each TTHERM

TTHERM_00058800

MNQYKLLSKKGEGTFSEVLKAQSLKEGKFVAIKCMKNHFNSIEQVQRLKEIQALRKVSPHPHIIKLI
EVLYDEPTGRLALVFELMDQNLYEAIRGKKQYLNQQKLNKHIKGEKRPAYYLSQRINNQKFDIFK
VTIPISVKFYMYQLLKAIDHMHKKGIFHRDIKPENILLLGDHVKLADFGSCKGIYSEHPYTEYISTR
WYRAPECLLTDGYYSSKMDLWGVGCVMFEVMSLFPLFPGNDELDQAHKIHNVLGTPNPKILEQF
QKHATHMELNFPPKKGTGIEKLAPHIPKDCVDLIYKLLCYDPEERINAEQALQHPYFRDLYQMDQE
NQLLLQSTITSKTKLSPSNNSNMYNRTLPENQLNLNQSTNNNTNNNHTNSTTRLEVSPQHNDPQYQ
SKNLNTMKIYSQISGQGGQNNSSNSNLNTTTKYNPYRESAENVFYPVLTIKKATNPYHKKKGQQA
ASMKKTQHQQYPSFNLNLKIESNYKDSGTDEDADNNNRSKPQFLPPIKNQMNPNHMMQMETKVN
MMKGQNSVKQKPNNQYPPYKQKKSKKFNEEYLVLGKKPVFNKQ

TTHERM_00161270

MEEAPHLNEDQLNEIYSWVDSIPLSRPKKNIARDFSDGVLMAEVVAHCIPRLVELHNYSAANSVKQ
KYYNWETLNKKVFSKLGLLVCKTDITNIVECKPEVIERVLKLVKNHIEAFNSDPQNFSPEKSQKSFN
QIKRGGALMDRGNNKQRFQNQDNQYQQSQQDEIIWEKDQTISELRETIEILELKMKKLEDLIKLKD
NKIATLQNRMAEHGLY

TTHERM_00170320

MDLQKSLTDNKFFVAGLALGSGLITSLLLKQFNTEEEKVERKGSLKTVKRKEVKKQRTVFLHQTD
NTIPKEILNASNTLMEEVKNLESRTCNILKVPNLIIEVPQDEIPCHLSPLYGENNYREVFGGGGFLPD
TAFVYTGSYVLSDAVIGKNQSLDQSKRYVRAGPRKHNYFNPQQVKAAIVTCGGLCPGLNVVIREIF
MCLYFNYGVRTIYGIKYGYKGFYQYDWQELTVKDVKEIQRLGGTVLGSSRGGFDKDKIVDNLVK
HGITHVYCLGGDGTHRGIQALFNEVRSRKLNISIVGIPKTIDNDIPIIDKSFGFETSVEEAVKAIQSAH
VEAHCAEYGVGLVRLMGRYAGFIAMEATNASRDAHVCLVPEFKFNLYGEEGVLEYVYQRLLSKR
NCIVVVAEGAGEAVLDAQFANTGEKDASGNAKLNDIGAFLQKEIVDYGKKKGMEVTLKYINPTY
MIRTVPANALDRKMCTQLAQNAVHGAMAGFTGFTVGHVNNRLCYIPLDEICKEGSGRRIKAEDRA
WQRLLASTGQPSFLKKEN

TTHERM_00185770

MNKYEVKGVVGEGAYGVVLKCMNKENGEYVAIKKFKESDEDEVVKKTTLREVKILRMLKQENI
VQLREAFRRKGKLYLVFEYVENNLLEILEERPNGLDQDDVRKYIYQLCKSISYCNSMDIIHRDIKPE
NLLISKDGTLKLCDFGFARVLPQKGGNLTDYVATRWYRAPELLLGYTDYGKEVDMWAVGCIMGE
LTDGQPLFPGQNEIDQLYVIQKVLGPLTAEQQEVFLKHPGFLGVKLPEISKPETIEKRYLGKLSKKA
LSFMKNLLKMDPSQRMSADEALQHPYFDGIRDKEIQQTIKHGRIEERVESATIKEPKQINQNGTNSG
FYNQQTQNASTIKPKDKPPQQQAIPPKNQTQQPPLGSNKQKALQQQPDQQQQRVNNFVKNGVRQ
EKNTPSHENNRIIKKDIYPHPQINFMQSSYSNIPDIYIKTKNAQNQQYNYEIQENYQPDFHNEDPDDS
QQPSVPQVLIRNDTHGNHNNMVNVSIEDNTPYSKFNERNKSKENIIPKGKKKKSQMSNIENFYQDN
AENSNNQEEDAEGRKSPLLTNNKLYGTSVAQQKKFRFNVMAQNIQQNEEAQEYPRLINIYQKQQQ
SRANSRSKFYKNGQLQKNANYVSSQQMNVLLHTEEGPDYEEKGLNLNSNTNNNNMNVNNNNISS
QNMMEQQVMGNEIFQSTSQLPYIQQSSSNYNPNFKFYNMYHQQQKQTQKPSGKMSTQHNNNTN
QNMNYDQYTYSMHPYDQQDQEMKNLNIIYNNNTYNYNINNSPLWNSISKKKI

TTHERM_00190830

MSRQIQSNSNKGEGSDYESHQSSDNQQKQGNGYFNNRNVNGQNQNGYQSNGYNNRGGYQKSYG

TNKYNNGNGYGNSNGGYFNNNNKPNGYFNANKKFGNEEQKTNVDNQKPYGYNQQNNNQSNKD
LNNSGYSDENHLGENLHDIDYTKVELKPFQKVFYQVGKSIHTDEEIATYQREKGIIIRSKHKEVPQP
FIKWNETKFPKYIMSVIEDSKFSEPMPIQAQYVTNKKQKKKYKMYECSFIPFPIVLSGHDLIGIAQTG
SGKTLSFMLPALVHINAQDPVKPGEGPIALVLAPTRELANQIQEQCFKFGSKCKISSVCVYGGAPKI
YQEKELRNGCDIVIATPGRLIDFLESNVIDLKRVTYLVLDEADRMLDMGFEPSIRKIVGQIRPDRQTL
MFSATWPQTVRRLALDFCHGDPIHIQIGDMENNVNNDIDQQVEIIDKSQKYDRVKEILSTMTRSDK
TIIFTQTKKDCDDLSKALQTDNIRNICIHGDKSQRDRDKVMDLFKTGRVNTLIATDVASRGLDVKDI
KLVINYDFPKQIEDYVHRVGRTGRAGAQGKAISFLDQYEDKKISKELVDVLKQNNQEISQDLLELS
EANYKGNYSNNYNKNRSSKPYSGGYNNNRNYYGNNNEMGFRNSNNGFSNNGYQNGNSNGGYR
SQQNNNGFQNNNYNRDFPKRQEDGNGGSEQPKKFGFFNPKKDN

TTHERM_00195990
MKGQSMYQNSNNTSGNNFSYMNYNQNQSIPSSYYQQQQNQQSSQQPLQQQNQNITSYSRSQSRH
SHHQSENVNISATNNTNNMQSQNQKYVKQSTSSMSHTQSNQVYQNSQLRENYNNLSNNSNAKSS
QIQSSQTGNIRPSTVANILNNNSTNKISQNTIQNLKQTNQVIPNNSKVNNENRVQNNTITSNQNTLTN
NTNNVISSSINANNNNQQSNNQQISSISASLKASTTNTKNQIQKKRKTFECGGTAFSVDDHYEFHKQ
IGHGAYGVVCSGVDTKCNKKIAIKKITNAFEDLVDAKRIVREIKMLKFFDHENIIALHDIIVPEQRTD
YNDIYIVTELMETDLHRVIYSRQELTDDHIQYFLYQTLRGMLYIHSANVMHRDLKPSNILVNKNCD
LKVCDLGLARGFEFDEEDGKTEYVVTRWYRAPEVILKASKYTKSIDVWSIGCIFAELLGRTPLFPG
KDYLEQIQRIIAILGTPTPEELSYITNEGALKYLKSLPKRTKQSWENLYPNANLVGLDLLSKMLTFNP
NDRYTIEECLAHPYFEGLHNPEDEPTCPEVFDWAWDDFELTKERLQGMVYDEAIDYQKVKATQQ
QQYQQQQSYENQC

TTHERM_00219410
MGSKDRTKNYREYRGSISKKYKNNQSGNYSKFQNRQNSRSRSRSQQQRNAPAQYQNSFQNNEKL
RYDNNNNRYQQNQYSQNDRDFRGGSVERRYDQGPNNQYSRNEGRDNNNFGRNNYKNDTGYSR
GGRDYSPQRSNGVDSRRKSLSQKNQRYYSQNQYSQGGNQAQSYPYQQNSFNQAQQSHYGNSYP
NHQQSSYNNNNNNSNSNYKKENVEDYYKNRDYKDWNNRDGNRNDRQNNNYSSNNNGNHNNS
NNSRFTNNPSGHSTINNSNNSYDKRNESYRQANTFPSDQYSRGNGSGNGNQNSFNKNDNNSNYNS
YNGNRSYKNEYQNSSQGHKQYGNNNNNYDNKRVSSQNYRNLTGGGGSYNNNGNFRRSNSNRNR
SNNNYKNYGRDNNNFRRNNSPRRNYNRYDNRNYERRDSQSRRTDKYQNLDRKYHHSDRSNKFQ
KANKSYYNSNRSNSIKYKGQKDNQYNRQGYQQSDGTSKKYYKDNKSKQTHSSSSKSSSSKDSFVS
GASKDRSKDAGHYKFKNGETFDDGRYVVVKHLSDGTFGRVLEVQDRQNGNKLYALKVIRAVER
YVDAAQIETEIIRKIQAADPENKYRIVKLESTFKHGANFCMIFEKLGLSLYELLKKNNYIGYKIRLVQ
SFFKQILESVGFLHSIHLTHTDLKPENILLASPELRLVVENNSQKDSPSTQYQSSDKTNATAGTVSEN
NQKNEHQNNSTGKRSSSSSSSSSSSSSSSGSHAKSSSKNSKKYWVPVSDEIKIIDFGGATFEDEHHSD
IINTRQYRAPEVILGCCKWNEVSDVWSIGCIIMELYSGELFFPTHEKIEHLAMIEKVCGPIPQWMAQ
RSDNSVNSHFKVTEEHFQQNGTYFDWPKNDTSRKSISNVKNLLSFDDILKPEHRILGDLVRKCLIID
PNQRITCAQALQHEFFSMKFD

TTHERM_00267860
MEKYNIISEIGRGAFGIVKKAQNKETKEIVAIKQMLQEYETWDECINLRELKSLRKLTHVNIIKLKE
VFRVKKQLSFVFEYVEKNIYKLYENAKQDGATSLPENTIKSIVYQIASALSYMHKHGFFHRDLKPE
NLLISSDGIVKLIDFGLAREVRSRPPYTDYVSTRWYRAPEILLRSTHYNSPVDIFALGCIMAELYLMK
PLFNGSSEIDQIQKITSVLGTPQKQDWPDGFILASTKYYTFPQYPAIPLSQVIPNCPPDALNLISEMLK
WDPQKRITAAKILQHPYFSNVELPEELTAESNSNQMIQSSNQPATLLGGASSNNAESVWNQSGNNN
ANLNQFQGNVNASQNVPANSRLGG

HSNTANQNSSQFGLSSINSTSSYHLGNYNNHQRQLSPQTNYDNNSLNTIKHGHGHFHYQSSSPDNR
YTQSQFRSKNIQYIIQNNRNNQRSQTDLIGLKNNFELNQQLQRQQQFSNQNQNQISKKQFDPIYEFL
NNDVRTINNPNTGSISTTNNVNSIASNNNNNNNNNLGGQIINPLEKGSRFSVPQGISSKNMIDDEIED
MEQEILGTYLPTGGNVSRVRNSNNNNTLSQNNNGSTLFGNNQENQASMFSNPRGGNRDRNVKDN
TLSNQDSNTLGYVPSAIARNQPSNLDIYRFDPKPQANTYNPSFAANNNMNNGSKINNDIYNFDNVI
TNNRKIQLPQ

TTHERM_00348940

MYIFNTFIQFCSQTKEAEKVQSTRNSFAIDQKNKDSILQSNISQPESTKNHFGYQNQLAHQQIQQQQ
LQQQQQSLQQMQQQQQQMTLQQSQNQLNQQHQQQLIMMIQQQQQQQHGQQLPNQTSSKKSNQ
QQQMQQNQIQQSQYSNVLNLLMQQQPTQPPQQHQQIYQNLNNSQQYTTMQQSSGQNYSNQSFTT
NRQNSNANYLYHQLNPPQSTKAADSNRRQNNPFSGTYSTKNTSAPQQLLKDILSKINKEQVQQQL
PYQQNQQVLRSSHMMNNSQNNSNSNVVILNNNTQNANNQNIYSSNYVNTNPDLSNIYNPNNLAA
MNPQPSSTKSNRIHSHYFPATTKNKSNRNPSLDKLNDSSFLRLKNVQSPKQRTIRTIMPGSNRDNSV
GNNSSINRNGYFQREYSESGVDASNQYILPASLLQQKTKRSGNSITGGVNNVIANNNNSYNNGSYS
KIQPHLPQNQTRQNGTPYVKKDLTIKTGNSIETLKDILSPTNNNANPKTTTATKDSKKHLNSNTLPI
QSYTTTHAQQIHQQKRLLQKRDDQKNKSNFINSSFQYIMASENNSVENTNRFSYNVMNNMVANN
QGQKLLKNKNNSMENQQFKNLSFTQPLNDQFKQTKENNNTQNSNNKILTVKQQQQKQQLQFIQK
GLGQNSSKQNQSSIQTSEGNENKSESNYESQGQNLDNFDQHKIDLSTASQTPFRNGNSTNEKKTLD
DNNTRNTERDEKKEEPLDWEKVKVPITPLEVLAKYQKYLTPQEKAEILEFKRIYFIGHNANKAAQE
NGNNDDTKGDYQYNVNDHIAYRYEIKEFLGKGSFGQVLKVFDHKRKEYAALKVIRNKQKFHEQA
LIELNILHYIKEKDYDNQTNIVKIRDFVIFRNHVCLVFELLSINLYDLLRNNKFQGLSLELIRRFAIQL
LNAISFLKENRIIHCDLKPENVLLKQPNKSGIKIADFGSSCFDDQVMYTYIQSRYYRSPEVILGIPYGT
EIDMWSFGCIIAELFLGYPIFPGEDENEQMGYIMELLGAPDSEFLKKCSRKKYFFDQNDQPLQIPNS
RGKIRMPNSKNWKSVIGCTDDNFIDFIKKCLVWEPSQRLTPRDALLHPWILDGLPEDIRVQHIKFLK
DKSFETASEKEFHKQSCLQAQYNADAQSAAMALALQKIYKYNNKKRILEKKSTTQGTVSEETRNS
VVN

TTHERM_00355780

MFKNQLLSLSLRIQKLRLCRQGNQLKVLQKSINQLLFKTQSIDQILQYLQQKKKTHQSILENQNQHI
HSIGLVSQLTNQKINKQKNQKAMEFNFIINKLVQLDQQGLGVYIPRASRSKVSSQQEQQLGQVLNT
MGERSAIAQGLKQVITNYDKVQGTDQRVYIVAEGRTCQGFLKVGQKNLFYRDMMGNIKEIKPLC
VLDFYVHESCQRQGYGKLLFEYMIQCEQTSPEKLAYDRPSPKLIAFLKKHYNLVKYIAQNNNFVVF
DQYFRSDASSQNKQNQNTRSYSQPYSDYSSQIPTNYPQQQQQQSNSKSYPYKQENNIDLMQHSSSR
NNKEFLNAGRAILSKEEIYKKDNLSQQNIENTLNNINNSQYSTKSQQQQQYQKDYQLDKYENNWG
ADKNIKKPPFPGQDRQLDKIQQKIQQTERELDVVNQQIIKQRQNLSQDPLTQNHRAQNVYNTNQFG
TSPWAQTGFNYYSTSSSNYGNHYTYYKK

TTHERM_00420030

MLSKLLKFNGATLQGTGKYYGLDRFGIVNKNILYNPSVAQLYEIATSCIPPSDPWTATNQISSTGAL
VAYSGQRTGRTPKEKRIVLNEASKDTIWWGDVNIPVCEESHQFCRDLTLKYLNQKKRVYIIDGYA
GWDPKYRLKVRVVCSRSYHALFMKNMLIRPTAEELARDFSDDKNIDFHIFNAGEMKAPVPIKDLT
SETTVQVNFKSNEIVILGTQYAGEMKKGVFGVMHYKMPQKGILSLHSSATEGAKGDITLFFGLSGT
GKTTLSADPNRKLLGDDEHCWSDDGVFNIEGGCYAKAIDLSREKEPEIFDAIRFGAVLENVKFLDK
STREVNYTDISITENTRACYPLEHIPNAKFPSVGGHPNNIIFLTCDAYGVLPPVAKLTPEQAMYHFIS
GYTAKVAGTEVGIVEPQATFSACFGEAFLPLHPTFYADMLAKKIQKHNTHVWIVNTGWSGGKYG

VGKRMSLKVTRKIIDAIHSGELEKAEYTNLELFNFKVPKQITGVDSSILMPKNTWTDKAAYDAELR
KLAGMFQKNFKKYQDKATKEVINAGPVL

TTHERM_00469230
MQQQNQQQAQQQLQQQQQQQQHQQAQQHQEKPQKPIQVPPGKHKFDCGQTIFVVDEKYE
YIKQIGYGAYGVVCSAVNKENGSKVAIKKVQNAYDDLIDAKRIVREIKLLKFFEHDNIISLLDVIKP
NAPVGDEDIYMVFDLMETDLHRVIYSRQELSDEHIQYFVYQILRGLLYIHSANVIHRDLKPSNILLN
KNCDLKICDFGLARGFEEPGESLTEYVITRWYRAPEVILNASEYNQAVDIWSVGCITAELLGRTPLF
PGENYLDQVQRVIAVLGTPTQEDMAYIGNQDAIKYIKSQPKRTKQSWQSLYPKANPLALDLLSKM
LTFNPEKRYTVKECIEHPYFEGLHYPDEEPECEQVFDWSWDKFTPTKEILQKMIYQEAKSFHPIH

TTHERM_00483640
MENLENLETIEDRYQKIKKIGEGTYGKVFQAKDLKTDQIVALKKVKNDYEEIGDEGIPSTALREISC
LKALDHPNVVKLVDVVYIMKKNKLYLVFEYIDYDLKAYQKKIGKIPEQTVKSYMHQILKGIEHCH
SRRIFHRDLKPQNILVNNKGDIKIADFGLGKIFGIPLNTITHEVETLWYRAPEILLGDKLYSLPVDVW
SIGCIFAELIIGHPLFHGDSEIDQIFKIFQFFGTPKSSKLEGTYNLKYWSNLFPRFKSQKDEILKPIIETD
PQAADLLVKLIDIEPAKRISVSQALKHPYFDNIQSPLNSQIFNF

TTHERM_00529780
MNQLQYNQNNDYQQNGDIFNQIQRPKPTQEYEIYKRLKLLGEGSFGKAYLVESQSDKSKWVIKQI
SLDAMSPEEKKESYKEAKILEQLNHPNIVKFKEIYKTKSGKLCIVMEYADGGDLSQKIQKQRGKYF
KEEQILDWFTQICLAMKHVHDRDLKGQNIFLTSQNICKLGDFGIARVLNKTFEKAKTMVGTPYYLS
PEIINSVPYSYKSDVWSIGVVLYEMCCLRPPFQGESLQNLALNIVKGQYQPIPNIYSQDLKKLVSNL
LQNRPESRYSIQQILGLPFIRNRIKNFLSETIKANEFSHTILHNQKIIIDDDFKNIKNVQGPPPPSVAYQ
MKEQQRNAQEKQDQQQQQKVLPSIEP
KQAAQKAPQNQKQNQIEKQQIRPQTPVSNQVRPINGKIPIQNQQYQAQNQRPSNIDVGPQNYPYGK
IPNSQQQTPIGKENSSDNKKIVRQPSKNNIVNVKYNSVPTEQEEKISQQPSKQTPPLPPQPQLKQNPS
QQKFQKAPQKPVVQKLQRPSTAVPQNPQRQIMTPKSQPGSKPFYPPKPIDRKQSAGQILVSNQVSK
GDQNNKPPISNNRPITAHNRNSTPQRISNTPPVQLQRNVRQEISQKRQDILNKVKDAQEQKKKKLE
DLKKAKMEASLKSKEEREQVRQNMYKDIKQKRQGSKPSQNVQQVQQQQQQNKKEEIQIEWMGT
LKPDQPPERPKYIDQKLEEQKIAAEKYNQEMALRKKKEEEIEKQKQLQEQMYEENKREREKKQRE
LFLMDMQENSNPQNYPQKGFAHRNVAPQNRQEIEYTQMLEEMKSLIDEPPCTEDNTTNDSDQQQN
INQNLNFEDVDRDDYDDQFDNNESQGNTNQKQKNNNIQQQQDEEEYIYTLPNNKQSINAIKENYV
KQIGKENLQKLQQAIKSEIKKNRKFVDLSPSDKIKILKQQVPSIKFDKVDLKINQLVDIVQIENS

TTHERM_00532680
MNMRIPTQNYDKEVLPTQADEGIGSRVHTVGSMKENEYQKKQQQKIQMTYFYNEKQQAGNYTY
TIVGDEDKVVGSGTFGVVYQATTKETGEVVAIKKVFQDKRYKNRELQMMKEIGNHPNVIKLRNH
YYSYGNSTDDVYLHLVMDFVPETLYKMIKYYSKKHKGNFPNILLKYYSYQMLRSLAYIHGINICH
RDIKPQNILVDPRNHNLKMCDFGSAKKLVPGESNISYICSRCYRAPELMFQATQYTHAIDVWSVGC
VIAEMVLGQPIFIGESSVDQLIEIIKVLGTPTQQQIFAMNPDHQGTKMPNIKPTPWTKVFQNCRIDPL
AIDLISKILVYNPEKRLKPLEALLHPFFEELRNPKCRINGKPLDNLFNFTEAEIGPNEELLQRLIPDWY
VQQLTEAQNQY

TTHERM_00537240
MSEDIEPHVLRRYEILSKLGRGAYGIVWKVYDKRTKQVLALKKIFDAFQNSTDAQRTFREIMFLQE
LDHENIIKVINVIRAKNDRDIYVVFEYMDTDLHAVIRVNILEDIHKQYIMYQIFRAIKYIHSGELIHRD
LKASNILVNSDCMVKVADFGLVRSIANQENGSTPILTEYIATRWYRAPEILLGIIELSQCSHTYTKGV
DMWSIGCILGQLLLGKPIFAGTSTLNQLELILQVTGKPTYEDIEAIQSDLAITMLEAVQNNPTQNTKT
LQQMIPMASDDALDLLQNLLQFNPKKRITAEQALSHPYVRQFHNPNDEPVCGRIIQIPLDDNKKYS
MRFYRDKLYYEISPRKNVQQNGKQDSQLCSNFSESQDQIQQPQQIQHQQQTVNEQQVLQQRTNNQ
QFQSKDNSIQNARNNSNSSNSSQVKTNQSTYNNIQSTLLRQKSNGIIQKQENGNMQQNQNIQSSDQ
QQNLAKEVIQQTSAVVAAQQAVLARKTQKQNSIQFQQNIYNEVQESIYQQVKTQKQQLYQQQLM
QQQQQQQKMQQQSQNQVISSKNQIQYQSSNINKKQNEENEIRDPSENSIKQQQKLYASPKAPSSN
QILSNNQNLQQTLTNNIMSTNNTNGIHNTISNGFAKRPTPFLQSKNSSQQTSNSTALNQKSSSNQNS
QQILQQQQVKQMLNSPNTQQMIQQQKSKEINQNIFSQQNIQNPHLLNGNQYISSNNTINNNNDGQI
DNVDEGITSPVSQNILKQSNSNRLIKPNNLKSNSSKALQSPTSNIVSPQHQQQSSVSSQSTQQSHPSS
AIQNQNSQVQQKPASINTSQKYLQRIIQGATPSTQQENNLKQAISHQNQNNIIQQQYQNYINSNLST
NNNNSNPSHQMNNQNVQNTNVQSQSNGYSQNNNGYPTQSNSGYQQLNNGYPQHTNIYSQNSSN
NIYTQQQQQQLSGQLKKGSTSINQYLTTTYKNPTQTMQQQMYKGNQIDFYSPNKSKVNTANYFG
NRASVTKTMAQLNQKNNSNSAIIPNSNTSNQIQNNNQLNTNNNNNSNKNYQTGQNGHIILNGNQS
VSHNIHAQNSVSAGNMANNNLNSLNGQSSTKNKYSYSNLLQNAQKLKANTPSYHQLNSPKY

TTHERM_00557780
MYNNMTSNKPNNRLNDEINQNHQKLDHKGPPKDLKIIDVVGNGTFGVVYKKLDDNINFNYLKMK
ALLNKEQVVAVKRVFQDKRYKNREREIIERLDHQNIIKCKDFFETKGEKKDEVYLNLVMEFMNFP
LSKVIRDKRKANQVMRIEDIRLYAYQMFKALSYLQSLNICHRDIKPQNILVNDKDDDKRLKICDFG
SAKQLIPNEANISYICSRYYRAPELIFKSQHYTNAIDMWSMGCVIAEMVLGEPLFPGENSLDQLVEII
RILGTPTRQQILMMNPTSQEFRFPQIKSYSWSKVFEKSLQKVSEDEKQKFNQLMDLISRVLVYAPSE
RPSPLQALAHPFFDELRIEEKKYRQTQLYLFDFPEAKEQEYQKQIEQLKLIPNGYSKKSYS

TTHERM_00575520
MSERTNMQGKLRKQKSEEKERHLKTNQQNLQEKLLEGEEKLTNQSRSGTIKMTLSCVFQQSYQSF
IKEIILIEALVANHQQKLSLGQQFQKQQRAAAKKTKIQPTTNKIDYKGRKQKFQQQTVIEKETKNTK
EEDHQKFLEAEKEFFQLKFIQTVKIIERMSSNEEIEPHILRKFEIIQKLGKGAYGIVWKAFCKKSKQIV
ALKKVFDAFHNATDAQRTFREVMFLQELNGHENIIRLINIIKAENNKDLYMVFDFMETDLHAVIRA
NILEEIHKQYVVYQILKAMKYIHSGELIHRDLKPSNILLNSECLVKLADFGLARSVAVTDDDDNPVR
TEYVATRWYRAPEILLGSTKYAKAVDMWSIGCIVGELITNRAIFPGNSTLNQIEKVLELTGKPTQED
IESLESQLAWNIISSINVQKKKTFHQMFPGASEEAIDLIKKLLCFNPKNRLTVEQALEHPYVKDFHCP
EEEIVCGRVIPISMNDNKKFSIKEYREALYNQIYQEKRIKRQQNRDIYLKQQQQITTTGNGQVSQAG
QSSQPVSSQQIPSSSQSTQSSVTQQQQQQQQEKSKNQYQSSSSQQQQAQQYQQQQSLQQQQVPQ
RQSSQSSQPQSQQSIQQNQQKYQQSQQQQQRYFDQEKEQYKPNAQQQYLQQQQQLQQQQALQQ
QQMQKSYYNSQNFNSSQVQVQQGNSQSTSQGNNYSGYKSGSQASTTNASSSAQYAGYDDKKVL
MNSSNQIQNKNSSQVYTQSQGANQASSSNTQQQYLPSSSNTIYGQGVTNSIAN
SNNLYNQAQLNGSKQQSSSLGHTQNKYSSTQAYQQAQPSKYMNQTQNAQKYVNNNNNSNGQLG
GSQNNIGGSQTNSQNHYQAAQQYLQQHMQQQQQQQQVQQPQTKYGVTYHHGHSQSSGSSNTFA
QYQQHYQQQQQRTTGVTANHQRQPSNNAQKVRKVEQKM

TTHERM_00827160
MSNNINPQSNKNNVSKQGQVDSTSKTQILQQIKQTQIKETQQQNLPNKDRGSQIYSTIRKKSRQED
QQDLYSVIGQNHRKLNIVEDDIIEKKIMDAYKLKEQLRNQLSHIDQRLENLSTQRIDNPKSIYSHPVL
KQYLETDEERKQQQASRFVKKIKSQQAHFAEQLKQKVEELDYKKKLEEAQVQRQIIQQKQNSEKE
REERLKRQIKLIKERKDLRIKQLNEFEELYKKKIKKDQSKRPLYIRLIKKFNMKEQFFKEITLKEKQE
KRKQYLQLAKIDFKQHEKNYFENLQKLEQIKKENREKQLQLLQQKQPEPNLIYDSSKLLKNEIVDN
QFHKKLETESIQRENALKIVEARNKFSNALKQHQQVSPSKQKQLEMEIIIDKSANPHKYITQVKKKN
ENDRYQEYLERGKLVRKKSADFSVRKSTSTMSLLNARSFQTPQMASKNALVSQDDEMIFKQQQK
VGFVKQEPGKERELGNLYLKSIRESSRKYQEYLNQRKQLTESSANEKLMIDIVNDSYQKEIKGPFSP
INTSSSHQYTLKSTQNSKLPIISNKQLKNKELGSSFSQQNINSQSENIARGQSSDKYSNINQDTKSNSE
QNEKKIFKKYNYMNEIREKYNLDQRKEKWEQIMESKKSVPILEKKKLLEEAYYKYDKEIMLKEQV
VRSHRPQELSLKEVEDRQILDDAYLKSIKAKIEMMKLTIDNDDFVKKEQPKVRQIFQNQEQYYNQY

TTHERM_01035490
MAFVIDKDSKNFHIEAAQPELERYQRTEKIGEGTYGIVYKAIDMQTNDIIALKKIRLEHEDEGVPST
AIREISLLKEIDHPNVIKLRDLVYGENKLYLIFDYLDHDLKKYLELNGGPLPPAVVKDYLFQLILGIA
VCHANRIVHRDLKPQNILINKKGSVQLADFGLARAFGLPLKTYTHEVVTLWYRPPEILLGQKQYST
PVDIWSIGCIFSEMAQKIPLFIGDSEIDQIFKIFRIMGTPSESTWPGVTQLPDFKNTFPRWNPIPLQKQC
PNICPKGIDLLTKMLQLDPTKRITAEEALDHPYFDDLDKSNYKLLD

TTHERM_01043090
MERDLLNPIYEEEKQKNKFKRLIQAPNSYFMDVKCAQCQNIQMIFSNAQSTIIFLTQSSSKSNKQNE
FGQKLLPQFQEKLKIILMYFQQLASLQSLSQIYKQIGQQTNNQNKQTNILHNQIFVNVAESYQTEAII
QNFRDIQFCLHPDVQQLINIKRSKIQKCKTFSNFIRIFNSLQSILSIQIYNYQKLNRCSAILCKPTGGKV
QIQAGCAFKIKN

TTHERM_01347900
MKKYKIGATLGDGTFGVVMKAVDETNGQVVAIKKMKSKKCPKWEEIIKLPEITSLMKFHHPNIVN
LYEIIKHNNELFFVFEYMDQNVYQMTKDREKPFTENQIRNIIYQTLQGLAYIHRHGYFHRDLKPENL
LESNNTIKIADFGLAREIRSKPPFTDYVSTRWYRAPEVILRANNYNSPIDIFAIGCIMAELYKLWPLFP
GQNELDQILQICKVMGTPSKEEWPEGYKLASGVGYNFPQYKPQPLQELIPNASPEAIDLLQKMLRY
SPQKRISAYAALQHPFFSCNIPIPESIKMKNDKILQEGSVLIDKRNFNDSSVKNDNSQDQKRQEIQQI
DKQQLLKNLKASQSIDNLINDLLQESQIINNNLENQFNSQKQQKSYIVKTGSQMRNSKYKPGTTCT
PYTAQN

Appendix B-3. Alignment between the *T. brucei* peptide and corresponding TTHERM_ amino acid sequence.

CRK1 - Tb10.70.7040

TTHERM_00483640 (Netphos score: 0.569)

```
pep             -------------------IGEGSYGVVFR------------------------------ 11
pro             MENLENLETIEDRYQKIKKIGEGTYGKVFQAKDLKTDQIVALKKVKNDYEEIGDEGIPST 60
                                   ****:** **:
```

MAPK11- Tb10.389.1730

TTHERM_00469230 (Netphos score: 0.923)

```
pep             --------------------------------------EESDQ----------------- 5
pro             IISLLDVIKPNAPVGDEDIYMVFDLMETDLHRVIYSRQELSDEHIQYFVYQILRGLLYIH 180
                                                      * **:

pep             ------------------------------------GEHMTDYVTMR------------- 16
pro             SANVIHRDLKPSNILLNKNCDLKICDFGLARGFEEPGESLTEYVITRWYRAPEVILNASE 240
                                                    ** :*:**  *
```

MAPK1- Tb10.6k15.2790

TTHERM_00575520 (Netphos score: 0.648)

```
pep             ---------CFN-------------TQGGDNDL-TEYIATR------------------- 18
pro             PSNILLNSECLVKLADFGLARSVAVTDDDDNPVRTEYVATRWYRAPEILLGSTKYAKAVD 360
                         *:              *:..** : ***:***
```

GSK3α - Tb927.7.2420

TTHERM_00532680 (Netphos score: 0.898)

```
pep             ---------------------------------------------NVPYIFSR------- 8
pro             SLAYIHGINICHRDIKPQNILVDPRNHNLKMCDFGSAKKLVPGESNISYICSRCYRAPEL 240
                                                             *:.** **
```

PK - Tb927.3.690

TTHERM_00267860 (Netphos score: 0.944)

```
pep             -----------------------------------ARPPYTDYVSTR------------- 12
pro             HKHGFFHRDLKPENLLISSDGIVKLIDFGLAREVRSRPPYTDYVSTRWYRAPEILLRSTH 180
                                                   :***********
```

PK - Tb927.7.3580

TTHERM_00529780 (Netphos score: 0.849)

```
pep             -----FN-------GG-----------------YNSSSVSGNG------VVK-------- 17
pro             MNQLQYNQNNDYQQNGDIFNQIQRPKPTQEYEIYKRLKLLGEGSFGKAYLVESQSDKSKW 60
                     :*       .*                 *: .:  *:*      :*:
```

MAPK2 - Tb10.70.2070

TTHERM_00537240 (Netphos score: 0.099)

```
pep             -------------------------------------------EQVARPVLTDYIMTR-- 15
pro             RAIKYIHSGELIHRDLKASNILVNSDCMVKVADFGLVRSIANQENGSTPILTEYIATRWY 180
                                                           *: : *:**:** **
```

PK - Tb11.01.4230

TTHERM_00219410 (Netphos score: 0.673)

```
pep             KVTY--ALPNQSR----------------------------------------------- 11
pro             QSHYGNSYPNHQQSSYNNNNNNSNSNYKKENVEDYYKNRDYKDWNNRDGNRNDRQNNNYS 240
                :  *  : **:.:
```

PK - Tb10.329.0030

TTHERM_00557780 (Netphos score: 0.647)

```
pep             DAQASDT------F--------------YVCTR--------------------------- 13
pro             NDKDDDKRLKICDFGSAKQLIPNEANISYICSRYYRAPELIFKSQHYTNAIDMWSMGCVI 240
                : : .*.      *              *:*:*
```

PK - Tb927.8.3770

TTHERM_00195990 (Netphos score: 0.935)

```
pep             -------------------------------------------------EDTQDPNKTHY 11
pro             IQYFLYQTLRGMLYIHSANVMHRDLKPSNILVNKNCDLKVCDLGLARGFEFDEEDGKTEY 420
                                                                 *  :: .**.*

pep             VTHR-------------------------------------------------------- 15
pro             VVTRWYRAPEVILKASKYTKSIDVWSIGCIFAELLGRTPLFPGKDYLEQIQRIIAILGTP 480
                *. *
```

PK - Tb09.211.0960

TTHERM 00058800 (Netphos score: 0.731)

```
pep      ----GLHVSQPLTEYVSTR----------------------------------------- 15
pro      GSCKGIYSEHPYTEYISTRWYRAPECLLTDGYYSSKMDLWGVGCVMFEVMSLFPLFPGND 240
             *:: .:* ***:***
```

CRK3 - Tb10.70.2210

TTHERM 01035490 (Netphos score: 0.919)

```
pep      --------------------------MDILGEGTYGVVYR-------------------- 14
pro      MAFVIDKDSKNFHIEAAQPELERYQRTEKIGEGTYGIVYKAIDMQTNDIIALKKIRLEHE 60
                                    : :******:**:
```

PK - Tb11.02.0640

TTHERM 00348940 (Netphos score: 0.308)

```
pep      -LFTYIQSR--------------------------------------------------- 8
pro      VMYTYIQSRYYRSPEVILGIPYGTEIDMWSFGCIIAELFLGYPIFPGEDENEQMGYIMEL1020
          ::******
```

ECK1 - Tb11.01.8550

TTHERM 00185770 (Netphos score: 0.830)

```
pep      -----------------------------------GNYTEYVATR--------------- 10
pro      DIIHRDIKPENLLISKDGTLKLCDFGFARVLPQKGGNLTDYVATRWYRAPELLLGYTDYG 180
                                            ** *:*****
```

MAPK9 - Tb10.61.1850

TTHERM 01347900 (Netphos score: 0.963)

```
pep      -------------------------------SRPPFTEYVSTR----------------- 12
pro      GYFHRDLKPENLLESNNTIKIADFGLAREIRSKPPFTDYVSTRWYRAPEVILRANNYNSP 180
                                        *:****:*****
```

MAPK2 - Tb10.61.0250

TTHERM_00195990 (Netphos score: 0.935)

```
pep             ---------------------------------------------------------DDQ 3
pro             RIVREIKMLKFFDHENIIALHDIIVPEQRTDYNDIYIVTELMETDLHRVIYSRQELTDDH 360
                                                                         **:

pep             CT----QT--------SA------------------------L--------------TEY 13
pro             IQYFLYQTLRGMLYIHSANVMHRDLKPSNILVNKNCDLKVCDLGLARGFEFDEEDGKTEY 420
                      **        **                        *              ***

pep             VVTR-------------------------------------------------------- 17
pro             VVTRWYRAPEVILKASKYTKSIDVWSIGCIFAELLGRTPLFPGKDYLEQIQRIIAILGTP 480
                ****
```

PK - Tb927.6.1780

TTHERM_00195990 (Netphos score: 0.935)

```
pep             ----------------------------GVGVN-----------------------VTSY 9
pro             IQYFLYQTLRGMLYIHSANVMHRDLKPSNILVNKNCDLKVCDLGLARGFEFDEEDGKTEY 420
                                            .: **                        *.*

pep             VVTR-------------------------------------------------------- 13
pro             VVTRWYRAPEVILKASKYTKSIDVWSIGCIFAELLGRTPLFPGKDYLEQIQRIIAILGTP 480
                ****
```

CRK2 - Tb927.7.7360

TTHERM_01035490 (Netphos score: 0.919)

```
pep             -----------------------------VGEGSYGIVYK-------------------- 11
pro             MAFVIDKDSKNFHIEAAQPELERYQRTEKIGEGTYGIVYKAIDMQTNDIIALKKIRLEHE 60
                                             :***:******
```

PK - Tb10.61.3140

TTHERM_00532680 (Netphos score: 0.898)

```
pep             ---------------------------------------LAADEPNVAYICSR------- 14
pro             SLAYIHGINICHRDIKPQNILVDPRNHNLKMCDFGSAKKLVPGESNISYICSRCYRAPEL 240
                                                       *...*.*::*****
```

Ribosomal Protein S27 - Tb11.01.1475

TTHERM_01043090 (Netphos score: 0.028)

```
pep             ---------------------LVQGPNSYFMDVK-------------------------- 13
pro             MERDLLNPIYEEEKQKNKFKRLIQAPNSYFMDVKCAQCQNIQMIFSNAQSTIIFLTQSSS 60
                                     *:*.*********
```

Glycosomal Phosphoenolpyruvate Carboxykinase - Tb927.2.4210

TTHERM_00420030 (Netphos score: 0.028)

```
pep             -----------------------------EQVILGTEYAGEMK----------------- 14
pro             IFNAGEMKAPVPIKDLTSETTVQVNFKSNEIVILGTQYAGEMKKGVFGVMHYKMPQKGIL 240
                                             * *****:******
```

TbPFK ATP-dependent Phosphofructokinase;6-phospho-1-fructokinase - Tb927.3.3270

TTHERM_00170320 (Netphos score: 0.799)

```
pep             -----------------------------------DKTD-YIMYNPR------PR----- 13
pro             YREVFGGGGFLPDTAFVYTGSYVLSDAVIGKNQSLDQSKRYVRAGPRKHNYFNPQQVKAA 180
                                                   *::. *:  .**      *:
```

ATP-dep DEAD/H RNA Helicase - Tb09.211.3510

TTHERM_00190830 (Netphos score: 0.046)

```
pep             ---FDVDFYDR--PR------------------------ 10
pro             NNGFQNNNYNRDFPKRQEDGNGGSEQPKKFGFFNPKKDN 699
                   *: : *:*  *:
```

Hypothetical Protein - Tb09.160.1160

TTHERM_00827160 (Netphos score: 0.056)

```
pep             ---------------IGGDY-------EWSNTLAR------------------------- 13
pro             QNEKKIFKKYNYMNEIREKYNLDQRKEKWEQIMESKKSVPILEKKKLLEEAYYKYDKEIM 660
                               *  .*       :*.: :
```

Hypothetical Protein - Tb927.4.3130

TTHERM 00161270

```
pep             --------------V--G---VTSG-----YANTQGR----------------------- 13
pro             QNFSPEKSQKSFNQIKRGGALMDRGNNKQRFQNQDNQYQQSQQDEIIWEKDQTISELRET 180
                              :  *   :  *     : * :.:
```

Hypothetical Protein- Tb927.3.1400

TTHERM 00355780 (Netphos score: 0.059)

```
pep             -----------------------------------------YLQSQ--NCR---PTGN-- 12
pro             SQQEQQLGQVLNTMGERSAIAQGLKQVITNYDKVQGTDQRVYIVAEGRTCQGFLKVGQKN 180
                                                         *: ::  .*:    .*:

pep             -----------------------------AGYG----------GGNGPASSAEVR----- 28
pro             LFYRDMMGNIKEIKPLCVLDFYVHESCQRQGYGKLLFEYMIQCEQTSPEKLAYDRPSPKL 240
                                              ***            ..* . *  *
```

Appendix B-4-1. NetPhos results for each TTHERM_ amino acid sequence corresponding to the first table

Pos	Context	Score	Prediction
TTHERM_00058800			
4	-MNQYKLLS	0.317	.
71	IEVLYDEPT	0.762	*Y*
90	DQNLYEAIR	0.662	*Y*
99	GKKQYLNQQ	0.288	.
117	KRPAYYLSQ	0.171	.
118	RPAYYLSQR	0.074	.
142	SVKFYMYQL	0.022	.
144	KFYMYQLLK	0.014	.
187	CKGIYSEHP	0.936	*Y*
192	SEHPYTEYI	0.898	*Y*
195	PYTEYISTR	0.731	*Y*
201	STRWYRAPE	0.030	.
212	LTDGYYSSK	0.595	*Y*
213	TDGYYSSKM	0.573	*Y*
298	VDLIYKLLC	0.153	.
303	KLLCYDPEE	0.010	.
319	LQHPYFRDL	0.013	.
324	FRDLYQMDQ	0.513	*Y*
353	NSNMYNRTL	0.145	.
394	NDPQYQSKN	0.975	*Y*
405	TMKIYSQIS	0.341	.
428	TTTKYNPYR	0.117	.
431	KYNPYRESA	0.157	.
440	ENVFYPVLT	0.200	.
452	ATNPYHKKK	0.256	.
471	QHQQYPSFN	0.158	.
484	IESNYKDSG	0.972	*Y*
540	PNNQYPPYK	0.191	.
543	QYPPYKQKK	0.534	*Y*
555	FNEEYLVLG	0.963	*Y*
TTHERM_00185770			
4	-MNKYEVKG	0.965	*Y*
15	GEGAYGVVL	0.703	*Y*
29	ENGEYVAIK	0.981	*Y*
77	KGKLYLVFE	0.044	.
82	LVFEYVENN	0.580	*Y*
106	DVRKYIYQL	0.054	.
108	RKYIYQLCK	0.102	.
116	KSISYCNSM	0.492	.
161	NLTDYVATR	0.830	*Y*
167	ATRWYRAPE	0.011	.
176	LLLGYTDYG	0.175	.
179	GYTDYGKEV	0.051	.
212	IDQLYVIQK	0.467	.
253	IEKRYLGKL	0.095	.
287	LQHPYFDGI	0.025	.
330	NSGFYNQQT	0.078	.

			Prediction
410	KKDIYPHPQ	0.907	*Y*
422	MQSSYSNIP	0.866	*Y*
429	IPDIYIKTK	0.627	*Y*
440	QNQQYNYEI	0.036	.
442	QQYNYEIQE	0.045	.
448	IQENYQPDF	0.019	.
492	DNTPYSKFN	0.905	*Y*
523	IENFYQDNA	0.942	*Y*
551	NNKLYGTSV	0.182	.
579	EAQEYPRLI	0.188	.
586	LINIYQKQQ	0.010	.
601	RSKFYKNGQ	0.338	.
612	KNANYVSSQ	0.916	*Y*
630	EGPDYEEKG	0.989	*Y*
677	SQLPYIQQS	0.413	.
685	SSSNYNPNF	0.192	.
692	NFKFYNMYH	0.015	.
695	FYNMYHQQQ	0.033	.
723	QNMNYDQYT	0.103	.
726	NYDQYTYSM	0.478	.
728	DQYTYSMHP	0.312	.
733	SMHPYDQQD	0.752	*Y*
747	LNIIYNNNT	0.403	.
752	NNNTYNYNI	0.365	.
754	NTYNYNINN	0.182	.
TTHERM_00195990			
7	GQSMYQNSN	0.531	*Y*
20	NNFSYMNYN	0.944	*Y*
23	SYMNYNQNQ	0.058	.
33	IPSSYYQQQ	0.035	.
34	PSSYYQQQQ	0.314	.
57	NITSYSRSQ	0.243	.
90	QNQKYVKQS	0.765	*Y*
107	SNQVYQNSQ	0.931	*Y*
116	LRENYNNLS	0.175	.
258	VDDHYEFHK	0.953	*Y*
269	GHGAYGVVC	0.110	.
332	QRTDYNDIY	0.916	*Y*
336	YNDIYIVTE	0.954	*Y*
351	HRVIYSRQE	0.038	.
363	DHIQYFLYQ	0.020	.
366	QYFLYQTLR	0.032	.
374	RGMLYIHSA	0.141	.
420	GKTEYVVTR	0.935	*Y*
426	VTRWYRAPE	0.008	.
438	KASKYTKSI	0.450	.
466	PGKDYLEQI	0.069	.
487	EELSYITNE	0.970	*Y*

496	GALKYLKSL	0.121	.
512	WENLYPNAN	0.946	*Y*
535	PNDRYTIEE	0.914	*Y*
545	LAHPYFEGL	0.019	.
581	QGMVYDEAI	0.014	.
587	EAIDYQKVK	0.858	*Y*
598	QQQQYQQQQ	0.084	.
604	QQQSYENQC	0.588	*Y*

TTHERM_00219410

10	RTKNYREYR	0.031	.
13	NYREYRGSI	0.008	.
21	ISKKYKNNQ	0.779	*Y*
29	QSGNYSKFQ	0.159	.
54	APAQYQNSF	0.150	.
66	EKLRYDNNN	0.979	*Y*
73	NNNRYQQNQ	0.891	*Y*
78	QQNQYSQND	0.598	*Y*
94	VERRYDQGP	0.654	*Y*
102	PNNQYSRNE	0.802	*Y*
118	GRNNYKNDT	0.205	.
124	NDTGYSRGG	0.954	*Y*
131	GGRDYSPQR	0.024	.
153	KNQRYYSQN	0.430	.
154	NQRYYSQNQ	0.233	.
159	SQNQYSQGG	0.825	*Y*
169	QAQSYPYQQ	0.033	.
171	QSYPYQQNS	0.265	.
184	QQSHYGNSY	0.673	*Y*
188	YGNSYPNHQ	0.659	*Y*
196	QQSSYNNNN	0.843	*Y*
207	SNSNYKKEN	0.880	*Y*
215	NVEDYYKNR	0.773	*Y*
216	VEDYYKNRD	0.980	*Y*
221	KNRDYKDWN	0.944	*Y*
239	QNNNYSSNN	0.877	*Y*
272	SNNSYDKRN	0.896	*Y*
279	RNESYRQAN	0.931	*Y*
290	PSDQYSRGN	0.913	*Y*
313	NNSNYNSYN	0.742	*Y*
316	NYNSYNGNR	0.066	.
322	GNRSYKNEY	0.544	*Y*
326	YKNEYQNSS	0.958	*Y*
336	GHKQYGNNN	0.399	.
343	NNNNYDNKR	0.898	*Y*
353	SSQNYRNLT	0.629	*Y*
363	GGGSYNNNG	0.836	*Y*
383	SNNNYKNYG	0.976	*Y*
386	NYKNYGRDN	0.027	.
403	PRRNYNRYD	0.030	.
406	NYNRYDNRN	0.425	.
411	DNRNYERRD	0.569	*Y*
424	RTDKYQNLD	0.979	*Y*
431	LDRKYHHSD	0.169	.

447	ANKSYYNSN	0.798	*Y*
448	NKSYYNSNR	0.873	*Y*
458	NSIKYKGQK	0.167	.
466	KDNQYNRQG	0.665	*Y*
471	NRQGYQQSD	0.574	*Y*
481	TSKKYYKDN	0.117	.
482	SKKYYKDNK	0.976	*Y*
519	DAGHYKFKN	0.671	*Y*
532	DDGRYVVVK	0.968	*Y*
559	GNKLYALKV	0.059	.
570	AVERYVDAA	0.971	*Y*
593	PENKYRIVK	0.951	*Y*
620	GLSLYELLK	0.044	.
628	KKNNYIGYK	0.465	.
631	NYIGYKIRL	0.053	.
689	PSTQYQSSD	0.905	*Y*
748	NSKKYWVPV	0.258	.
780	NTRQYRAPE	0.030	.
809	IMELYSGEL	0.130	.
863	QNGTYFDWP	0.835	*Y*

TTHERM_00267860

4	MEKYNIIS	0.949	*Y*
39	MLQEYETWD	0.109	.
81	FVFEYVEKN	0.855	*Y*
87	EKNIYKLYE	0.658	*Y*
90	IYKLYENAK	0.130	.
111	KSIVYQIAS	0.113	.
119	SALSYMHKH	0.446	.
160	SRPPYTDYV	0.917	*Y*
163	PYTDYVSTR	0.944	*Y*
169	STRWYRAPE	0.030	.
181	RSTHYNSPV	0.663	*Y*
198	MAELYLMKP	0.625	*Y*
239	ASTKYYTFP	0.778	*Y*
240	STKYYTFPQ	0.454	.
245	TFPQYPAIP	0.426	.
287	LQHPYFSNV	0.040	.
383	STSSYHLGN	0.484	.
388	HLGNYNNHQ	0.090	.
401	PQTNYDNNS	0.873	*Y*
418	GHFHYQSSS	0.139	.
427	PDNRYTQSQ	0.873	*Y*
439	KNIQYIIQN	0.573	*Y*
490	FDPIYEFLN	0.969	*Y*
570	ILGTYLPTG	0.029	.
637	NTLGYVPSA	0.766	*Y*
653	NLDIYRFDP	0.170	.
664	QANTYNPSF	0.023	.
685	NNDIYNFDN	0.852	*Y*

TTHERM_00348940

2	MYIFNT	0.399	.
54	NHFGYQNQL	0.315	.

143	QQSQYSNVL	0.717	*Y*
166	HQQIYQNLN	0.633	*Y*
175	NSQQYTTMQ	0.727	*Y*
186	SGQNYSNQS	0.438	.
202	SNANYLYHQ	0.561	*Y*
204	ANYLYHQLN	0.407	.
230	FSGTYSTKN	0.703	*Y*
260	QQLPYQQNQ	0.358	.
300	NQNIYSSNY	0.803	*Y*
304	YSSNYVNTN	0.973	*Y*
315	LSNIYNPNN	0.093	.
339	IHSHYFPAT	0.024	.
400	NRNGYFQRE	0.108	.
405	FQREYSESG	0.597	*Y*
416	ASNQYILPA	0.616	*Y*
449	NNNSYNNGS	0.804	*Y*
454	NNGSYSKIQ	0.824	*Y*
473	NGTPYVKKD	0.924	*Y*
525	PIQSYTTTH	0.088	.
559	SSFQYIMAS	0.706	*Y*
576	NRFSYNVMN	0.311	.
678	SESNYESQG	0.986	*Y*
751	VLAKYQKYL	0.020	.
754	KYQKYLTPQ	0.238	.
770	FKRIYFIGH	0.144	.
793	TKGDYQYNV	0.834	*Y*
795	GDYQYNVND	0.205	.
803	DHIAYRYEI	0.010	.
805	IAYRYEIKE	0.068	.
830	KRKEYAALK	0.256	.
855	NILHYIKEK	0.050	.
861	KEKDYDNQT	0.967	*Y*
892	SINLYDLLR	0.058	.
963	DQVMYTYIQ	0.200	.
965	VMYTYIQSR	0.308	.
970	IQSRYYRSP	0.141	.
971	QSRYYRSPE	0.070	.
982	LGIPYGTEI	0.008	.
1002	LFLGYPIFP	0.036	.
1016	EQMGYIMEL	0.142	.
1037	SRKKYFFDQ	0.122	.
1140	LQAQYNADA	0.013	.
1157	LQKIYKYNN	0.095	.
1159	KIYKYNNKK	0.139	.

TTHERM_00469230

62	VDEKYEYIK	0.974	*Y*
64	EKYEYIKQI	0.358	.
70	KQIGYGAYG	0.213	.
73	GYGAYGVVC	0.056	.
97	VQNAYDDLI	0.198	.
140	DEDIYMVFD	0.983	*Y*
155	HRVIYSRQE	0.038	.
167	EHIQYFVYQ	0.057	.
170	QYFVYQILR	0.026	.
178	RGLLYIHSA	0.441	.
223	SLTEYVITR	0.923	*Y*
229	ITRWYRAPE	0.007	.
241	NASEYNQAV	0.288	.
269	PGENYLDQV	0.522	*Y*
290	EDMAYIGNQ	0.581	*Y*
299	DAIKYIKSQ	0.631	*Y*
315	WQSLYPKAN	0.153	.
338	PEKRYTVKE	0.665	*Y*
348	IEHPYFEGL	0.256	.
354	EGLHYPDEE	0.454	.
384	QKMIYQEAK	0.244	.

TTHERM_00483640

14	IEDRYQKIK	0.953	*Y*
25	GEGTYGKVF	0.569	*Y*
49	VKNDYEEIG	0.980	*Y*
84	VDVVYIMKK	0.315	.
92	KNKLYLVFE	0.140	.
97	LVFEYIDYD	0.254	.
100	EYIDYDLKA	0.383	.
105	DLKAYQKKI	0.021	.
120	TVKSYMHQI	0.010	.
181	ETLWYRAPE	0.048	.
193	GDKLYSLPV	0.052	.
244	LEGTYNLKY	0.217	.
248	YNLKYWSNL	0.913	*Y*
300	LKHPYFDNI	0.256	.

TTHERM_00529780

6	NQLQYNQNN	0.494	.
12	QNNDYQQNG	0.975	*Y*
31	PTQEYEIYK	0.665	*Y*
34	EYEIYKRLK	0.849	*Y*
49	FGKAYLVES	0.035	.
79	KKESYKEAK	0.971	*Y*
100	FKEIYKTKS	0.949	*Y*
113	IVMEYADGG	0.174	.
130	QRGKYFKEE	0.070	.
194	VGTPYYLSP	0.111	.
195	GTPYYLSPE	0.609	*Y*
206	NSVPYSYKS	0.219	.
208	VPYSYKSDV	0.028	.
220	GVVLYEMCC	0.042	.
246	VKGQYQPIP	0.610	*Y*
253	IPNIYSQDL	0.035	.
273	PESRYSIQQ	0.794	*Y*
333	PSVAYQMKE	0.242	.
405	QNQQYQAQN	0.268	.
422	GPQNYPYGK	0.022	.
424	QNYPYGKIP	0.441	.
461	VNVKYNSVP	0.877	*Y*
538	SKPFYPPKP	0.843	*Y*

651	RQNMYKDIK	0.824	*Y*
701	ERPKYIDQK	0.972	*Y*
716	AAEKYNQEM	0.489	.
741	QEQMYEENK	0.921	*Y*
769	NPQNYPQKG	0.261	.
789	QEIEYTQML	0.940	*Y*
834	DRDDYDDQF	0.920	*Y*
865	DEEEYIYTL	0.986	*Y*
867	EEYIYTLPN	0.918	*Y*
883	IKENYVKQI	0.460	.

TTHERM_00532680

152	PETLYKMIK	0.969	*Y*
157	KMIKYYSKK	0.689	*Y*
158	MIKYYSKKH	0.013	.
173	ILLKYYSYQ	0.025	.
174	LLKYYSYQM	0.005	.
176	KYYSYQMLR	0.038	.
184	RSLAYIHGI	0.081	.
229	SNISYICSR	0.898	*Y*
235	CSRCYRAPE	0.028	.
247	QATQYTHAI	0.116	.
337	KILVYNPEK	0.008	.
395	IPDWYVQQL	0.281	.
406	AQNQY----	0.155	.

TTHERM_00537240

13	VLRRYEILS	0.021	.
24	GRGAYGIVW	0.094	.
31	VWKVYDKRT	0.019	.
88	DRDIYVVFE	0.923	*Y*
93	VVFEYMDTD	0.543	*Y*
114	IHKQYIMYQ	0.031	.
117	QYIMYQIFR	0.017	.
125	RAIKYIHSG	0.735	*Y*
174	ILTEYIATR	0.099	.
180	ATRWYRAPE	0.011	.
199	CSHTYTKGV	0.197	.
243	GKPTYEDIE	0.981	*Y*
309	LSHPYVRQF	0.049	.
336	DNKKYSMRF	0.187	.
341	SMRFYRDKL	0.305	.
346	RDKLYYEIS	0.422	.
347	DKLYYEISP	0.946	*Y*
433	NQSTYNNIQ	0.765	*Y*
508	QQNIYNEVQ	0.327	.
516	QESIYQQVK	0.962	*Y*
527	KQQLYQQQL	0.084	.
559	NQIQYQSSN	0.555	*Y*
589	QQKLYASPK	0.456	.
716	NGNQYISSN	0.880	*Y*
821	TSQKYLQRI	0.085	.
856	IQQQYQNYI	0.056	.
859	QYQNYINSN	0.443	.
894	QSNGYSQNN	0.857	*Y*
901	NNNGYPTQS	0.305	.
916	LNNGYPQHT	0.289	.
923	HTNIYSQNS	0.403	.
932	SNNIYTQQQ	0.869	*Y*
954	SINQYLTTT	0.220	.
959	LTTTYKNPT	0.702	*Y*
971	QQQMYKGNQ	0.074	.
979	QIDFYSPNK	0.106	.
991	NTANYFGNR	0.044	.
1043	SNKNYQTGQ	0.507	*Y*
1092	TKNKYSYSN	0.868	*Y*
1094	NKYSYSNLL	0.726	*Y*
1111	NTPSYHQLN	0.886	*Y*
1119	NSPKY----	0.878	*Y*

TTHERM_00557780

2	---MYNNMT	0.596	*Y*
49	FGVVYKKLD	0.181	.
60	INFNYLKMK	0.234	.
84	QDKRYKNRE	0.520	*Y*
118	KDEVYLNLV	0.910	*Y*
152	DIRLYAYQM	0.012	.
154	RLYAYQMFK	0.030	.
162	KALSYLQSL	0.242	.
209	ANISYICSR	0.647	*Y*
214	ICSRYYRAP	0.187	.
215	CSRYYRAPE	0.052	.
227	KSQHYTNAI	0.542	*Y*
292	QIKSYSWSK	0.026	.
326	RVLVYAPSE	0.042	.
353	EEKKYRQTQ	0.931	*Y*
359	QTQLYLFDF	0.013	.
371	KEQEYQKQI	0.514	*Y*
385	IPNGYSKKS	0.076	.
390	SKKSYS---	0.693	*Y*

TTHERM_00575520

62	FQQSYQSFI	0.196	.
111	NKIDYKGRK	0.549	*Y*
186	GKGAYGIVW	0.448	.
251	NKDLYMVFD	0.955	*Y*
277	IHKQYVVYQ	0.057	.
280	QYVVYQILK	0.032	.
288	KAMKYIHSG	0.427	.
337	VRTEYVATR	0.648	*Y*
343	ATRWYRAPE	0.011	.
355	GSTKYAKAV	0.593	*Y*
462	LEHPYVKDF	0.284	.
494	SIKEYREAL	0.051	.
499	REALYNQIY	0.881	*Y*
503	YNQIYQEKR	0.696	*Y*
517	NRDIYLKQQ	0.426	.

575	SKNQYQSSS	0.946	*Y*
588	QAQQYQQQQ	0.088	.
623	NQQKYQQSQ	0.675	*Y*
633	QQQRYFDQE	0.090	.
641	EKEQYKPNA	0.982	*Y*
649	AQQQYLQQQ	0.072	.
671	MQKSYYNSQ	0.177	.
672	QKSYYNSQN	0.547	*Y*
698	QGNNYSGYK	0.043	.
701	NYSGYKSGS	0.364	.
718	SSAQYAGYD	0.539	*Y*
721	QYAGYDDKK	0.380	.
743	SSQVYTQSQ	0.741	*Y*
761	TQQQYLPSS	0.043	.
770	SNTIYGQGV	0.816	*Y*
785	SNNLYNQAQ	0.819	*Y*
807	TQNKYSSTQ	0.662	*Y*
813	STQAYQQAQ	0.603	*Y*
821	QPSKYMNQT	0.706	*Y*
831	NAQKYVNNN	0.983	*Y*
860	SQNHYQAAQ	0.540	*Y*
866	AAQQYLQQH	0.063	.
887	PQTKYGVTY	0.645	*Y*
891	YGVTYHHGH	0.073	.
908	TFAQYQQHY	0.144	.
912	YQQHYQQQQ	0.133	.

TTHERM_01035490

24	ELERYQRTE	0.670	*Y*
35	GEGTYGIVY	0.919	*Y*
39	YGIVYKAID	0.186	.
91	RDLVYGENK	0.696	*Y*
97	ENKLYLIFD	0.687	*Y*
102	LIFDYLDHD	0.184	.
110	DLKKYLELN	0.253	.
126	VVKDYLFQL	0.016	.
178	PLKTYTHEV	0.014	.
187	VTLWYRPPE	0.009	.
199	GQKQYSTPV	0.040	.
304	LDHPYFDDL	0.135	.
313	DKSNYKLLD	0.654	*Y*

TTHERM_01043090

10	LNPIYEEEK	0.802	*Y*
29	APNSYFMDV	0.028	.
86	IILMYFQQL	0.006	.
100	LSQIYKQIG	0.480	.
129	VAESYQTEA	0.609	*Y*
185	SIQIYNYQK	0.030	.
187	QIYNYQKLN	0.016	.

THERM_01347900

4	MKKYKIGA	0.815	*Y*
67	IVNLYEIIK	0.186	.
82	FVFEYMDQN	0.620	*Y*
88	DQNVYQMTK	0.301	.
108	RNIIYQTLQ	0.466	.
116	QGLAYIHRH	0.048	.
122	HRHGYFHRD	0.033	.
159	PFTDYVSTR	0.963	*Y*
165	STRWYRAPE	0.030	.
177	RANNYNSPI	0.155	.
194	MAELYKLWP	0.320	.
228	WPEGYKLAS	0.400	.
236	SGVGYNFPQ	0.471	.
241	NFPQYKPQP	0.372	.
267	KMLRYSPQK	0.083	.
276	RISAYAALQ	0.117	.
381	QQKSYIVKT	0.146	.
394	RNSKYKPGT	0.805	*Y*
403	TCTPYTAQN	0.451	.

Appendix B-4-2. NetPhos results for each TTHERM_ amino acid sequence corresponding to the second table.

TTHERM_00161270

Pos	Context	Score	Pred
16	LNEIYSWVD	0.802	*Y*
58	ELHNYSAAN	0.095	.
68	VKQKYYNWE	0.529	*Y*
69	KQKYYNWET	0.034	.
158	QDNQYQQSQ	0.780	*Y*
215	EHGLY----	0.316	.

TTHERM_00170320

Pos	Context	Score	Pred
116	LSPLYGENN	0.666	*Y*
121	GENNYREVF	0.371	.
138	TAFVYTGSY	0.067	.
142	YTGSYVLSD	0.729	*Y*
161	QSKRYVRAG	0.799	*Y*
171	RKHNYFNPQ	0.814	*Y*
203	FMCLYFNYG	0.668	*Y*
206	LYFNYGVRT	0.007	.
212	VRTIYGIKY	0.221	.
216	YGIKYGYKG	0.458	.
218	IKYGYKGFY	0.116	.
222	YKGFYQYDW	0.092	.
224	GFYQYDWQE	0.014	.
270	ITHVYCLGG	0.018	.
340	HCAEYGVGL	0.070	.
351	LMGRYAGFI	0.035	.
379	KFNLYGEEG	0.243	.
387	GVLEYVYQR	0.078	.
389	LEYVYQRLL	0.090	.
443	EIVDYGKKK	0.273	.
455	VTLKYINPT	0.810	*Y*
460	INPTYMIRT	0.627	*Y*
504	NRLCYIPLD	0.064	.

TTHERM_00190830

Pos	Context	Score	Pred
17	EGSDYESHQ	0.978	*Y*
33	QGNGYFNNR	0.699	*Y*
47	NQNGYQSNG	0.943	*Y*
52	QSNGYNNRG	0.755	*Y*
58	NRGGYQKSY	0.511	*Y*
62	YQKSYGTNK	0.286	.
67	GTNKYNNGN	0.627	*Y*
73	NGNGYGNSN	0.896	*Y*
80	SNGGYFNNN	0.982	*Y*
90	KPNGYFNAN	0.755	*Y*
112	NQKPYGYNQ	0.176	.
114	KPYGYNQQN	0.069	.
131	NNSGYSDEN	0.959	*Y*
146	HDIDYTKVE	0.830	*Y*
159	QKVFYQVGK	0.097	.
174	EIATYQREK	0.073	.
204	KFPKYIMSV	0.740	*Y*
224	IQAQYVTNK	0.435	.
234	QKKKYKMYE	0.490	.
237	KYKMYECSF	0.030	.
321	SVCVYGGAP	0.049	.
328	APKIYQEKE	0.123	.
363	KRVTYLVLD	0.107	.
446	KSQKYDRVK	0.887	*Y*
534	LVINYDFPK	0.015	.
543	QIEDYVHRV	0.387	.
566	FLDQYEDKK	0.957	*Y*
598	SEANYKGNY	0.937	*Y*
602	YKGNYSNNY	0.955	*Y*
606	YSNNYNKNR	0.296	.
615	SSKPYSGGY	0.238	.
619	YSGGYNNNR	0.898	*Y*
625	NNRNYYGNN	0.171	.
626	NRNYYGNNN	0.832	*Y*
646	SNNGYQNGN	0.978	*Y*
655	SNGGYRSQQ	0.755	*Y*
669	QNNNYNRDF	0.046	.

TTHERM_00355780

Pos	Context	Score	Pred
48	QILQYLQQK	0.017	.
110	GLGVYIPRA	0.019	.
151	VITNYDKVQ	0.062	.
162	DQRVYIVAE	0.062	.
183	KNLFYRDMM	0.385	.
202	VLDFYVHES	0.090	.
212	QRQGYGKLL	0.059	.
219	LLFEYMIQC	0.020	.
233	EKLAYDRPS	0.884	*Y*
248	LKKHYNLVK	0.048	.
253	NLVKYIAQN	0.127	.
266	VFDQYFRSD	0.446	.
284	NTRSYSQPY	0.327	.
288	YSQPYSDYS	0.638	*Y*
291	PYSDYSSQI	0.117	.
299	IPTNYPQQQ	0.030	.
312	NSKSYPYKQ	0.129	.
314	KSYPYKQEN	0.357	.
348	KEEIYKKDN	0.973	*Y*
370	NNSQYSTKS	0.698	*Y*
380	QQQQYQKDY	0.080	.
384	YQKDYQLDK	0.095	.
389	QLDKYENNW	0.949	*Y*
454	AQNVYNTNQ	0.143	.
471	TGFNYYSTS	0.080	.
472	GFNYYSTSS	0.061	.
479	SSSNYGNHY	0.926	*Y*

483	YGNHYTYYK	0.310	.
485	NHYTYYKK-	0.028	.
486	HYTYYKK--	0.082	.

TTHERM_00420030

19	GTGKYYGLD	0.091	.
20	TGKYYGLDR	0.012	.
34	KNILYNPSV	0.053	.
42	VAQLYEIAT	0.263	.
70	ALVAYSGQR	0.011	.
118	LTLKYLNQK	0.118	.
126	KKRVYIIDG	0.468	.
131	IIDGYAGWD	0.122	.
138	WDPKYRLKV	0.383	.
150	CSRSYHALF	0.046	.
218	LGTQYAGEM	0.028	.
232	GVMHYKMPQ	0.027	.
297	EGGCYAKAI	0.198	.
336	REVNYTDIS	0.970	*Y*
349	TRACYPLEH	0.007	.
377	TCDAYGVLP	0.802	*Y*
393	EQAMYHFIS	0.144	.
399	FISGYTAKV	0.074	.
433	HPTFYADML	0.768	*Y*
460	SGGKYGVGK	0.639	*Y*
488	EKAEYTNLE	0.964	*Y*
522	DKAAYDAEL	0.278	.
540	NFKKYQDKA	0.837	*Y*

TTHERM_00827160

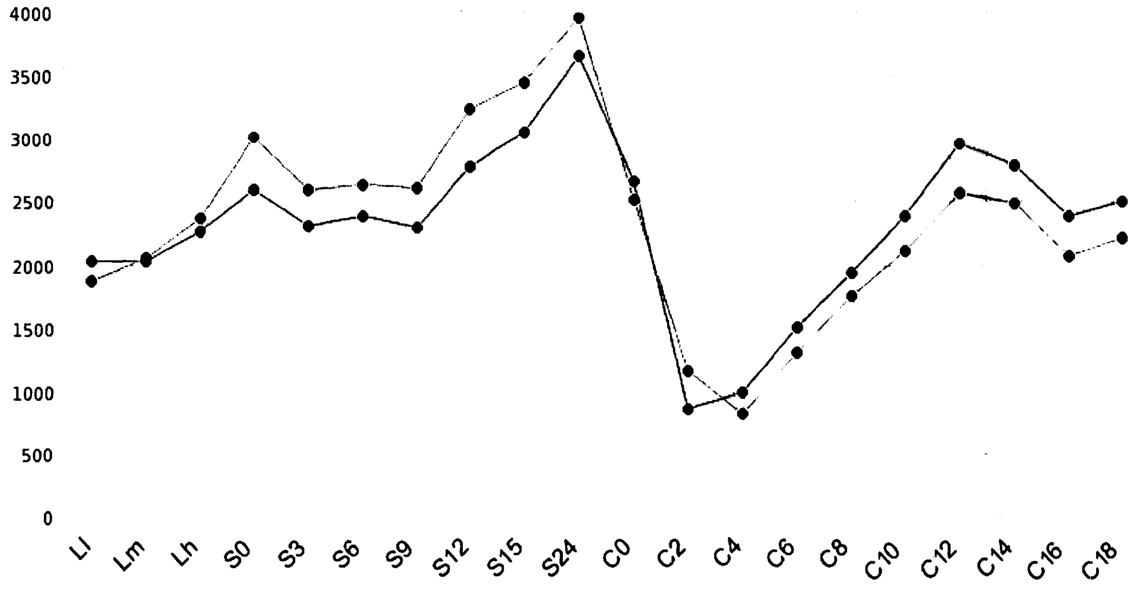
55	GSQIYSTIR	0.205	.
71	QQDLYSVIG	0.865	*Y*
97	IMDAYKLKE	0.506	*Y*
129	PKSIYSHPV	0.510	*Y*
137	VLKQYLETD	0.030	.
176	EELDYKKKL	0.974	*Y*
229	FEELYKKKI	0.941	*Y*
243	KRPLYIRLI	0.413	.
272	KRKQYLQLA	0.202	.
287	HEKNYFENL	0.564	*Y*
320	PNLIYDSSK	0.838	*Y*
392	NPHKYITQV	0.153	.
405	ENDRYQEYL	0.964	*Y*
408	RYQEYLERG	0.247	.
482	LGNLYLKSI	0.017	.
493	SSRKYQEYL	0.610	*Y*
496	KYQEYLNQR	0.165	.
521	VNDSYQKEI	0.477	.
540	SSHQYTLKS	0.394	.
588	SSDKYSNIN	0.991	*Y*
610	IFKKYNYMN	0.077	.
612	KKYNYMNEI	0.254	.
620	IREKYNLDQ	0.056	.
652	LEEAYYKYD	0.698	*Y*
653	EEAYYKYDK	0.854	*Y*
655	AYYKYDKEI	0.016	.
689	LDDAYLKSI	0.365	.
727	NQEQYYNQY	0.741	*Y*
728	QEQYYNQY-	0.446	.
731	YYNQY----	0.080	.

TTHERM_01043090

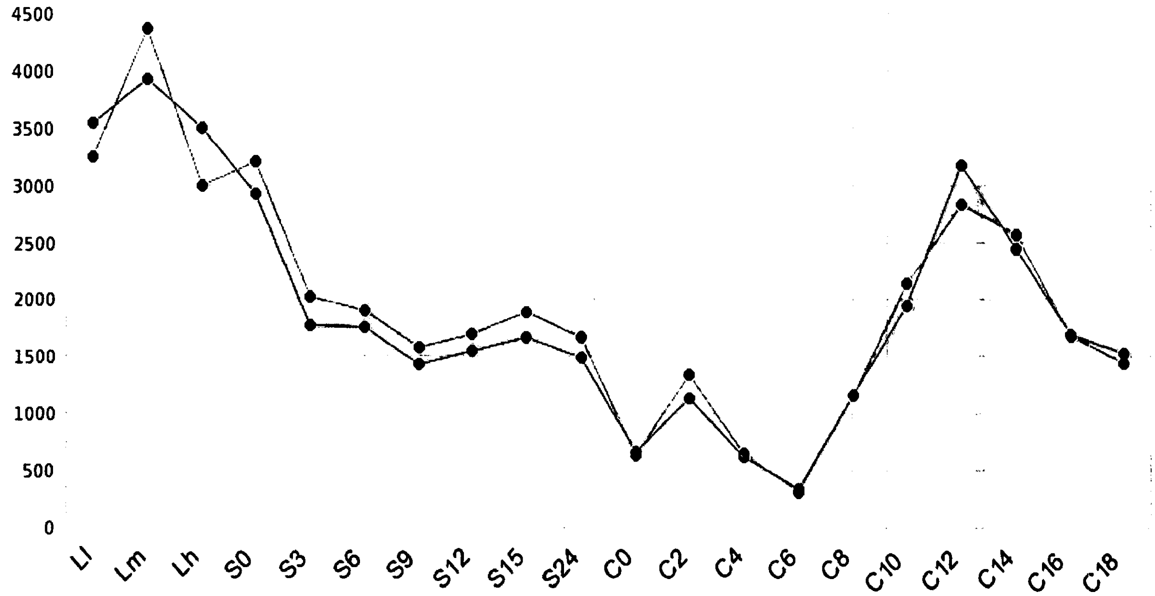
10	LNPIYEEEK	0.802	*Y*
29	APNSYFMDV	0.028	.
86	IILMYFQQL	0.006	.
100	LSQIYKQIG	0.480	.
129	VAESYQTEA	0.609	*Y*
185	SIQIYNYQK	0.030	.
187	QIYNYQKLN	0.016	.

Appendix B-5. Microarray gene expression profile for each TTHERM_ . Vegetative cells (L) were collected at ~1X10^5 cells/ml, and ~3.5X10^{5}cells/ml and ~1*10^6 cells/ml, starved cells (S) were collected between 0-18 hrs and conjugating cells (C) were collected between 0-18 hrs.

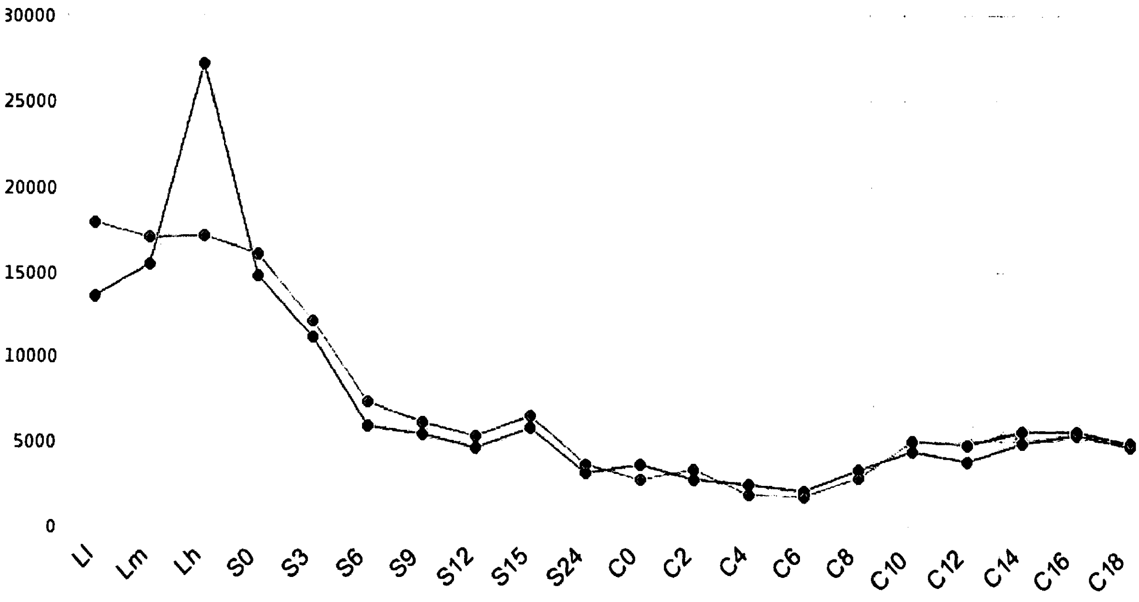
TTHERM_00058800



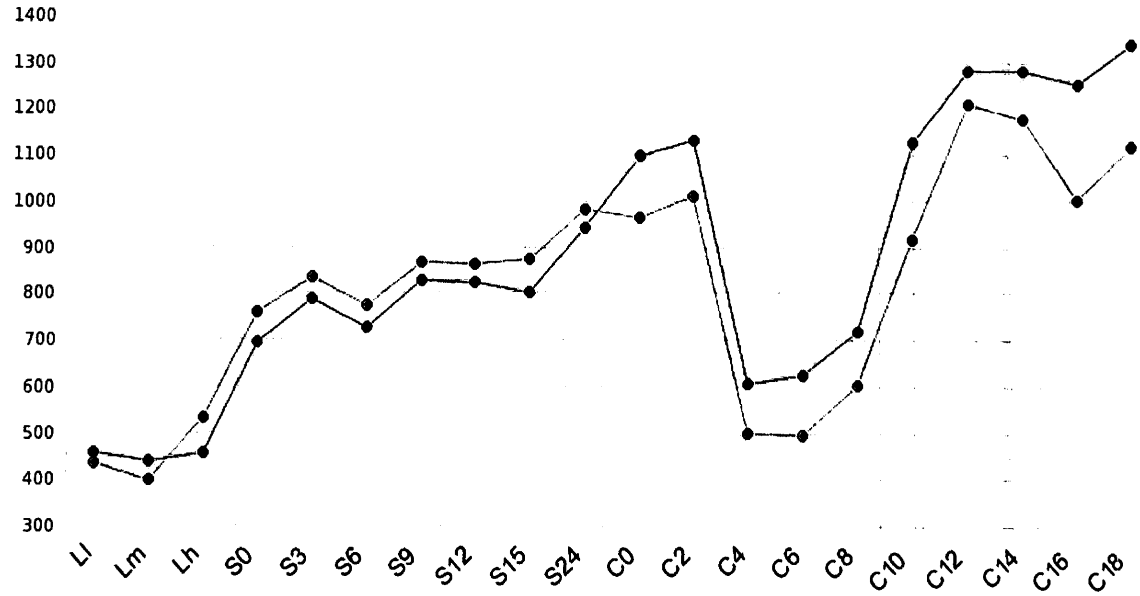
TTHERM_00161270



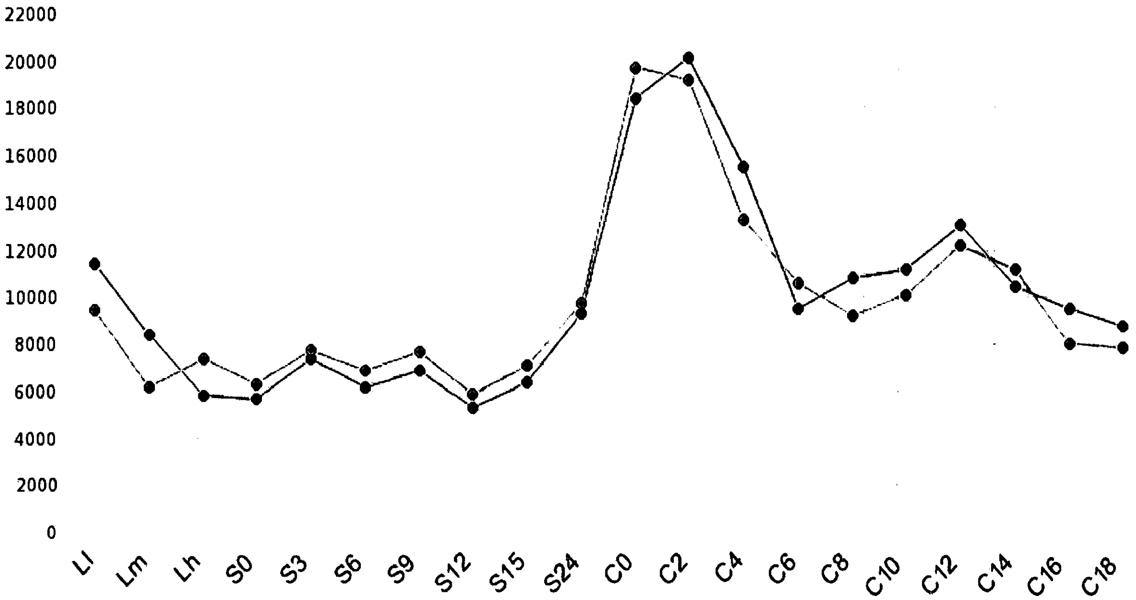
TTHERM_00170320



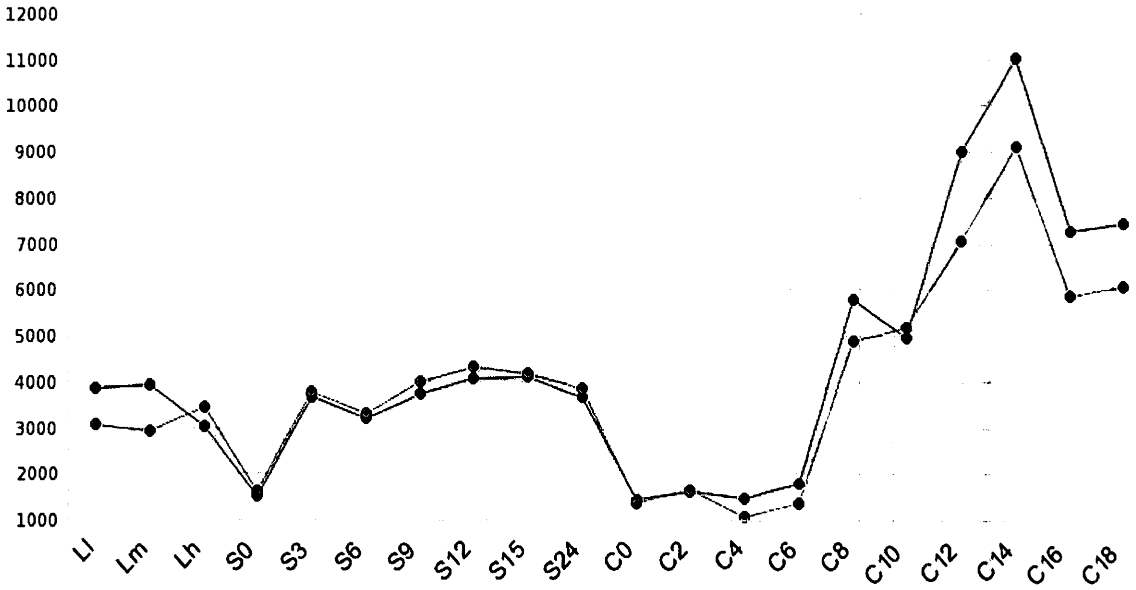
TTHERM_00185770



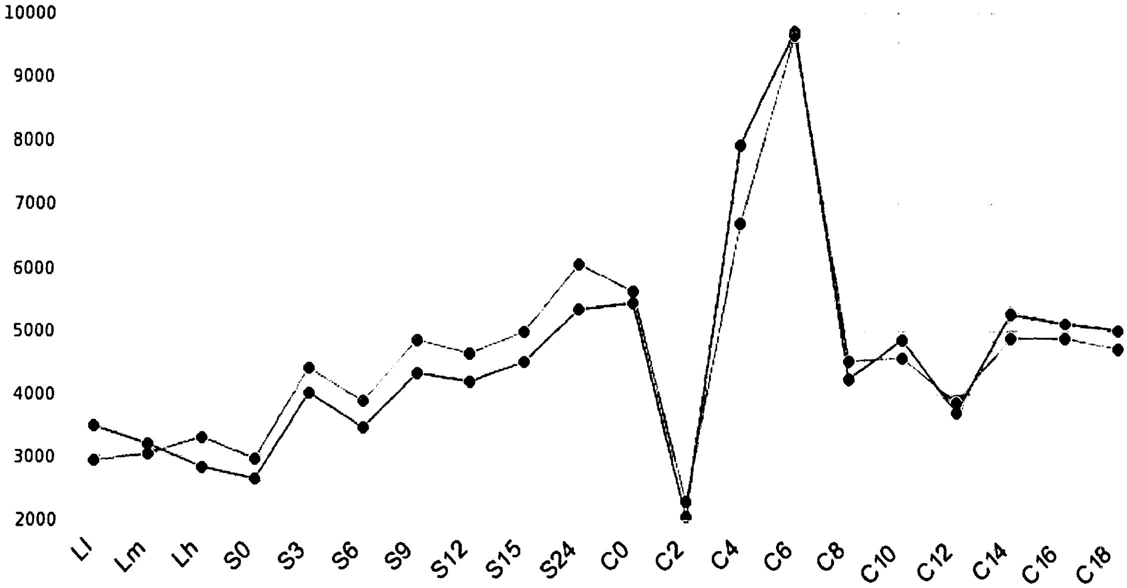
TTHERM_00190830



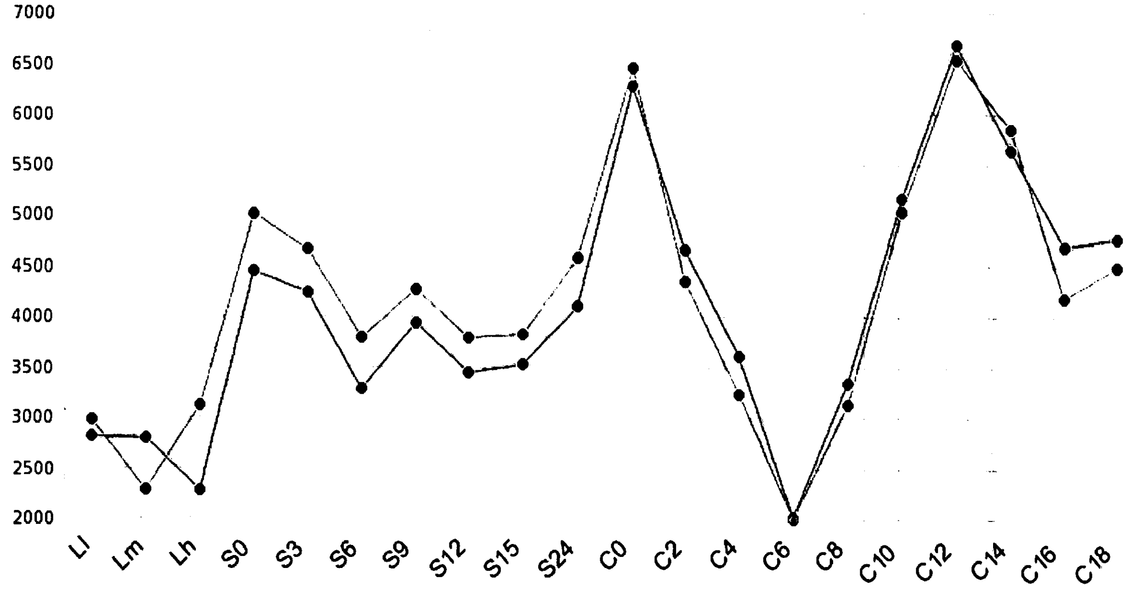
TTHERM_00195990



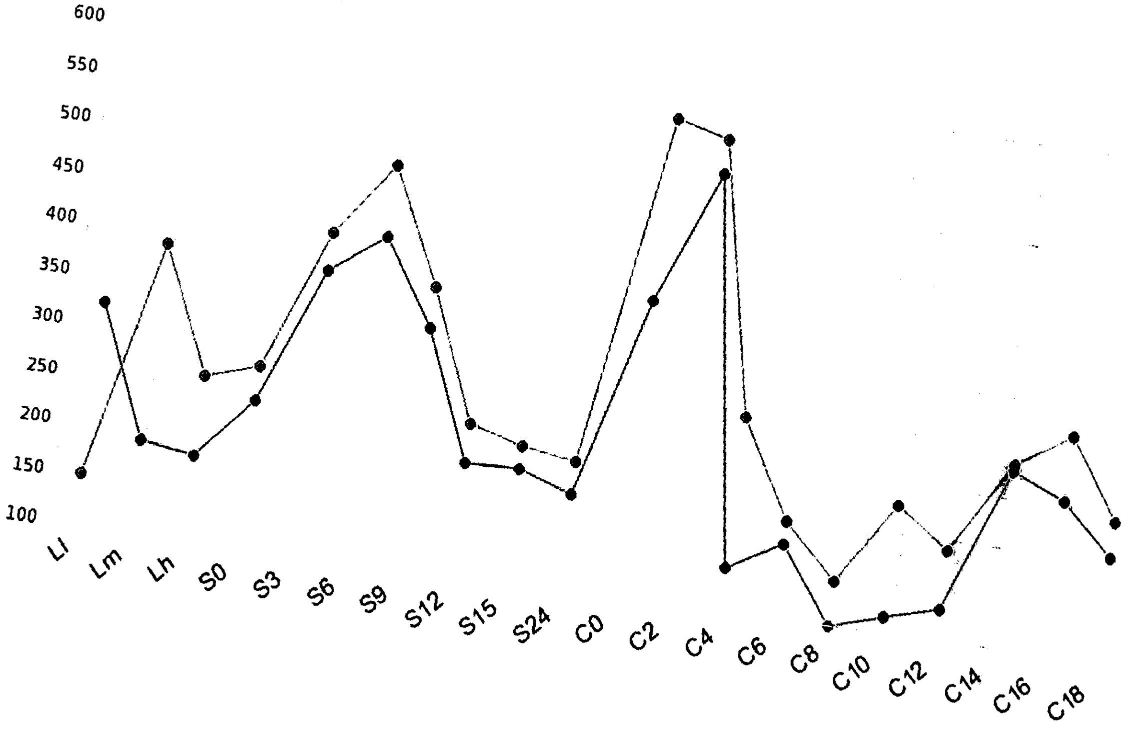
TTHERM_00219410



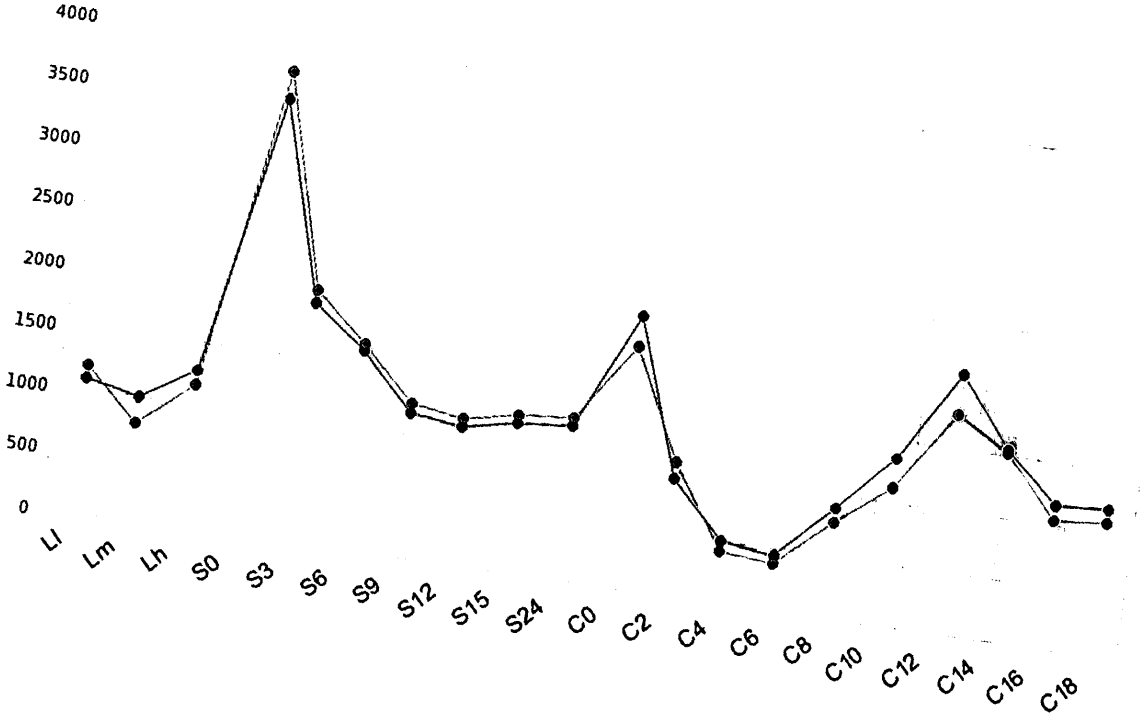
TTHERM_00267860



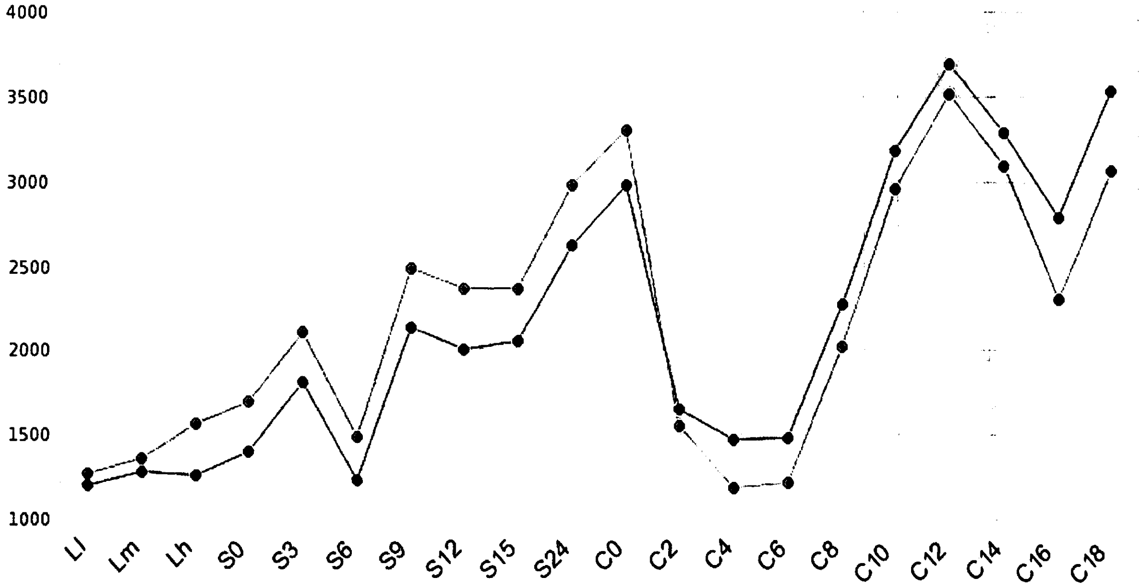
TTHERM_00316060



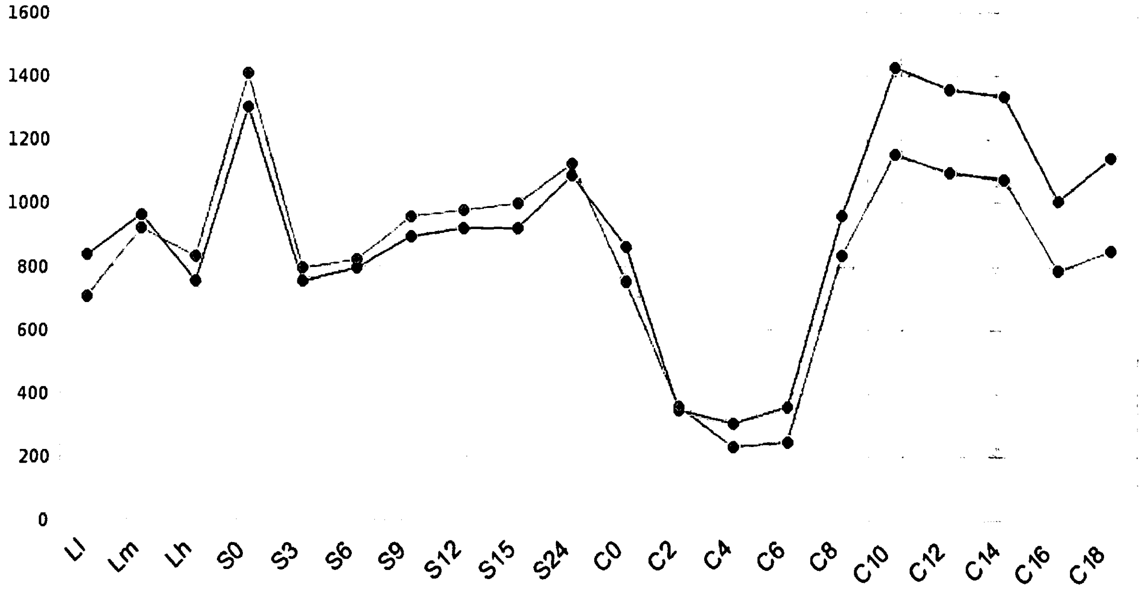
TTHERM_00339690



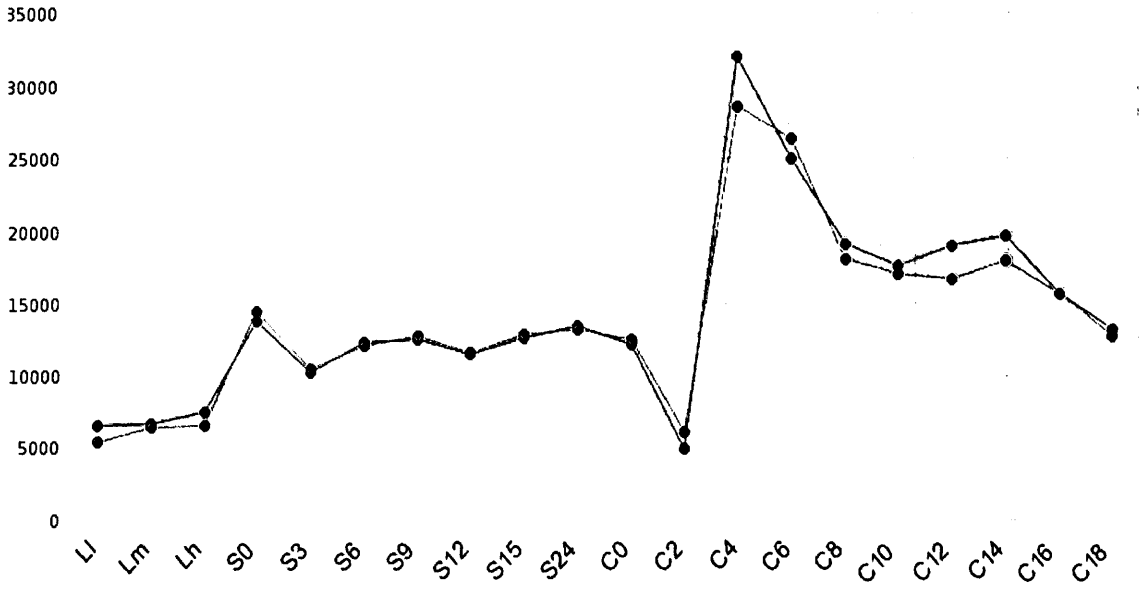
TTHERM_00348940



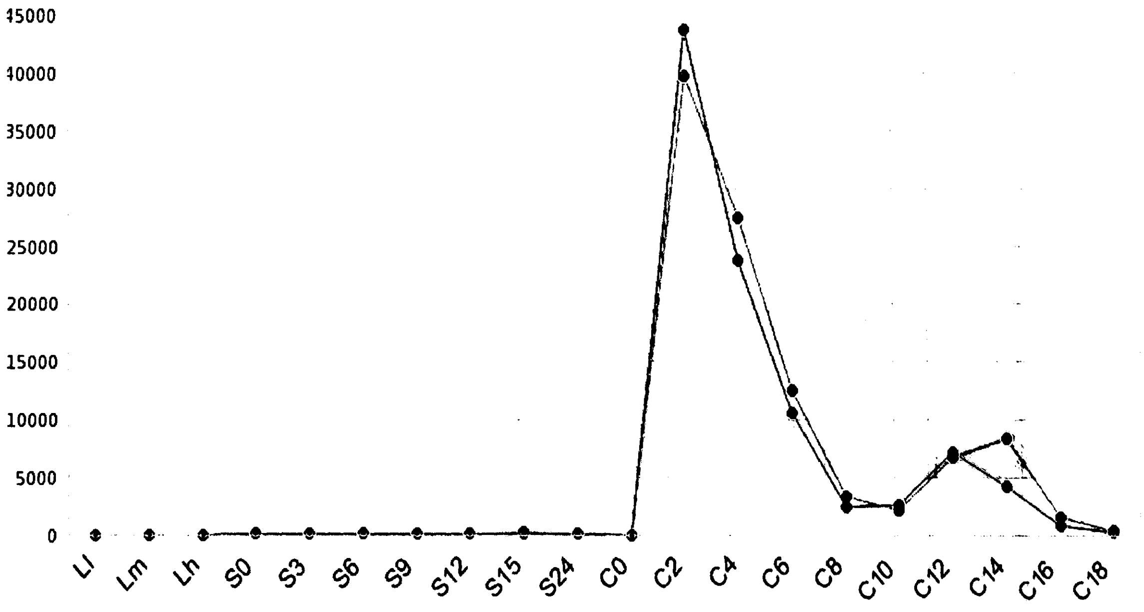
TTHERM_00355780



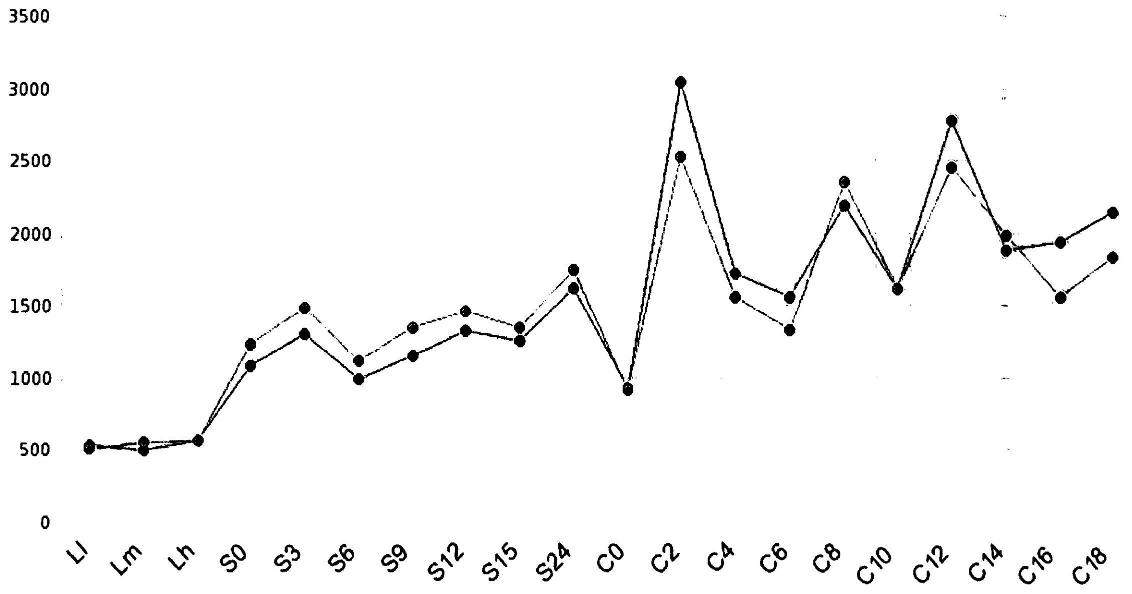
TTHERM_00469230



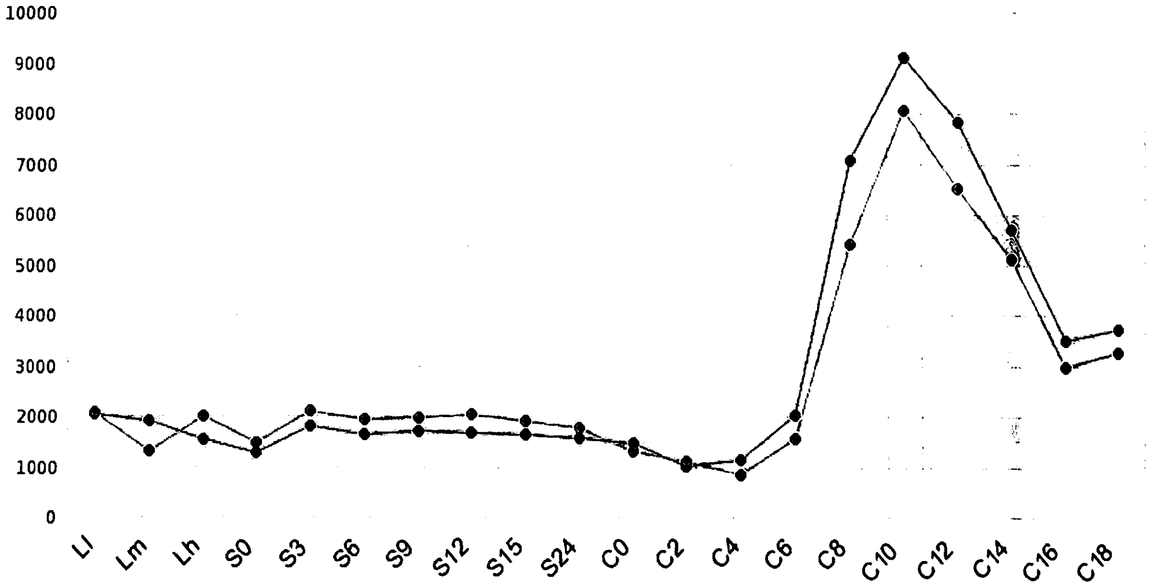
TTHERM_00483640



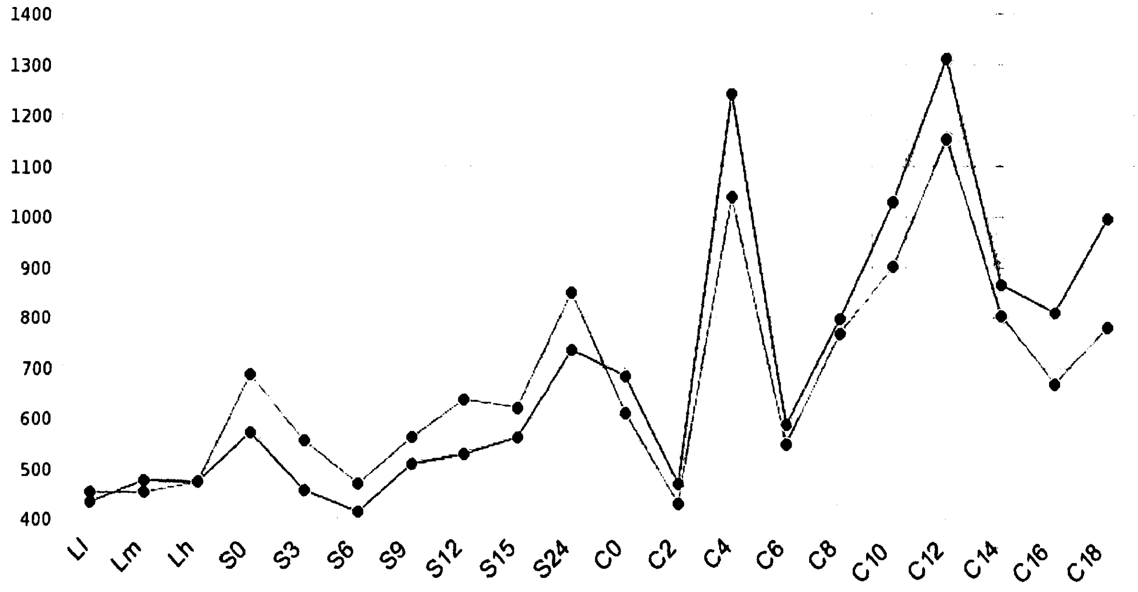
TTHERM_00529780



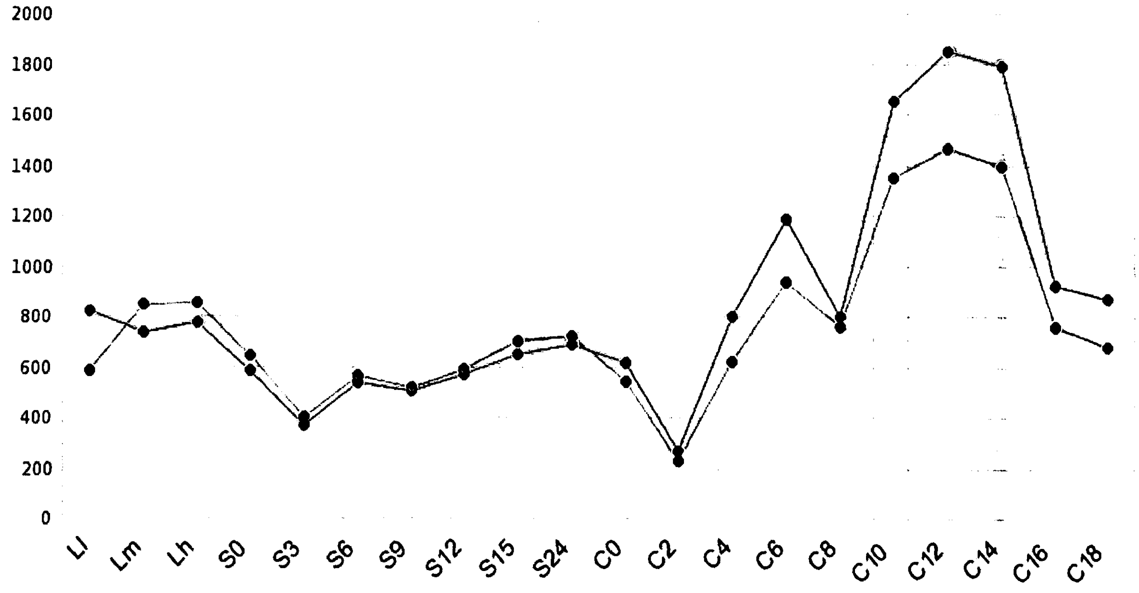
TTHERM_00532680



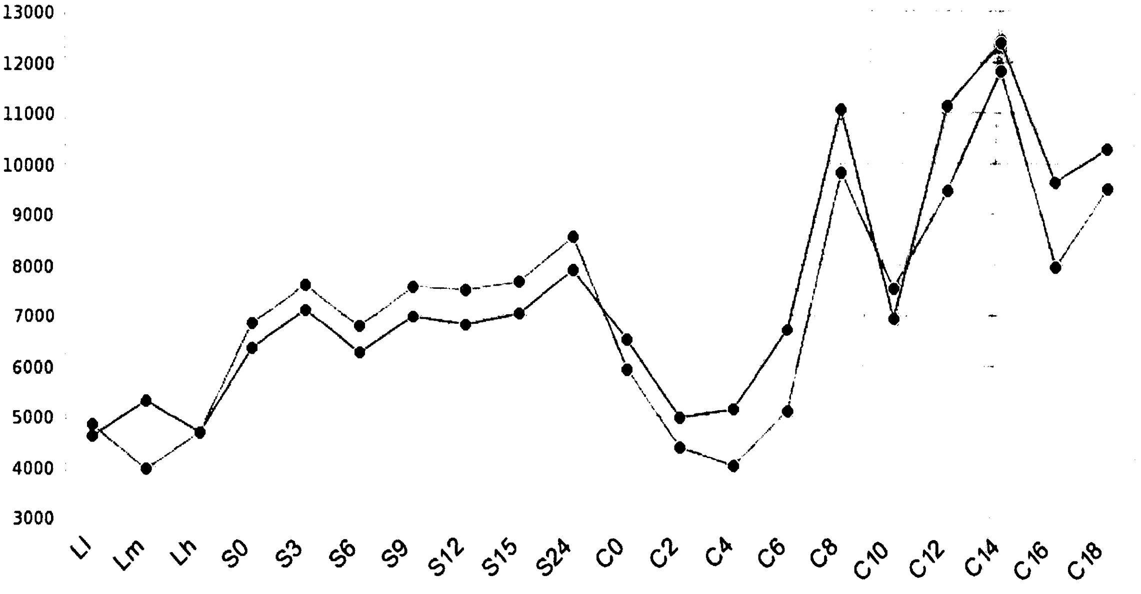
TTHERM_00537240



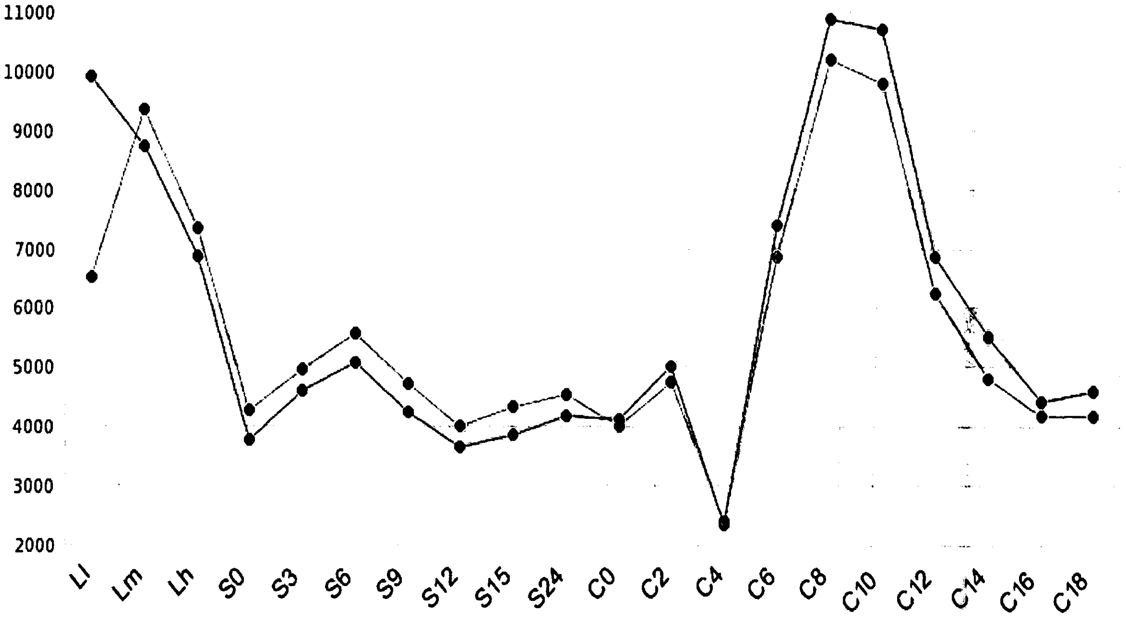
TTHERM_00557780



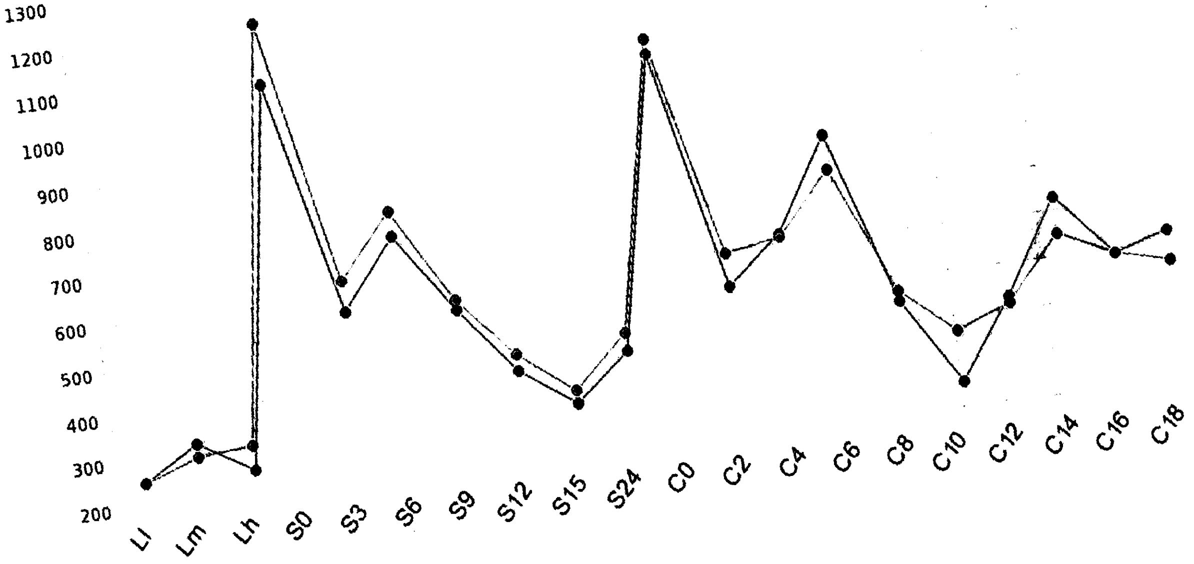
TTHERM_00575520



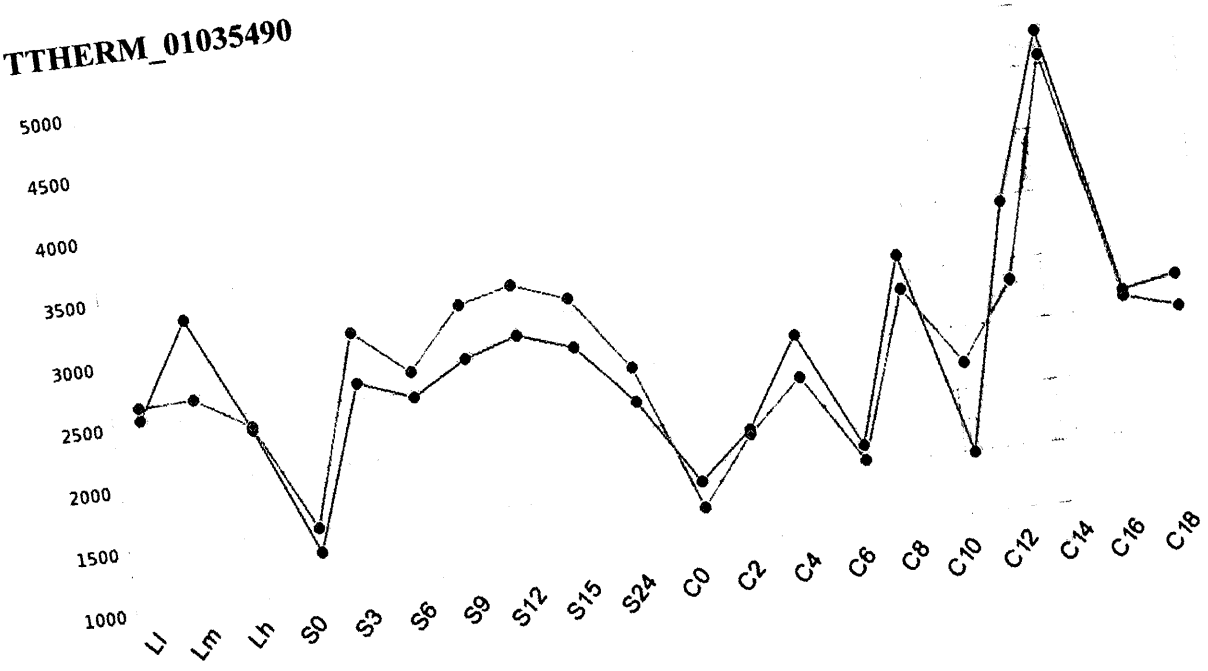
TTHERM_00780700



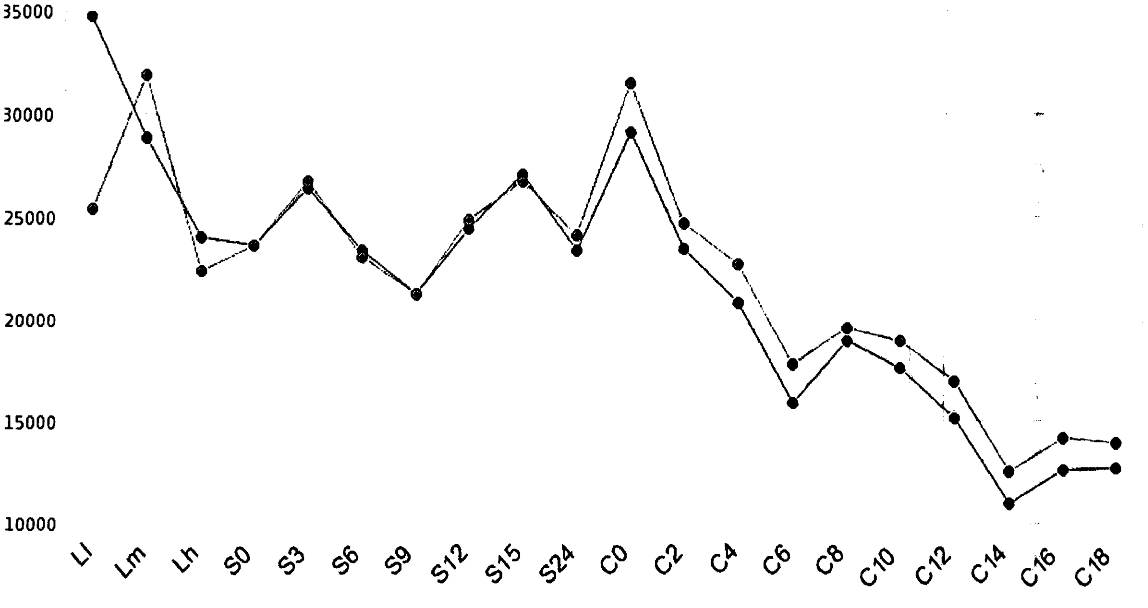
TTHERM_00827160



TTHERM_01035490



TTHERM_01043090



TTHERM_01347900