MITOGEN-ACTIVATED PROTEIN KINASE-ACTIVATED PROTEIN KINASE 5

Structure, function and inhibition

Inger Lindin
A dissertation for the degree of Philosophiae Doctor – June 2014

Inger Lindin has been enrolled in BioStruct, the Norwegian national graduate school in Structural biology and in the PhD school in Molecular and Structural Biology at UiT-The Arctic University of Norway.
Contents

Acknowledgements ................................................................................................................................. 3
List of papers ........................................................................................................................................... 4
Abbreviations .......................................................................................................................................... 5
Introduction ............................................................................................................................................. 9
MAPK signaling pathways .................................................................................................................... 9
Conventional MAPKs .......................................................................................................................... 11
  ERK1/2 ........................................................................................................................................... 11
  ERK5 ............................................................................................................................................... 12
  p38 (α, β, γ, δ) ............................................................................................................................... 13
  JNK 1/2/3 ....................................................................................................................................... 14
Atypical MAPKs .................................................................................................................................. 16
  ERK3/4 ........................................................................................................................................... 16
  ERK7/8 ........................................................................................................................................... 17
  NLK ................................................................................................................................................. 17
MAPKAPKs ......................................................................................................................................... 19
  RSK ................................................................................................................................................. 20
  MSK ................................................................................................................................................ 22
  MNK ............................................................................................................................................... 23
  MK2 and MK3 ................................................................................................................................ 25
  MK5 ............................................................................................................................................... 26
General three-dimensional structure of protein kinases .................................................................. 29
Inhibitors ........................................................................................................................................... 31
Methodical considerations .................................................................................................................... 33
  Homology modeling .......................................................................................................................... 34
  Docking .............................................................................................................................................. 35
  ROC curves ......................................................................................................................................... 37
  Virtual screening ................................................................................................................................ 37
  Molecular dynamics .......................................................................................................................... 38
  Kinase assay ....................................................................................................................................... 40
Aims of study ......................................................................................................................................... 42
Summary of papers ............................................................................................................................... 43
Acknowledgements

This work has been carried out at UiT-The Arctic University of Norway from 2010 to 2014 under the supervision of Professor Ingebrigt Sylte (Medical Pharmacology and Toxicology, Department of Medical Biology, Faculty of Health Sciences) and Professor Ugo Moens (Research Group for Molecular Inflammation, Department of Medical Biology, Faculty of Health Sciences).

I would like to start by thanking my supervisors Professor Ingebrigt Sylte and Professor Ugo Moens for introducing me to the field of computational chemistry and protein kinases. It has been an interesting journey filled with encouraging discussions regarding both my work and the thesis preparation.

I also would like to express my sincere gratitude to my co-supervisor Associate Professor Aina Westrheim Ravna for always being open for my questions and discussions, and for the continual help and support through my PhD. The hot cocoa, served in the freezing graphics room, was really nice!

I wish to thank my co-supervisor Dr. Sergiy Kostenko for introducing me to kinase assays and Dr. Ymingjiang Wuxiuier for guiding me through the jungle that is molecular dynamics.

Also, I would like to express my gratitude for the collaboration with Dr. Irina Kufareva and Professor Ruben Abagyan during the work on homology models and the Stallo support team during MD simulations on the super computer.

I also want to thank my coworkers in the coffee-room at Plan 8, with whom I have been able to share my daily mood with and who have been both warm and helpful in issues of both professional and private character.

Finally, I would like to thank my husband Stian for his everlasting love, support, encouragement and understanding, and my children, Maja and Mathias, for showing me that there are much more important things in life than work.

Tromsø, June 2014

Inger Lindin
List of papers

Paper I


Paper II


Paper III

### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASK</td>
<td>Apoptosis signal-regulating kinase</td>
</tr>
<tr>
<td>Abl</td>
<td>Abelson murine leukemia viral oncogene homolog 1</td>
</tr>
<tr>
<td>ADP</td>
<td>Adenine diphosphate</td>
</tr>
<tr>
<td>AGC family</td>
<td>Protein kinase A, G and C family</td>
</tr>
<tr>
<td>ATF</td>
<td>Activating transcription factor</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenine triphosphate</td>
</tr>
<tr>
<td>AUC</td>
<td>Area under curve</td>
</tr>
<tr>
<td>Bad</td>
<td>Bcl-2-associated death promoter</td>
</tr>
<tr>
<td>Bax</td>
<td>Bcl-2-associated X protein</td>
</tr>
<tr>
<td>BDNF</td>
<td>Brain-derived neurotrophic factor</td>
</tr>
<tr>
<td>BMK</td>
<td>Big MAPK</td>
</tr>
<tr>
<td>CaMK</td>
<td>Calcium- and calmodulin-dependent kinase</td>
</tr>
<tr>
<td>cAMP</td>
<td>Cyclic adenosine monophosphate</td>
</tr>
<tr>
<td>CD-domain</td>
<td>Common docking domain</td>
</tr>
<tr>
<td>cKIT</td>
<td>Proto-oncogene c-KIT, transmembrane receptor tyrosine kinase</td>
</tr>
<tr>
<td>cPLA2</td>
<td>Cytosolic phospholipases A2</td>
</tr>
<tr>
<td>CRM1</td>
<td>Chromosome region maintenance 1, exportin 1</td>
</tr>
<tr>
<td>Cx43</td>
<td>Connexin 43</td>
</tr>
<tr>
<td>CTKD</td>
<td>C-terminal kinase domain</td>
</tr>
<tr>
<td>DAPK</td>
<td>Death-associated protein kinase</td>
</tr>
<tr>
<td>DFG motif</td>
<td>Asp-Phe-Glu motif</td>
</tr>
<tr>
<td>DLK</td>
<td>Dual-leucine-zipper-bearing kinase</td>
</tr>
<tr>
<td>D-domain</td>
<td>Docking motif</td>
</tr>
<tr>
<td>eIF4G</td>
<td>Eukaryotic translation initiation factor 4 gamma</td>
</tr>
<tr>
<td>eIF4H</td>
<td>Eukaryotic translation initiation factor 4H</td>
</tr>
<tr>
<td>EGCG</td>
<td>Epigallocatechin gallate</td>
</tr>
<tr>
<td>EGF</td>
<td>Epidermal growth factor</td>
</tr>
<tr>
<td>Elk-1</td>
<td>ETS domain-containing protein Elk-1</td>
</tr>
<tr>
<td>EPS</td>
<td>Electrostatic potential surfaces</td>
</tr>
<tr>
<td>ERK</td>
<td>Extracellular signal-regulated kinase</td>
</tr>
<tr>
<td>ETS</td>
<td>E26 transformation-specific</td>
</tr>
<tr>
<td>FGF-2</td>
<td>Fibroblast growth factor</td>
</tr>
<tr>
<td>GPCR</td>
<td>G protein-coupled receptor</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>G-CSF</td>
<td>Granulocyte-colony stimulating factor</td>
</tr>
<tr>
<td>HIPK</td>
<td>Homeodomain-interacting protein kinase</td>
</tr>
<tr>
<td>hnRNP</td>
<td>Heterogenous nuclear RNA-binding protein A1</td>
</tr>
<tr>
<td>HSF1</td>
<td>Heat shock transcription factor 1</td>
</tr>
<tr>
<td>HSP</td>
<td>Heat shock protein</td>
</tr>
<tr>
<td>HTS</td>
<td>High throughput screening</td>
</tr>
<tr>
<td>h-Tid1</td>
<td>Human DnaJ protein</td>
</tr>
<tr>
<td>HuR</td>
<td>Human antigen R</td>
</tr>
<tr>
<td>IL-6</td>
<td>interleukin 6</td>
</tr>
<tr>
<td>JNK</td>
<td>c-Jun N-terminal kinase</td>
</tr>
<tr>
<td>kDa</td>
<td>kilodalton</td>
</tr>
<tr>
<td>LPS</td>
<td>Lipopolysaccharides</td>
</tr>
<tr>
<td>MAPK</td>
<td>Mitogen-activated protein kinase</td>
</tr>
<tr>
<td>MAPKK</td>
<td>MAPK kinase</td>
</tr>
<tr>
<td>MAPKKK</td>
<td>MAPK kinase kinase</td>
</tr>
<tr>
<td>MAPKAPK</td>
<td>MAPK-activated protein kinase</td>
</tr>
<tr>
<td>MD</td>
<td>Molecular dynamics</td>
</tr>
<tr>
<td>MEF2</td>
<td>myocyte enhancer factor-2</td>
</tr>
<tr>
<td>MEK</td>
<td>MAPK/ERK kinase</td>
</tr>
<tr>
<td>MEKK</td>
<td>MEK kinase</td>
</tr>
<tr>
<td>MK</td>
<td>MAPK-activated protein kinase</td>
</tr>
<tr>
<td>MKK</td>
<td>Mitogen-activated protein kinase kinase</td>
</tr>
<tr>
<td>MLK</td>
<td>Mixed lineage kinase</td>
</tr>
<tr>
<td>MNK</td>
<td>MAPK-interacting kinase</td>
</tr>
<tr>
<td>MSK</td>
<td>Mitogen- and stress-activated kinase</td>
</tr>
<tr>
<td>NES</td>
<td>Nuclear export signal</td>
</tr>
<tr>
<td>NF-AT</td>
<td>Nuclear factor of activated T-cells</td>
</tr>
<tr>
<td>NGF</td>
<td>nerve growth factor</td>
</tr>
<tr>
<td>NLK</td>
<td>Nemo-like kinase</td>
</tr>
<tr>
<td>NLS</td>
<td>Nuclear localization signal</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>NTKD</td>
<td>N-terminal kinase domain</td>
</tr>
<tr>
<td>p-38</td>
<td>MAPK 11-14</td>
</tr>
<tr>
<td>PAK</td>
<td>p21-activated kinase</td>
</tr>
<tr>
<td>PBR</td>
<td>Polybasic region</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein Data Bank</td>
</tr>
<tr>
<td>PDGF</td>
<td>Platelet-derived growth factor</td>
</tr>
<tr>
<td>PDGFR</td>
<td>Platelet-derived growth factor receptor</td>
</tr>
<tr>
<td>PDK1</td>
<td>Phosphoinositide-dependent protein kinase 1</td>
</tr>
<tr>
<td>PKA</td>
<td>Protein kinase A</td>
</tr>
<tr>
<td>PKI</td>
<td>Protein kinase inhibitor peptide</td>
</tr>
<tr>
<td>P-loop</td>
<td>Phosphate binding loop</td>
</tr>
<tr>
<td>PR-domain</td>
<td>Proline-rich domain</td>
</tr>
<tr>
<td>PRAK</td>
<td>p38-regulated/activated protein kinase</td>
</tr>
<tr>
<td>QSAR</td>
<td>Quantative structure-activity relationship</td>
</tr>
<tr>
<td>Raf</td>
<td>Rapidly accelerated fibrosarcoma</td>
</tr>
<tr>
<td>RET/PTC3</td>
<td>Activated form of the RET proto-oncogene</td>
</tr>
<tr>
<td>RI</td>
<td>Regulatory subunit</td>
</tr>
<tr>
<td>RMSD</td>
<td>Root Mean Square Deviations</td>
</tr>
<tr>
<td>RMSF</td>
<td>Root Mean Square Fluctuations</td>
</tr>
<tr>
<td>ROC</td>
<td>Receiver Operating Characteristic</td>
</tr>
<tr>
<td>RSK</td>
<td>Ribosomal S6 kinase</td>
</tr>
<tr>
<td>Sap1a</td>
<td>ETS domain transcription factor</td>
</tr>
<tr>
<td>SAPK</td>
<td>Stress-activated protein kinase</td>
</tr>
<tr>
<td>SFTPB</td>
<td>Surfactant protein B</td>
</tr>
<tr>
<td>SGK</td>
<td>serum- and glucocorticoid-induced protein kinase</td>
</tr>
<tr>
<td>SH</td>
<td>src homology</td>
</tr>
<tr>
<td>STAT3</td>
<td>Signal transducer and activator of transcription 3</td>
</tr>
<tr>
<td>SPC</td>
<td>Simple point charge</td>
</tr>
<tr>
<td>TAD</td>
<td>Transactivation domain</td>
</tr>
<tr>
<td>TAO</td>
<td>Thousand-and-one amino acid</td>
</tr>
<tr>
<td>TAK</td>
<td>Transforming growth factor β-activated kinase</td>
</tr>
<tr>
<td>TCF/LEF</td>
<td>T-cell factor / lymphoid enhancer factor</td>
</tr>
<tr>
<td>TGF-β</td>
<td>Transforming growth factor β</td>
</tr>
<tr>
<td>TNF-α</td>
<td>Tumor necrosis factor α</td>
</tr>
<tr>
<td>Tpl</td>
<td>Tumor progression loci 2</td>
</tr>
<tr>
<td>TSC2</td>
<td>tuberous sclerosis complex 1</td>
</tr>
<tr>
<td>VEGF</td>
<td>vascular endothelial growth factor</td>
</tr>
<tr>
<td>VLS/VS</td>
<td>Virtual ligand screening/virtual screening</td>
</tr>
</tbody>
</table>
Introduction

Signal transduction is one of the fundamental processes of living cells and can be considered a coordinated relay of messages derived from extracellular cues to intracellular effectors (Scott and Pawson 2009).

Living cells are exposed to many different physical and chemical stimuli from their environment. Chemical stimuli can be alteration of nutrients, growth factors and cytokines as well as drugs and neurotransmitters, while physical stimuli can be sudden changes in osmolarity, heat, pH, radiation and mechanical stress. All these signals control many aspects of cell function, and different cells require different sets of signals in order to survive. Cells receive signals by binding signal molecules to cellular receptor proteins. This often initiates a phosphorylation cascade, which ultimately results in some sort of cellular response. This phosphorylation cascade is mediated by enzymes called protein kinases. Humans have 518 different protein kinases, often referred to as the human kinome (Manning, Whyte et al. 2002). Of these, 478 belong to a single superfamily whose catalytic domains are related in sequence. These kinases can be clustered into groups, families and sub-families of increasing sequence similarity.

One group of phosphorylation cascades is the mitogen-activated protein kinase (MAPK) signaling pathways. The focus of this thesis is on the MAPK signaling pathways and particularly on one of its members, the Mitogen-activated protein kinase-activated protein kinase 5 (MAPKAPK5).

MAPK signaling pathways

Mitogen-activated protein kinases (MAPKs) constitute major signaling pathways in cells, and are involved in processes controlling gene expression, cell division, cell survival, apoptosis, metabolism, differentiation and motility. MAPK pathways are divided into the conventional and atypical signaling pathways (see figure 1). The conventional mammalian pathways consist of a cascade of three serine/threonine kinases referred to as MAPK kinase kinase, MAPK kinase and MAPK. The MAPKs are divided into four different subfamilies: the extracellular signal-regulated kinases 1/2 (ERK1/2), the c-JUN N-terminal kinases 1-3 (JNK1-3) also called stress-activated protein kinases (SAPK$\alpha$, $\beta$ and $\gamma$), the p38 MAPKs (p38 $\alpha$, $\beta$, $\gamma$ and $\delta$), and the big MAPKs (BMK1/ERK5). The atypical MAPK pathways are
not organized in the normal three-tiered kinase cascade, and includes ERK3/4, ERK7/8 and Nemo-like kinase (NLK) (Cargnello and Roux 2011). Both the conventional and the atypical pathways can phosphorylate non protein kinase substrates, as well as other protein kinases called mitogen-activated protein kinase-activated protein kinases (MAPKAPK). The MAPKAPK family comprises 11 Ser/Thr kinases, which according to sequence similarities can be subdivided into four groups: ribosomal S6 kinases (RSKs), mitogen and stress activated kinases (MSKs), MAPK-interacting kinases (MNKs) and MKs. The last includes MK2, MK3, and MK5 (Gaestel 2006). The MAPK signaling cascades of both conventional and atypical MAPK leading to phosphorylation and activation of MAPKAPKs are depicted in figure 1.

![MAPK signaling cascade diagram](image)

**Figure 1: Mammalian MAPK signaling cascades leading to the activation of MAPKAPK.**

Different cellular stimuli, such as stress, mitogens, cytokines and UV-irradiation, promote the activation of various MAPK pathways. These will in turn phosphorylate and activate the MAPKAPKs. Solid lines indicate validated activation, while dotted lines indicate that activation has been reported but still needs validation. The figure is modified from (Cargnello and Roux 2011).
The following sections will give a description of the protein kinases involved both in the conventional and the atypical MAPK signaling pathways.

**Conventional MAPKs**

**ERK1/2**

Extracellular signal-regulated kinases 1 and 2 (ERK1/2) were originally found to be phosphorylated on Tyr and Thr in response to growth factors (Cooper, Bowen-Pope et al. 1982, Kazlauskas and Cooper 1988, Ray and Sturgill 1988). Human ERK1 consist of 379 amino acids while ERK2 consists of 360 amino acids, and they are approximately 83% amino acid identical. These were the first MAPKs to be cloned (Boulton, Yancopoulos et al. 1990, Boulton, Nye et al. 1991).

ERK1/2 are co-expressed in all tissues examined, with highest levels in brain, skeletal muscle, thymus and heart (Boulton, Yancopoulos et al. 1990). In quiescent cells, ERK1/2 have a cytoplasmic localization. Upon extracellular stimulation a significant proportion of ERK1/2 will, however, translocate to the nucleus (Chen, Sarnecki et al. 1992). They are activated by various stimuli, including growth factors, insulin, heterodimeric G protein-coupled receptors, cytokines, osmotic stress and microtubule disorganization (Boulton, Yancopoulos et al. 1990, Raman, Chen et al. 2007). ERK1/2 activation initiates a phosphorylation cascade where Raf kinases or c-Mos (MAPKKK) binds to and phosphorylate the dual-specificity kinases MEK1/2 (MAPKK), which in turn phosphorylate ERK1/2 within the conserved Thr-Glu-Tyr (TEY) motif in the activation loop. This allows ERK1/2 to phosphorylate its substrates (Cargnello and Roux 2011). Their substrates include nuclear transcription factors (NF-AT, MEF2, STAT3, Elk-1, c-Fos, c-Myc), cytoskeletal proteins (neurofilaments and paxillin), cytoplasmic proteins (death-associated protein kinase (DAPK), tuberous sclerosis complex 1 (TSC2)), and several MAPKAPKs (RSK1/2/3, MSK1/2, MNK1/2) (Yoon and Seger 2006, Cargnello and Roux 2011, Roskoski 2012). Through these substrates ERK1/2 are involved in a large variety of cellular processes like cell adhesion, cell cycle progression, cell migration, cell survival, differentiation, metabolism, proliferation, and transcription (Roskoski 2012).

Even though ERK1 and ERK2 are quite similar in both amino acid sequence (see overall domain organization in figure 2), expression pattern and biological function the respective knockout phenotypes are markedly different, with ERK2 deficient mice dying early in development, suggesting that ERK1 cannot compensate for ERK2 activity (Yao, Li et al. 2003), while ERK1 deficient mice are...
viable, and have only minor defects such as a deficit in thymocyte maturation (Pages, Guerin et al. 1999).

**ERK5**

ERK5, also known as big MAP kinase 1 (BMK1), is twice as big as other MAPKs and was discovered by two different groups in 1995 (Lee, Ulevitch et al. 1995, Zhou, Bao et al. 1995). Zhou *et al.* identified the upstream kinase of MEK1 and used a two-hybrid screen to discover ERK5 as a binding partner, while Lee *et al.* screened a human placenta cDNA library using degenerated PCR to identify a novel MAPK gene, whose gene product was termed big MAPK1 (BMK1) due to its size compared to ERK1/2.

ERK5 consists of 816 amino acids, with a primary structure quite distinct from other MAPK members (see figure 2). The N-terminal part starts with a region important for cytoplasmic targeting followed by a kinase domain. The kinase domain is 66% identical to the ERK2 kinase domain and contains the TEY activation motif (Zhou, Bao et al. 1995). The kinase domain can be separated into a region important for MEK5 interaction and a region important for oligomerization (Yan, Luo et al. 2001). The kinase domain also contains a common docking (CD) domain which allows docking with certain docking (D)-domain-containing substrates (Tanoue and Nishida 2002). The kinase domain is followed by a rather unique extended C-terminal tail. Within this C-terminal tail a nuclear localization signal (NLS) domain, important for nuclear targeting, is localized (Yan, Luo et al. 2001, Buschbeck and Ullrich 2005). Further, two proline-rich (PR) domains termed PR1 and PR2 are localized in this region, and are believed to be potential binding sites for Src-homology 3 (SH3)-domain containing proteins (Zhou, Bao et al. 1995, Yan, Luo et al. 2001). A myocyte enhancer factor 2 (MEF2)-interacting region is also localized in the C-terminal tail (Yan, Luo et al. 2001). Recently a potent transcriptional activation domain was identified, which through auto-phosphorylation enables it to regulate gene transcription (Kasler, Victoria et al. 2000). Finally it has been shown that truncation of the C-terminal tail gives rise to increased kinase activity, indicating an auto-inhibitory function of the tail (Buschbeck and Ullrich 2005). ERK5 lack a normal nuclear export (NES) domain and has a constant nuclear localizing activity (Yan, Luo et al. 2001, Kondoh, Terasawa et al. 2006). However, it has also been proposed to possess nuclear export activity (Raviv, Kalie et al. 2004, Buschbeck and Ullrich 2005). It is anticipated that nonphosphorylated ERK5 exists in a folded state where the C-terminal and N-terminal ends interact, thereby either masking the effect of NLS signal or creating a NES signal (Kondoh, Terasawa et al. 2006).
ERK5 is expressed to various extents in all tissue examined, with particular high levels in brain, thymus and spleen (Yan, Carr et al. 2003). It was initially shown to be activated by stress stimuli (oxidative stress and hyperosmolality) but not platelet-derived growth factor (PDGF) (Abe, Kusuhara et al. 1996). Subsequently, it was proven to be activated by a plethora of extracellular stimuli, such as: vascular endothelial growth factor (VEGF), epidermal growth factor (EGF), fibroblast growth factor (FGF-2) and PDGF (Kato, Tapping et al. 1998, Hayashi and Lee 2004, Kesavan, Lobel-Rice et al. 2004). ERK5 may also be activated by trophic factors in neurons (brain-derived neurotrophic factor, BDNF) (Cavanaugh, Ham et al. 2001), nerve growth factor (NGF) (Kamakura, Moriguchi et al. 1999, Watson, Heerssen et al. 2001, Finegan, Wang et al. 2009) and inflammatory cytokines like interleukin 6 (IL-6) (Carvajal-Vergara, Tabera et al. 2005).

Upon stimulation, MEKK2/3 and Tlp-2 may stimulate MEK5, which will activate and phosphorylate ERK5 on Thr and Tyr residues in the conserved TEY motif in the activation loop (Mody, Campbell et al. 2003). This renders ERK5 capable of further phosphorylating its substrates like the MEF2 family of transcriptional factors, ETS domain transcription factor (Sap1a), c-Myc, serum- and glucocorticoid-induced protein kinase (SGK), connexin 43 (Cx43) and Bad (Hayashi and Lee 2004, Wang, Finegan et al. 2006), and the MAPKAPK RSK (Ranganathan, Pearson et al. 2006).

ERK5 is essential for early embryonic development, and is required for normal development of vascular system as well as cell survival (Regan, Li et al. 2002, Sohn, Sarvis et al. 2002, Yan, Carr et al. 2003). It has also been shown to be implicated in cancer/tumor development and heart function (Wang and Tournier 2006).

\textbf{p38 (α, β, γ, δ)}


The four p38 MAPKs are encoded by different genes and have different tissue expression patterns, with p38α being ubiquitously expressed at significant levels in most cell types, whereas the others seem to be expressed in a more tissue-specific manner; for example p38β in brain, p38γ in skeletal muscle and p38δ in endocrine glands (Cuadrado and Nebreda 2010).
In mammalian cells the p38s are strongly activated by various environmental stress factors and inflammatory cytokines, including oxidative stress, UV irradiation, hypoxia, ischemia, interleukin-1 (IL-1) and tumor necrosis factor alpha (TNF-α) (Cuadrado and Nebreda 2010). The p38 module plays a critical role in normal immune and inflammatory responses (Cuadrado and Nebreda 2010), and is also important in cell proliferation and survival (Thornton and Rincon 2009).

Upon stimuli the MAPKKks MEKK1-3, MLK2/3, ASK1, Tpl2, TAK1 and TAO1/2 activates MKK3/6. MKK3/6 will then activate and phosphorylate the different p38 isoforms in the Thr-Gly-Tyr (TGY) motif of the activation loop (Cuadrado and Nebreda 2010). Once the p38 isoforms are activated they are free to phosphorylate their various substrates. The substrates include cPLA2, MNK1/2, MK2/3, HuR, Bax and tau in the cytoplasm, and ATF1/2/6, MEF2, Elk-1, GADD153, Ets1, p53 and MSK1/2 in the nucleus (Cuadrado and Nebreda 2010). The role of MK5 as a p38 substrate is still under debate (Shiryaev and Moens 2010).

JNK 1/2/3

The first c-Jun N-terminal kinase (JNK) was identified as a cycloheximide-activated MAP-2 kinase (Kyriakis and Avruch 1990, Hibi, Lin et al. 1993). Subsequently it was found that stress stimuli promote JNK phosphorylation of Thr and Tyr residues (Kyriakis, Brautigan et al. 1991, Derijard, Hibi et al. 1994).

Three isoforms of JNK have been identified; JNK1-3 (also termed stress-activated protein kinase (SAPK)-γ, SAPK-α and SAPK-β, respectively). The JNKs are more than 85% identical in sequence and are encoded by three different genes, which will give rise to at least 10 different spliced variants (Derijard, Hibi et al. 1994, Kyriakis, Banerjee et al. 1994, Gupta, Barrett et al. 1996). For overall domain organization see figure 2.

JNK3 is mainly found in neuronal tissue, testis and cardinal myocytes, while JNK 1 and 2 are ubiquitously expressed (Bode and Dong 2007).

JNKs are strongly activated in response to various cellular stresses like heat shock, ionizing radiation, oxidative stress, DNA-damaging agents, cytokines, UV irradiation, DNA and protein synthesis inhibitors and growth factor deprivation. To a lesser extent JNKs are also activated by growth factors, GPCR ligands and serum (Bode and Dong 2007). Upon stimuli several MAPKKks (MEKK1-4, MLK1-3, Tpl-2, DLK, TAO1/2, TAK1 and ASK1/2) become activated and may phosphorylate and activate mitogen-activated protein kinase kinase (M KK) 4 and/or 7 by dual phosphorylation. The activated
MKK4/7 will then directly phosphorylate and activate JNKs. Complete activation of JNKs requires dual phosphorylation of threonine and tyrosine residues within a threonine/proline/tyrosine motif located in kinase domain VIII. MKK4 preferentially phosphorylates tyrosine 185, whereas MKK7 prefers threonine 183 (Cagnello and Roux 2011).

Activated JNKs are shown to relocalize from the cytoplasm to the nucleus (Mizukami, Yoshioka et al. 1997), where they interact with various transcriptional factors like c-Jun, p53, ATF-2, NF-ATc1, Elk-1, HSF-1, STAT3, c-Myc, and JunB (Weston and Davis 2002, Raman, Chen et al. 2007, Bogoyevitch, Ngoei et al. 2010). Through these substrates JNKs contribute to multiple physiological processes, including cell proliferation (Sabapathy, Hochdinger et al. 2004), survival (Hess, Pihan et al. 2002) and apoptosis (Dhanasekaran and Reddy 2008). Studies with primary murine embryonic fibroblast, isolated from JNK1/JNK2 knockout mice, show that JNK is required for UV-induced apoptosis. Absence of JNK causes a defect in the mitochondrial death signal pathway, including the failure to release cytochrome c. This demonstrates JNK’s role in the intrinsic apoptotic pathway (Tournier, Hess et al. 2000).

Figure 2: Overall domain organization of conventional MAPKs.
All conventional MAPK contain a Thr kinase domain flanked by a C- and N-terminal domain of various lengths. TEY, TGY and TPY are the phosphorylation motifs within the kinase domain. In addition ERK5 contains a nuclear localization signal (NLS) and a transactivation domain (TAD). Figure modified from (Cagnello and Roux 2011).
Atypical MAPKs

ERK3/4

ERK3 was first cloned in 1991 by the group of Boulton by screening a rat cDNA library using a probe derived from ERK1 (Boulton, Nye et al. 1991). In 1992 ERK4 was isolated using a similar method by the group of Gonzales (Gonzalez, Raden et al. 1992). These two kinases have a very similar protein structure with 76% amino acid identity in the kinase domain. ERK3 and 4 are considered to be atypical MAPKs because their activation loop does not contain the classical Thr-Xxx-Tyr motif (see figure 3). Instead, they contain a single phospho-acceptor site within the Ser-Glu-Gly motif. Secondly ERK3/4 contains the sequence Ser-Pro-Arg instead of the highly conserved Ala-Pro-Glu (APE motif) in subdomain VIII of the kinase domain. ERK3/4 are the only MAPK with an Arg residue in this position. In addition, ERK3/4 contain a unique C-terminal extension important for subcellular targeting (Julien, Coulombe et al. 2003), and ERK3 has an N-terminal part which in involved in degradation (Coulombe, Rodier et al. 2004). For overall domain organization see figure 3.

ERK3 mRNA has the highest expression in the skeletal muscle, followed by the brain. It is also found in heart, placenta, lung, liver, pancreas, kidney, and skin fibroblasts, while ERK4 is mainly found in heart and brain tissue (UniProt).

To date, little is known about the substances or kinases that activate ERK3 or ERK4, but recent studies have shown that a group of p21-activated kinases (PAK1-3) can phosphorylate ERK3/4 both in vivo and in vitro (De la Mota-Peynado, Chernoff et al. 2011, Deleris, Trost et al. 2011). This phosphorylation results in activation of ERK3/4 and the subsequent activation of MK5, their only known substrate.

Little is known about the biological roles of ERK4, as MK5 is the only known substrate and erk4 -/- mice display no obvious phenotype (Rousseau, Klinger et al. 2010). ERK3 has been shown to be involved in a number of biological functions, including cell proliferation, cell cycle progression and cell differentiation (Boulton, Nye et al. 1991, Coulombe, Rodier et al. 2003, Julien, Coulombe et al. 2003, Klinger, Turgeon et al. 2009, Tanguay, Rodier et al. 2010). A study by Guaman et al. on ERK3 deficient mice also showed ERK3 to be vital for terminal differentiation of type II cells, SFTPB production, and fetal pulmonary maturity (Cuevas Guaman, Sbrana et al. 2014).
**ERK7/8**

ERK7 was first cloned in 1999 by the group of Rosner (Abe, Kuo et al. 1999) by PCR amplification of rat brain cDNA. More recently ERK7 was used as a probe in a human cDNA screening where the human analog ERK8 was identified (Abe, Saelzler et al. 2002). ERK7 and ERK8 display 69% amino acid identity, with 82% identity in the kinase domain. They have an N-terminal kinase domain with the classical TEY motif (Thr-Glu-Tyr) in the activation segment, similar to that of most other MAP kinases. In addition, both possess a unique C-terminal extension of 204 residues, which contains an NLS sequence and two proline-rich regions possibly important for SH3 domain ligands. See figure 3 for overall domain organization.

ERK7/8 are expressed in all adult tissues, with predominant expression seen in lungs and kidneys (Abe, Saelzler et al. 2002). Both ERK7 and ERK8 appear to be regulated by auto-phosphorylation of the TEY motif in the activation segment and not by activation by any known MAPKK (Abe, Kuo et al. 1999, Abe, Kahle et al. 2001). ERK8, however, seem to be regulated by certain stimuli of conventional MAPK, including serum and hydrogen peroxide (Abe, Saelzler et al. 2002, Klevernic, Stafford et al. 2006). In addition RET/PTC3, an activated form of the RET proto-oncogene, is able to activate ERK8 (Iavarone, Acunzo et al. 2006). Similar to ERK3, both the expression and activity of ERK7/8 seem to be tightly regulated by the ubiquitin-proteasome pathway (Kuo, Duke et al. 2004).

The physiological substrates and function of ERK7/8 are unknown, but studies suggest they may be involved in cell proliferation (Abe, Kuo et al. 1999), chloride transport (Qian, Okuhara et al. 1999), nuclear receptor signaling (Henrich, Smith et al. 2003, Saelzler, Spackman et al. 2006) and autophagy (Zacharogianni, Kondylis et al. 2011, Colecchia, Strambi et al. 2012). Knockout models of ERK7 and ERK8 are lacking.

**NLK**

Human Nemo-like kinase (NLK) was first identified in 1998 by the group of Erikson (Brott, Pinsky et al. 1998). In the kinase domain, NKL displays a 45% amino acid identity with the kinase domain of ERK2. It is considered to be an atypical MAPK kinase because its N- and C-terminal extensions are not present in other MAPK kinases. The N-terminal extension has a unique sequence highly enriched in alanine, glutamine and histidine residues. The extension itself is not well conserved between NLK orthologs and the function is still unknown. The C-terminal extension is well conserved and may contribute in the interaction with substrate (Ishitani, Ninomiya-Tsuji et al. 1999, Yamada, Ohkawara...
et al. 2003, Yamada, Ohnishi et al. 2006). In addition to the unconventional N- and C-terminal extensions, NLK lacks the tyrosine phosphorylation site in the activation loop and has a Thr-Gln-Glu (TQE) motif instead of the classical TEY motif. See figure 3 for overall domain organization.

Mouse NLK is expressed in most adult tissues with highest level in brain and lymphoid organs (Brott, Pinsky et al. 1998). NLK is activated by stimuli of the Wnt pathway (Kanei-Ishii, Ninomiya-Tsuji et al. 2004), and by cytokines like IL-6, granulocyte colony-stimulating factor (G-CSF) and transforming growth factor β (TGF-β) (Ohkawara, Shirakabe et al. 2004, Kojima, Sasaki et al. 2005). The MAPKK TAK-1 has been reported to activate NLK (Ishitani, Ninomiya-Tsuji et al. 1999, Ohkawara, Shirakabe et al. 2004, Smit, Baas et al. 2004). Homeodomain-interacting protein kinase 2 (HIPK2) is the only MAPKK, which has shown potential to activate NLK. Whether this activation happens by phosphorylation of the activation loop, or through promotion of auto-phosphorylation, is still unclear (Kanei-Ishii, Ninomiya-Tsuji et al. 2004).

So far transcription factors of the T-cell factor/lymphoid enhancer factor (TCF/LEF) family (Ishitani, Ninomiya-Tsuji et al. 1999), as well as STAT3 (Kojima, Sasaki et al. 2005), have been identified as substrates of NLK. Over the past years, NLK has been shown to play crucial roles in the regulation of diverse signaling pathways, including Wnt/β-catenin and Notch signaling pathways, and to be involved in embryonic patterning, nervous system development, and cancer cell proliferation (Ishitani and Ishitani 2013). Knockout studies with mice have illustrated that when expressed in undifferentiated osteoblasts, NLK is a negative regulator of skeletal homeostasis possibly by targeting signals that regulate osteoclastogenesis and bone resorption (Canalis, Kranz et al. 2014).

**Figure 3: Overall domain organization of the atypical MAPK.** All atypical MAPK contain a Thr/Ser kinase domain flanked by a C- and N-terminal domain of various lengths. SEG, TEY and TQE are the phosphorylation motifs within the kinase domain. In addition ERK3/4 contains a conserved C34 region, ERK 7/8 contains a nuclear localization signal (NLS) and NLK contains a sequence rich in Ala, His and Glu (AHQr). The figure is modified from (Cargnello and Roux 2011).
MAPKAPKs

Both the conventional and the atypical MAPK pathways can phosphorylate non protein kinase substrates and other protein kinases referred to as mitogen-activated protein kinase-activated protein kinases (MAPKAPK). MAPKAPKs belong to the Ca\(^{2+}\)/calmodulin family of protein kinases and comprise 11 Ser/Thr kinases. According to sequence similarities, all MAPKAPKs can be subdivided into four groups: ribosomal S6 kinases (RSKs), mitogen and stress activated kinases (MSKs), MAPK-interacting kinases (MNKs) and MKs (Gaestel 2006). This relationship is nicely illustrated by Roux and Blenis (Roux and Blenis 2004) in figure 4.

Figure 4: Phylogenetic tree of Ca\(^{2+}\)/calmodulin family members. Modified figure from (Roux and Blenis 2004). Red ring indicates the protein kinase MK5, which is the main focus of this thesis.
RSK

RSK was first identified in *Xenopus laevis* extract (Erikson and Maller 1985), and since then several orthologs have been identified throughout metazoans. The human RSK family contains four isoforms (RSK1 (Roux 2007), RSK2 and RSK3 (Julien L.-A. 2007) and RSK4 (Roux 2007)), which are 73 to 80% identical in sequence to each other, with the most divergent regions in the N- and C-terminal (Romeo, Zhang et al. 2012). The RSKs contain two functional and non-identical kinase domains, one N-terminal kinase domain (NTKD) and one C-terminal kinase domain (CTKD) (Jones, Erikson et al. 1988, Fisher and Blenis 1996). These kinase domains are connected by a linker region of approximately 100 amino acids containing essential regulatory domains. The NTKD belongs to the AGC family of kinases and are responsible for substrate phosphorylation, while the CTKD belongs to the Ca\(^{2+}\)/calmodulin-dependent kinase family and is responsible for activation of NTKD through auto-phosphorylation of the hydrophobic motif in the linker region (Bjorbaek, Zhao et al. 1995, Fisher and Blenis 1996, Vik and Ryder 1997). See figure 5 for overall domain organization.

All RSKs can be activated by ERK1/2 through the C-terminal located D domain (Gavin and Nebreda 1999, Smith, Poteet-Smith et al. 1999). The D domain consists of Leu-Ala-Gln-Arg-Arg, where Leu and Arg are essential (Roux, Richards et al. 2003). Two additional basic residues C-terminal of the D domain may also contribute to the ERK1/2 docking, but these are not essential (Roux, Richards et al. 2003). In RSKs four out of six known phosphorylation sites are important for activation (Ser221, Ser363, Ser380 and Thr573 in human RSK1) (Dalby, Morrice et al. 1998). Upon mitogen stimulation, ERK1/2 phosphorylates both Thr573 located in the activation loop of CTKD and Thr359/Ser363 in the linker region (Sutherland, Campbell et al. 1993, Smith, Poteet-Smith et al. 1999, Ranganathan, Pearson et al. 2006). Mitogen activation also give auto-phosphorylation at Ser380 within the hydrophobic motif (Vik and Ryder 1997), creating a docking site for phosphoinositide-dependent protein kinase 1 (PDK1) (Frodin, Jensen et al. 2000). PDK1 association with RSKs leads to phosphorylation of Ser221 in the activation loop of NTKD, resulting in full activation of the enzyme (Jensen, Buch et al. 1999, Richards, Fu et al. 1999).

Recent studies have described a cross-talk between protein kinase A (PKA) and the ERK1/2 signaling pathway, involving both PKA activity and cellular distribution of RSK (Chaturvedi, Poppleton et al. 2006, Chaturvedi, Cohen et al. 2009). Inactive RSK interacting with the regulatory subunit (RI) of PKA will sensitize PKA to cAMP, while activation of RSK will promote its interaction with the catalytic subunit of PKA. This decreases cAMP’s ability to stimulate PKA. At the same time interaction between PKA and active RSK1 will ensure RSK1s nuclear localization.
RSK1-3 are ubiquitously expressed in all human tissues tested (Zeniou, Ding et al. 2002). However, there are tissue variations between the three RSKs, suggesting isoform specific function. RSK1 has been found to be predominantly expressed in the kidneys, lungs and pancreas, whereas both RSK2 and 3 are highly expressed in skeletal muscle, heart and pancreas (Alcorta, Creus et al. 1989, Moller, Xia et al. 1994, Zeniou, Ding et al. 2002). In the brain, the three RSKs are expressed in different areas (Zeniou, Ding et al. 2002, Heffron and Mandell 2005). The expression of RSK4 is lower than for the three others, and is found expressed in the brain, heart, cerebellum, kidneys and skeletal muscles (Dummler, Hauge et al. 2005). In resting cells RSK isoforms are found in the cytoplasm, but upon stimulation the RSKs are translocated to the nucleus (Chen, Sarnecki et al. 1992, Zhao, Bjorbaek et al. 1995). However, activated RSK2 is found both in the cytoplasm and the nucleus (Chen, Sarnecki et al. 1992).

Studies have shown that RSK isoforms regulate nuclear signaling, cell-cycle progression, cell proliferation, cell growth and protein synthesis, cell migration and cell survival (Roux and Blenis 2004, Doehn, Hauge et al. 2009).

**Figure 5: Domain organization of RSK.** RSK is characterized by the presence of two separate kinase domains, N-terminal kinase domain (NTKD) and C-terminal kinase domain (CTKD), which are connected by a linker region. ERK1/2 bind to ERK docking motif and phosphorylates Thr573 in the activation loop of CTKD and Thr359/Ser363 in the linker region. This phosphorylation triggers auto-phosphorylation of Ser380 within the hydrophobic motif and enables the docking of PKA which phosphorylates Ser221 in the activation loop of NTKD. Modified from (Anjum and Blenis 2008).
Mitogen- and stress-activated protein kinase 1 and 2 (MSK1 and MSK2) were discovered simultaneously by two groups in 1998/1999 by genome-wide homology searches (Deak, Clifton et al. 1998, New, Zhao et al. 1999). At the same time, a group in Basel, Switzerland discovered MSK2 in a two-hybrid screen using p38 as bait (Pierrat, Correia et al. 1998). The human versions of MSK1 and MSK2 are 63% identical in sequence and display about 40% identity to the RSK sequences. MSK1 and MSK2 contain two different kinase domains within the same polypeptide, a feature shared with the RSKs (see figure 6). The N-terminal kinase domain belongs to the AGC family of kinases, while the C-terminal kinase domain has a CaMK-like sequence and highest sequence homology to the kinase domain of MK2/3 (40% amino acid identity) (Roux and Blenis 2004). In addition to these kinase domains MSK1/2 contain an N-terminal tail, a linker region between the two kinase domains and a C-terminal tail.

MSK1 and MSK2 are ubiquitously expressed in all tissues examined, with predominant expression in the brain, heart, placenta, and skeletal muscles (Deak, Clifton et al. 1998).

The C-terminus tail of MSK1 and MSK2 contains a functional bipartite NLS (Lys-Arg-Xaa₁₄-Lys-Arg-Arg-Lys-Gln-Lys in MSK2) resulting in an almost exclusively nuclear localization in both serum starved cells and stimulated cells (Deak, Clifton et al. 1998, Pierrat, Correia et al. 1998). Consistent with this, MSK1 and -2 are shown to regulate mainly nuclear events (Arthur 2008, Vermeulen, Vanden Berghe et al. 2009). Despite that MSKs are not being translocated after activation, they are still found to regulate the localization of their upstream activators p38α and ERK1 (Pierrat, Correia et al. 1998).

Depending on the cell type and stimulus, MSK can be activated both by the ERK1/2 and the p38 MAPK cascade.
**Figure 6: Domain organization of MSK.** MSKs are characterized by having two separate kinase domains, N-terminal kinase domain (NTKD) and C-terminal kinase domain (CTKD), which are connected by a linker region. ERK or p38 phosphorylate Ser360, Thr581 and Thr700 (marked by star in figure). The activated CTKD proceeds to phosphorylate Ser212 in NTKD, Ser376 in hydrophobic motif and Ser381 in linker region. The activated NTKD can then phosphorylate MSKs substrates. Modified from (Vermeulen, Vanden Berghe et al. 2009).

**MNK**

MAPK-interacting kinases 1 and 2 (MNK1 and MNK2) were discovered at the same time by two groups in 1997 as a result of screening for interaction partners of ERKs (Fukunaga and Hunter 1997, Waskiewicz, Flynn et al. 1997). MNK1 and MNK2 display approximately 70% sequence identity (Roux and Blenis 2004). Their single kinase domain belongs to the CaMK family, and is most similar in sequence to the CTKD of RSKs, MK2 and MK3 (Waskiewicz, Flynn et al. 1997). In comparison to other MAPKAPKs kinase domains MNK1/2 has a DFD (Asp-Phe-Asp) motif in the activation segment where most other kinases have DFG (Asp-Phe-Glu). In addition the kinase domain of MNK1/2 has two short inserts, one immediately after the DFD motif and the other following the APE motif in the activation segment (Buxade, Parra-Palau et al. 2008). Two isomers exist of both MNK1 and MNK2, one long and one short that lacks the C-terminal MAPK-binding motif (see figure 7). All four variants contain very closely related catalytic domains and a polybasic region (PBR) N-terminal to the catalytic domain responsible for binding importin α and the translational factor scaffold protein eIF4G (Pyronnet, Imataka et al. 1999, Waskiewicz, Johnson et al. 1999). The activation loop of all isoforms contains phosphorylatable threonine residues followed by prolines, which correspond to the phosphorylation site for MAPK. In the C-terminal end only MNK1a contains a functional NES and the MAPK docking domain is only present in MNK1a and MNK2a. Despite of this MNK1b may still promote activity (see figure 7) (Buxade, Parra-Palau et al. 2008)
Figure 7: Schematic representation of the four known MNKs. All four contain a polybasic region (PBR) which binds importin α and eIF4G as well as the catalytic domain. Only MNK1a contains a functional NES and MNK1a and MNK2a contain a MAPK docking site. The shorter MNK1b and MNK2b lack both NES and MAPK docking site, although MNK1b may still promote activity. Figure modified from (Buxade, Morrice et al. 2008).

Since only MNK1a contains a functional NES, this isoform is mainly localized in the cytoplasm, while MNK1b and MNK2b, which lack the NES, will reside in the nucleus. For MNK2a the C-terminus may interfere with binding of its polybasic region to importin α and eIF4G and keep the protein in the cytoplasm (Scheper, Parra et al. 2003).

MNKs are expressed in all adult tissues examined, with the exception of brain, where the level is greatly reduced (Waskiewicz, Flynn et al. 1997). MNKs are activated either by ERK1/2 or p38. ERK1/2 activates MNK2 in response to growth factors and cytokines, while p38 may phosphorylate both MNK1 and MNK2 when cells are exposed to cytotoxic stress (Waskiewicz, Flynn et al. 1997). So far a number of substrates have been identified for the MNKs, including eIF4G and eIF4H, heterogeneous nuclear RNA-binding protein A1 (hnRNP), polypyrimidine-tract binding protein-associated splicing factor (PSF), cytoplasmic phospholipase A2 and sprouty (Waskiewicz, Flynn et al. 1997, Pyronnet, Imataka et al. 1999, Hefner, Borsch-Haubold et al. 2000, Buxade, Parra et al. 2005, DaSilva, Xu et al. 2006, Buxade, Morrice et al. 2008). The exact biological roles of the MNKs are still somewhat unclear, but an involvement in the immune system, inflammatory response, and cell survival/proliferation has been established (Buxade, Parra-Palau et al. 2008, Moens, Kostenko et al. 2013).
MK2 and MK3

MAPK-activated protein kinase 2 (MK2) was first discovered in 1992 by Stokoe et al. as an ERK1/2 activated protein kinase that could phosphorylate heat shock protein 25 (Hsp25) and Hsp27 (Stokoe, Campbell et al. 1992). However, two years later two independent groups determined that MK2 was in fact activated by p38 as a response to stress stimuli (Freshney, Rawlinson et al. 1994, Rouse, Cohen et al. 1994). MAPK-activated protein kinase 3 (MK3) was discovered a few years later by two independent groups (McLaughlin, Kumar et al. 1996, Sithanandam, Latif et al. 1996). The former group identified MK3 by using a two-hybrid screen for p38-interacting proteins, while the latter did it by analyzing genes commonly deleted in small-cell lung cancer.

MK2 and MK3 are highly homologous (MK2 has a 75% amino acid similarity with MK3), which indicated that these two enzymes are closely related. Their kinase domain is most identical to CaMK (35-40% identical) and the CTKD of the RSK isoforms. Vertebrate MK2/MK3 contain a proline-rich N-terminal region involved in interaction with the Src homology 3 (SH3) of c-Abl in vitro (Plath, Engel et al. 1994). The C-terminal end of MK3 contains a functional bipartite NLS (Lys-Lys-Xaa_{10}-Lys-Arg-Arg-Lys-Lys), which is also present in MK5. The NLS in both MK2 and MK3 encompasses a D-domain (Leu-Leu-Lys-Arg-Arg-Lys in MK2) that is important for the interaction with p38α and β (Smith, Poteet-Smith et al. 2000). MK2 and MK3 also contain a functional NES (Met-Thr-Ser-Ala-Leu-Ala-Thr-Met-Arg-Val) N-terminal to their NLS, which is regulated by phosphorylation (Engel, Kotlyarova et al. 1998). See figure 8 for overall domain organization.

MK2 and MK3 are ubiquitously expressed in all tissues examined with predominance to heart, skeletal muscles and kidneys (Stokoe, Campbell et al. 1992, Engel, Plath et al. 1993, Sithanandam, Latif et al. 1996). MK2 is however expressed at a significantly higher level than MK3 (Ronkina, Kotlyarova et al. 2007). Both the mk2 and mk3 genes give rise to two alternative isoforms (Stokoe, Campbell et al. 1992, Chevalier and Allen 2000, Moise, Dingar et al. 2010). The shorter isoforms (MK2S and MK3S) lack part of the C-terminal region and therefore miss the nuclear export/import signal and the MAPK binding domain. In addition MK3S also lacks some catalytic subdomains. MK2S and MK3S are predominantly localized to the cytoplasm (Zu, Wu et al. 1994, Moise, Dingar et al. 2010), while MK2 and MK3 are mainly found in the nucleus of quiescent cells (Engel, Kotlyarova et al. 1998, Neufeld, Grosse-Wilde et al. 2000) and these kinases are transported to the cytoplasm upon stress stimulation (Ben-Levy, Hooper et al. 1998, Engel, Kotlyarova et al. 1998).

MK2 and MK3 become activated by various stress conditions that stimulate different p38 isoforms, such as UV irradiation, heat shock, oxidative stress, hyperosmolarity and cytokines (Freshney,
Activated p38 phosphorylates MK2 and MK3 at residue T222 (human MK2 numbering) located in the T-loop, at S272 within the catalytic domain and at a regulatory phosphorylation site T334 which is located in a hinge region between the catalytic core and the auto-inhibitory helix (Ben-Levy, Leighton et al. 1995, Engel, Schultz et al. 1995). In addition to these amino acids, some minor (auto)-phosphorylation sites were also described (Ben-Levy, Leighton et al. 1995). Through this p38 cascade, MK2 and MK3 participate in diverse cellular processes such as cytokine production, endocytosis, reorganization of cytoskeleton, cell migration, cell cycle control, chromatin remodeling and gene expression (Guay, Lambert et al. 1997, Kotlyarov, Neininger et al. 1999, Hannigan, Zhan et al. 2001, Wu, Hannigan et al. 2004, Yannoni, Gaestel et al. 2004, Manke, Nguyen et al. 2005, Voncken, Niessen et al. 2005, Rousseau, Dolado et al. 2006, Ronkina, Kotlyarova et al. 2007). Animal studies by Kotlyarov et al. showed that MK2 deficient mice have increased stress resistance and survive LPS-induced endotoxic shock (Kotlyarov, Neininger et al. 1999). Also, various cell studies, with cells obtained from MK2 deficient mouse, have shown MK2’s involvement in inflammation (reviewed in Moens et al. (Moens, Kostenko et al. 2013)).

MK5

Mitogen-activated protein kinase-activated protein (MAPKAP) kinase 5, also called MK5, was originally discovered by the research group of Ni, as a novel murine kinase that could be phosphorylated and activated by ERK and p38 but not by Jun N-terminal kinase (JNKs) in vitro (Ni, Wang et al. 1998). The same year the research group of Han also described a protein kinase activated downstream of p38 and called it PRAK for p38 regulated/activated protein kinase (New, Jiang et al. 1998). This was the human analog of MK5. The human mk5 gene codes for two different splice variants of MK5; one with 471 amino acids (MK5A) and one with 473 amino acids (MK5B). The reason for the two nearly identical variants of one protein remains to be elucidated. Five splice variants have been described for mouse MK5, but the biological roles of the different isoforms are unknown (Dingar, Benoit et al. 2010).

MK5 is found only in vertebrates (Gaestel 2006) and is ubiquitously expressed throughout the human body. It has a predominant expression in the heart, skeletal muscles, pancreas and lungs (New, Jiang et al. 1998, Ni, Wang et al. 1998, Perander, Keyse et al. 2008, Gerits, Shiryaev et al. 2009). In resting cells the protein resides predominantly in the nucleus but is able to shuttle between the nucleus and the cytoplasm. Nucleocytoplasmic shuttling is controlled through interaction with PKA, Cdc15A and
the upstream kinases ERK3/4 and p38 (Seternes, Johansen et al. 2002, New, Jiang et al. 2003, Schumacher, Laass et al. 2004, Kant, Schumacher et al. 2006, Gerits, Mikalsen et al. 2007, Deleris, Rousseau et al. 2008, Hansen, Bartek et al. 2008, Gong, Ming et al. 2010, Kostenko, Shiryaev et al. 2011). The in vivo interaction between p38 and MK5 is, however, under some debate, and is currently not completely resolved (reviewed in (Shiryaev and Moens 2010)).

MK5 shares 45% and 46% sequence identity with MK2 and MK3, respectively (Ni, Wang et al. 1998). The fact that the sequence identity is at approximately 45% and not 75%, as between MK2 and 3, might indicate that MK5 is a more distant homolog to these proteins (see figure 4), and that it originated earlier during evolution from a common ancestral protein.

A comparison of MK2, MK3 and MK5 (figure 8), reveals that the primary structure shares a high homology in the kinase domain, where the important ATP-binding site is found. MK2 and MK3 contain an N-terminal proline rich region that interacts with Src-homology-3 (SH3) domains in vitro (Plath, Engel et al. 1994). This region is not present in MK5. The C-terminal end of MKs contains sequence regions involved in sub-cellular targeting of kinases to either the nucleus (NLS) or the cytoplasm (NES). In MK2 and MK3 these regions are located in distinct areas while they are overlapping in MK5. The NES signal consists of MK5 residue 345 to 354 (Leu-Lys-Val-Ser-Leu-Lys-Pro-Leu-His-Ser) and triggers CRM1-dependent nuclear export, as shown by leptomycin B experiments (Seternes, Johansen et al. 2002). It can also be pointed out that Leu337, which is located within a PKI-like NES sequence N-terminal of the NES signal, is important for auto-regulation of MK5 activity as mutation of this residue renders MK5 more active than wild-type MK5 (Seternes, Johansen et al. 2002). The NLS signal consists of residue 360-365 (Leu-Arg-Lys-Arg-Lys-Leu) in the primary sequence of MK5 and is found to be functional because alanine substitution disrupts the nuclear localization of MK5 (Seternes, Johansen et al. 2002). The NLS, in MK2, MK3 and MK5, overlaps a D-domain (Ile-Leu-Arg-Lys-Arg-Lys-Leu-Leu) and is believed to be involved in the interaction with p38α and p38β (Tanoue, Adachi et al. 2000, Tanoue, Maeda et al. 2001, Seternes, Johansen et al. 2002). P38 binding to the D-domain of MK5 will mask the NLS signal and promote the nuclear export of MK5 (Seternes, Johansen et al. 2002). This ability of p38 to translocate MK5 does not depend on activation by p38, but only on the binding of p38 to MK5. This has been confirmed by using a kinase dead mutant of p38α (p38αAGF) (Raiingeaud, Gupta et al. 1995), and by co-expression of p38 and MK5 in NIH 3T3 cells (Seternes, Johansen et al. 2002). In addition to these structural differences between the MKs, MK5 also differs by its unique C-terminal sequence believed to be involved in ERK3/4 interactions (Perander, Keyse et al. 2008).
Figure 8: Schematic primary structure of MK2, MK3 and MK5. MK2 and MK3 contain an N-terminal proline rich region which is important in SH3 domain binding. This region is not found in MK5. All three MKs share the same CaMK-like domain which contains the activation loop with the phosphorylation site important in ATP binding. In addition, MK2 and MK3 contain a phosphorylation site outside the kinase domain. The MKs share high homology in the CaMK kinase domain. All three kinases have a nuclear localization and export signal located in the C-terminal end. In MK2 and MK3 these regions are in distinct areas while they overlap in MK5. In addition to this, MK5 has a unique C-terminal end believed to be involved in interaction with ERK3/4. Figure modified from (Cargnello and Roux 2011).

A lot of experimental work has been performed to elucidate the biological roles of MK5. Studies by Tak and coworkers (Tak, Jang et al. 2007), Moens and coworkers (Gerits, Mikalsen et al. 2007, Kostenko, Johannessen et al. 2009, Kostenko, Shiryaev et al. 2011), and Choi and coworkers (Choi, Choi et al. 2012) have established the relationship between MK5, hTid-1 and Hsp27 in F-actin rearrangement and cell migration. Several studies have shown the importance of MK5 in cell cycling and proliferation. PRAK was reported to suppress oncogenic ras-induced proliferation (Chen, Hitomi et al. 2000), while overexpression of MK5 inhibited proliferation of NIH3T3 cells (Li, Zhang et al. 2008). MK5 was also found to be essential for ras-induced senescence and thereby to act as a tumor suppressor (Sun, Yoshizuka et al. 2007). Later it was also discovered that MK5 may act as a tumor suppressor through down-regulation of Myc (Kress, Cannell et al. 2011). MK5 may also be involved in inhibition of cell proliferation through ERK3 interaction (Coulombe, Rodier et al. 2003, Julien, Coulombe et al. 2003). Moreover, MK5 may repress cell invasiveness (Stohr, Kohn et al. 2012). Recent studies demonstrated that MK5 can act as a tumor promoter (Yoshizuka, Lai et al. 2012). The authors reported that MK5 stimulates angiogenesis (Yoshizuka, Chen et al. 2012). The same group also unveiled a role of MK5 in cell growth arrest induced by energy starvation (Zheng, Wang et al. 2012).
Furthermore, animal studies have suggested that MK5 is involved in neurological processes controlling anxiety and locomotion (Gerits, Van Belle et al. 2007). Despite all these described functions the exact biological role of MK5 still remains elusive, also because mk5-/- mice have no obvious phenotype (Shi, Kotlyarov et al. 2003).

General three-dimensional structure of protein kinases

The overall three-dimensional structure of protein kinases is highly conserved (See figure 9). They consist of a small N-terminal lobe and a larger C-terminal lobe connected by a hinge region, which forms the outer rim of the ATP binding pocket. The first description of a protein kinase structure was of PKA by Knighton et al. (Knighton, Zheng et al. 1991). This structure showed that the N-terminal lobe is made mainly of five antiparallel β-sheets except one α-helix, referred to as the αC-helix. The two first β-sheets (β1 and β2) are connected by a conserved glycine-rich loop (GxGxxG). The glycine-rich loop is the most flexible part of the N-lobe, and helps positioning the β- and γ-phosphates of ATP for catalysis. The β3-strand typically contains an AxK sequence, where the lysine residue (Lys51 of MK5) couples the α- and β-phosphates of ATP to the αC-helix. Near the center of the αC-helix a conserved glutamate residue (Glu62 of MK5) can be found (Taylor and Kornev 2011, Roskoski 2012). A salt bridge between the β3 lysine and the αC glutamate is a prerequisite for the formation of the active state of the kinase (called the αC-in conformation). The absence of this salt bridge indicates that the kinase is in inactive state (Kornev, Haste et al. 2006). The C-terminal lobe compromises mainly of α-helixes and a few conserved β-sheets which contain most of the catalytic residues involved in the phospho-transfer between the kinase and its substrates. The activation segment, located in the C-terminal lobe, is defined as the sequence starting with the Asp-Phe-Gly (DFG) motif and ending with the conserved Ala-Pro-Glu (APE) motif (APQ in MK5) (Taylor and Kornev 2011, Roskoski 2012). In this activation segment the primary phosphorylation site (Thr182 of MK5) can be found. This primary phosphorylation site makes ionic contacts with arginine residues of the basic Arg-Asp (RD) pocket upon phosphorylation and ensures a active conformation of the kinase (Johnson, Noble et al. 1996). Protein kinases may contain one or more additional phosphorylation sites in the activation segment, but these do not interact with the RD pocket and are called secondary phosphorylation sites. One of the most important catalytic amino acids is the aspartic acid (Asp169 of MK5) of the DFG motif (Buechler and Taylor 1988). It makes contact with all three phosphate groups of ATP either directly or through the coordination of two magnesium ions, and is important for the correct positioning of the ATP γ-phosphate for transfer to the substrate (Madhusudan, Akamine et al. 2011).
The orientation of Asp169 is highly dependent both on phosphorylation of the activation loop, and upon the phenylalanine of the DFG motif. The phenylalanine makes hydrophobic contact with both the αC-helix and the catalytic loop (HRD motif) thereby facilitating the correct orientation of Asp169 and accommodating the αC-helix’s interaction between Glu62 and Lys51 (Kornev, Haste et al. 2006). The catalytic loop is a segment of residues flanked by β-sheet 6 and 7 and sits at the bottom of the ATP pocket in the C-terminal lobe. It contains a conserved HRD motif with a catalytic aspartic acid residue (Asp148 of MK5) which is required for optimal phosphotransfer from ATP to the substrate (Huse and Kuriyan 2002).

Figure 9: The Cα-trace of the MK5 homology model based on the X-ray structure of MK3 in complex with an inhibitor (PDB id: 3FHR) described in Lindin et al. (Lindin, Wuxiuer et al. 2013). In the figure conserved regions are marked with both text and color coding; hinge region with gatekeeper residue (green), P-loop (yellow), αC-helix (dark blue), activation segment with DFG and APQ motif (purple), HRD catalytic loop (red) and Lys51 of β-sheet 3 (light blue).
Inhibitors

Protein phosphorylation is a key step in many crucial cellular processes like cell proliferation, differentiation and apoptosis. Because of this, there has been a growing interest in the discovery of small molecule kinase inhibitors for novel drug research and development, as well as for the identification of experimental tools for the understanding of the biological roles of protein kinases.

The majority of kinase inhibitors that have been developed so far are known as type I inhibitors. They target the ATP binding site of the kinase in its active conformation, in which the activation loop is phosphorylated. Normally N1 and N6 atoms of the ATP adenine ring form hydrogen bonds with backbone carbonyl and amide groups of the hinge region. Interestingly, one or both of these bonds are typically replicated by type I inhibitors (Rockey and Elcock 2006). The glycine rich loop covers and anchors the non-transferable phosphates of ATP, and often also binds to inhibitors. The β3 lysine helps anchoring the α- and β-phosphate of ATP when the kinase is in the active state, and is stabilized by a conserved catalytic glutamate residue (Glu62 in MK5) in the center of the αC-helix. This lysine residue might also interact with an inhibitor. An example of type I inhibitor is the compound GLPG0259, which was developed as a MK5 inhibitor for use in inflammatory diseases. It was shown to diminish inflammation in the mouse collagen-induced arthritis model and it was safe and well-tolerated in healthy subjects (Namour, Vanhoutte et al. 2012, Westhovens, Keyser et al. 2013). However, the inhibitor was stopped in phase II studies on patients with active rheumatoid arthritis (RA) where it was revealed that GLPG0259 did not display any beneficial effects compared with present drug therapy.

It is often a concern that it is difficult to achieve selectivity when targeting an inhibitor to the ATP pocket due to the high structural conservation between the 518 human kinases. It could however be achieved by utilizing the naturally occurring structural idiosyncrasies in the ATP-binding pocket (Rockey and Elcock 2006). Development of irreversible inhibitors that form covalent bonds with cysteine in the ATP-binding pocket is currently gaining interest (Liu, Sabnis et al. 2013). Irreversible kinase inhibitors may have potential advantages including prolonged pharmacodynamics and suitability for rational drug design. Reversible ATP competitors have to compete with high intracellular ATP concentrations, typically 1-10mM (Beis and Newsholme 1975). The use of irreversible protein kinase inhibitors can overcome this competition and therefore lower dosages can be used (Garuti, Roberti et al. 2011).

In addition to type I of inhibitors, it has become increasingly more common to target the inactive (DFG-out) conformation. These type II inhibitors don’t compete with ATP and exploit the unique
conformational changes in an inactive kinase. When the inhibitor binds to the closed conformation of the activation loop it prevents binding of both ATP and the substrate. The flip from DFG-in to DFG-out introduces a conformational change in the kinase and exposes an additional hydrophobic site close to the DFG-motif, which is possible for inhibitors to exploit (Zuccotto, Ardini et al. 2010). An example of a type II inhibitor that has reached the market is imatinib 1 (Gleevec, Novartis) (Schindler, Bornmann et al. 2000). Imatinib 1 is a cKIT, Abl and PDGFR inhibitor approved for chronic myeloid leukemia treatment.

Type III inhibitors typically bind close outside the ATP-binding site, at an allosteric site, and modulate the kinase activity in an allosteric manner (Zhang, Yang et al. 2009). The most well-characterized allosteric kinase inhibitor is CI-1040, which inhibits MEK1 and MEK2 by occupying a pocket adjacent to the ATP binding site (Ohren, Chen et al. 2004).

Type IV inhibitors bind to allosteric pockets further away from the ATP pocket than the type III inhibitors, like the substrate binding pocket (Cox, Shomin et al. 2011, Lamba and Ghosh 2012). Both Type III and type IV inhibitors obtain specificity more easily than type I and II, because these targeted pockets are structurally more diverse between the kinases. Another type of inhibitors are the bisubstrate inhibitors (Type V) which consist of two conjugated fragments, each targeted to a different binding site of a bisubstrate enzyme (Cox, Shomin et al. 2011, Lamba and Ghosh 2012).
Methodical considerations

The completion of the human genome project has resulted in an increasing number of sequenced proteins. In order to understand the function and potential therapeutic use of all these proteins it is important to understand both the physical and chemical properties of both the proteins themselves and the interactions with their interaction partners. This is easier done if the molecular structures are constructed and visualized in three-dimensional models. Traditionally this is done either by X-ray crystallography or NMR spectroscopy, but these methods are both costly and time consuming. An alternative to these traditional experimental methods are molecular modeling. Molecular modeling encompasses all theoretical methods and computational techniques used to model or mimic the behavior of molecules and molecular systems. Typically they are used in the fields of computational chemistry, drug design, computational biology and materials science for studying molecular systems ranging from small chemical systems to large biological molecules and material assemblies. The simplest calculations can be performed by hand, while computers are required to perform molecular modelling of any reasonably sized system.

Molecular modelling comprises both molecular mechanics and quantum mechanics. Molecular mechanics is used for calculations concerning atomic nuclei, while quantum mechanics is used for calculations concerning electronic systems. Molecular mechanics considers the atomic structure of molecules to be a collection of atomic masses that interact with each other via harmonic forces and ignore the electronic motions. Its calculations are based upon Newton’s Classical Mechanics equations, and use the Born-Oppenheimer approximation, which states that atomic nuclei move much slower than electrons, so the vibrational and rotational motions of a molecule can be separated from the electronic motion. Molecular dynamics (MD), Monte Carlo and global energy minimization are examples of calculations where molecular mechanics are used.

Quantum mechanics enable the calculation of the energy of an electronic system. Geometries and structures of small molecules can be predicted, but protein molecules are too big to be solved by the Schrödinger equation $E\psi = H\psi$ (which states that the Hamiltonian operator acts on a certain wave function $\psi$, and the result is proportional to the same wave function in a stationary state and the proportionality constant, $E$, is the energy of the stat $\psi$). Quantum mechanics is useful for calculation of electrostatic potentials (ESP) of small molecules.

In the following sections the various molecular modeling techniques (homology modeling, docking, virtual ligand screening (VLS), MD) used in this thesis will be discussed. Also included is the experimental method used to verify the results from VLS.
**Homology modeling**

In order to understand the mechanism of protein function it is important to understand the three-dimensional structure. Determining the protein structures by experimental methods like X-ray crystallography and NMR spectroscopy is both time consuming and by far successful for all proteins. To aid in this work computational methods can be applied. One of these methods is homology modeling, which predicts the three-dimensional structure of the target based on sequence similarity to one or more proteins with known structure (template). This method is especially applicable for proteins of the same family because they frequently have noticeable similarities and thus share the three-dimensional architecture. This means if only one member of a family is solved structurally it might be possible to get a structural prediction of the other members as well.

Homology modeling consists of the following four steps (Fiser 2010):

1. Identify homologs with known structure in the Protein Data Bank (PDB).
2. Align the query sequence with template sequence.
3. Build the models based on the alignment.
4. Assess, refine and test the models.

The accuracy of the homology model is highly dependent on the sequence similarity between the template and target proteins. For water soluble proteins a template with 30% sequence similarity to the target is considered borderline of what can be considered as realistic modeling (Xiang 2006). The accuracy of the homology model also depends on the quality of the available template X-ray crystal structure, the alignment created and the performance of the computer program chosen for modeling.

In the present study the homology macro of ICM was employed for building the model. The homology macro of ICM (Abagyan 1994) constructs the backbone of the target molecule by homology from core sections of the template molecule. Core sections are defined by the average of Cα atom positions in these regions, and side chain torsion angles are then predicted by simultaneous global optimization of the energy for all non-identical residues. Loops are subsequently searched for among several thousand structures in the PDB data bank (Berman, Westbrook et al. 2000) and matched in regard to sequence similarity and steric interaction with the surroundings of the model. Best fitting loops were selected by calculating maps around loops and scoring their relative energies.

The created models were refined using the RefineModel macro of ICM which globally optimizes side chains and anneals the backbone. The macro includes: (1) Monte Carlo fast (Abagyan and Totrov
1994) simulation for sampling of the conformational space of side chains, (2) iterative annealing of
the backbone with tethers, which are harmonic restraints pulling an atom in the model to a static
point in space represented by a corresponding atom in the template and (3) a second Monte Carlo
fast simulation on side chains. The Monte Carlo method is a complete search of the conformational
space of a protein or part of a protein, and at each stage the actual conformation is modified
randomly in order to obtain a new one. Each iteration samples the conformational space of a
molecule with the ICM global optimization procedure, and consists of a random move followed by a
local energy minimization, and then a complete energy calculation. Based on the energy and the
temperature, the iteration is accepted or rejected (Abagyan and Totrov 1994).

In order to analyze and validate the models constructed in this study, the SAVS Metaserver
(http://nihserver.mbi.ucla.edu/SAVS/) was employed. This server is one of the tools for checking the
stereo chemical quality of computationally constructed models, and includes the programs ProCheck
(Laskowski, MacArthur et al. 1993), What_Check (Hooft, Vriend et al. 1996), and Errat (Colovos and
Yeates 1993). ProCheck checks the stereo chemical quality of the protein structure by analyzing the
residue-by-residue geometry and the overall structure geometry, while What_Check performs
extensive checking of many stereo chemical parameters of the residues in the model based on a
subset of protein verification tools from the WHAT IF program (Vriend 1990). Errat analyzes the
statistics of non-bonded interactions between different atom types and plots the value of the error
function versus position of a 9-residue sliding window, calculated by a comparison with statistics
from highly refined structures.

It is also wise to check the overall 3D structure of the models created against their templates, and
calculate RMSD values. In the present study this was done using the DaliLite program
(http://www.ebi.ac.uk/Tools/structure/dalilite) (Holm and Park 2000).

Docking

Molecular docking has become an important tool both for the study of molecular interactions and for
drug discovery. It provides a more direct and rational approach, with the advantage of being a low
cost and effective method compared to more traditional experimental methods (Bailey and Brown
2001, Moitessier, Englebienne et al. 2008). The aim of molecular docking is to give a prediction of the
ligand-receptor complex structure using computational methods. This can be achieved through two
interrelated steps: (1) Sampling of ligand conformation in the active site of the protein and (2)
ranking of these conformations using a scoring function (Meng, Zhang et al. 2011). Both the protein
and the ligand have six degrees of translational, rotational and conformational freedom that creates a vast amount of possible binding modes. For a computer it would be impossible to calculate all possible poses and for this reason various sampling algorithms have been created and are used in the different molecular docking software available.

In the present study the molecular docking module of ICM version 3.5 (www.molsoft.com) was applied both for testing the validity and predictability of homology models (Paper I) and in a larger virtual ligand screen with compounds from publically available databases (Paper III). ICM uses a Monte Carlo global optimization procedure (Abagyan and Totrov 1994) for predicting the binding poses for a set of ligands in the space of grid potential maps calculated for the protein pocket. The procedure follows four steps (Neves, Totrov et al. 2012): (1) a random move is introduced to one of the rotational, translational or conformational variables of the ligand within the binding pocket, (2) differentiable terms of the energy function are minimized, (3) desolvation energy is calculated and, (4) the Metropolis selection criterion is used to either accept or reject the final minimized conformation (Metropolis, Rosenbluth et al. 1953). The procedure is repeated until the maximum level of steps is achieved. The maximum number of steps is determined by the number of rotatable bonds in the ligand multiplied by a user defined value of thoroughness. Calculation time for the whole global optimization procedure is greatly reduced by using pre-calculated grid maps accounting for hydrogen bonding potential, van der Waals potential, hydrophobic potential and electrostatic potential. The total ligand binding modes are scored according to the quality of the complex and a user-defined number of the top-scoring poses are re-ranked using the full ICM scoring function. This score ($\Delta G_{\text{score}}$) is calculated as a weighted ($\alpha_1$-$5$) sum of all ligand – target interactions:

$$
\Delta G_{\text{score}} = \Delta E_{\text{IntFF}} + T\Delta S_{\text{Tor}} + \alpha^1 \Delta E_{\text{HBond}} + \alpha^2 \Delta E_{\text{HBDesol}} + \alpha^3 \Delta E_{\text{SoEl}} + \alpha^4 \Delta E_{\text{HPhob}} + \alpha^5 \Delta E_{\text{Size}}
$$

Where $\Delta E_{\text{IntFF}}$ represents van der Waals interactions and internal force field of ligand, $T\Delta S_{\text{Tor}}$ represents free energy changes due to conformational energy loss upon ligand binding, $\Delta E_{\text{HBond}}$ represents hydrogen bonding interactions, $\Delta E_{\text{HBDesol}}$ represents hydrogen bond donor-acceptor desolvation energy, $\Delta E_{\text{SoEl}}$ represents solvation electrostatic energy upon ligand binding, $\Delta E_{\text{HPhob}}$ represents hydrophobic free energy gain and $\Delta E_{\text{Size}}$ is a size correction term proportional to the number of ligand atoms.
ROC curves

The overall predictability of models created by homology modeling was evaluated using ROC (receiver operating characteristic) curves (Swets, Dawes et al. 2000, Fawcett 2006). Docked inhibitors were labeled 1 (as true inhibitors) and docked decoys were labeled 0 (as false inhibitors). Score values were then analyzed using the ROC curve script incorporated into ICM. Results were displayed as ROC curves, and the area under curve (AUC) was calculated. A diagonal ROC curve would signify that the model gives no preference to true inhibitors over decoys or vice versa. A curve closer to the left hand border and the top border indicates a greater accuracy of the model (a higher ratio of true positives to false positives). This is illustrated in figure 10.

![ROC curve illustration](image)

**Figure 10: ROC curve illustration.** Red line indicates a random result, while blue line indicates a greater accuracy.

Virtual screening

Virtual screening (VS) is a computational complement to the experimental high throughput screening (HTS) used for identification and optimization of bioactive compounds. It is meant to reduce the vast expense in both time and resources used in the experimental approach. VS require knowledge about the spatial and energetic criteria responsible for the binding of a particular candidate ligand to the receptor under investigation, and is also highly dependable to the quality of the knowledge available and the computer algorithms used (Klebe 2006). However, the compounds being studied doesn’t
need to exist and one doesn’t need to take material cost, solubility, aggregate formation and other disturbing factors for experiments into account in the initial computational screen (Klebe 2006).

VS can be divided into three different categories; a target based approach, a ligand based approach, or a combination of these two. Ligand based approach uses structure-activity data from known actives to identify similar compounds in commercially available or in house databases. This can be achieved by a variety of methods, including similarity and substructure searching, quantitative structure-activity relationship (QSAR), pharmacophore matching or three-dimensional shape matching (Scior, Bender et al. 2012). The structure based approach utilizes the three-dimensional shape of the target, either determined by X-ray crystallography, NMR or homology modeling. The candidate ligands are docked into the three-dimensional structure and ranked according to binding affinity or binding site complementary. It is also a possibility to combine the two approaches. This has been done in paper III. Here we searched for bioactive compounds in commercially available databases using sub-structures from known actives, and presided to dock the retrieved compounds into a homology model and ranking them according to binding affinity.

When doing VS there are a few considerations to be taken into account. First of all the three-dimensional structures available are not always correct. Even an X-ray structure will only show one of all possible conformations a native protein can obtain, and it is important to check that the protein has is in the conformation we would like to target. It is also important to assess whether the correct binding pocket is used. In nature ligands might bind to alternative binding pockets. It is also important to decide if the ligands tested should be flexible as well as the target protein during docking, and if either of them should be protonated in one or more positions. Water molecules are often an important participant of ligand binding, and it is wise to check with the proteins crystal structure or its protein family members if any water is present during binding of bioactive compounds. After the initial screening it can be important to check for drug likeness and solubility of the compounds obtained. This is done in order to facilitate further work on the experimental side.

**Molecular dynamics**

Biological systems are dynamic in nature; analyzing their motion at the molecular and atomic level is therefore essential for the understanding of key biological phenomena. One of the tools for theoretical studies of time dependent movements in a biomolecule is molecular dynamics (MD). MD is based on Newton’s second law of motion, $F_i = m_i a_i$, where $F_i$ is the sum of all forces exerted on atom $i$ that results in its acceleration $a_i$, and $m_i$ is its mass. From the knowledge of the force of each
atom, it is possible to determine the acceleration of each atom in the system. The acceleration is defined as the second derivative of the position with respect to time:

$$a_i = \frac{dv_i}{dt} = \frac{d^2r_i}{dt^2}$$

In other words acceleration is the rate of change of the velocity ($v_i$), which in turn, is the rate of change of position ($r_i$). Integration of the equations of motion yields a trajectory which describes the position, velocities and the acceleration of particles as they vary over time. The average values of properties can be determined from this trajectory.

In the late 1950’s Alder and Wainwright were the first to introduce the MD method in a study of the interactions of hard spheres (Alder and Wainwright 1957, Alder and Wainwright 1959). In contemporary literature, one routinely finds molecular dynamics simulations of solvated proteins, protein-DNA complexes as well as lipid systems addressing a variety of issues including the thermodynamics of ligand binding and the folding of small proteins.

In the present study MD was used for analyzing the interaction between MK5 and an inhibitor, and between MK5 and its possible interaction partner p38α. Also, the overall movement and internal hydrogen bonding network in MK5 during MD of MK5 alone, was analyzed.

MD simulations were performed using the Desmond program (D. E. Shaw Research New York, NY, USA). The complexes to be analyzed were optimized using the protein preparation wizard in Maestro V9.1 (Maestro 9.1 2010) as implemented in the Schrödinger suit programs by assigning bond orders, adding hydrogen and correcting wrong bond types. The molecular systems were then neutralized by adding chloride. In Desmond, the volume of space in which the simulation takes place (the global cell) was divided into regular 3D simulation boxes of 10 Å × 10 Å × 10 Å. Each box was assigned to a single Desmond process. These boxes constituted the total simulation space with a total volume of 90 Å × 90 Å × 90 Å. The molecular systems were solvated by simple point charge (SPC) orthorhombic water box.

All complexes were refined and energy optimized using the default quick relaxation protocol of the Desmond program. In short, two rounds of steepest descent minimization were performed with a maximum of 2000 steps with and without restraints (force constant of 50 kcal/mol/Å on all solute atoms). These minimizations were followed by a series of four short MD simulations: (1) 12 ps MD simulation at a temperature of 10 K in the Berendsen NVT ensemble (constant number of particles, volume, and temperature) with solute heavy atoms restrained (force constant of 50 kcal/mol/Å); (2) 12 ps simulation was performed at 10 K with the same restraints as in (1), but with the Berendsen
NPT ensemble (constant number of particles, pressure, and temperature); (3) Using the same restraint as in A and B, a 12 ps simulation was performed in which the temperature was raised to 300 K using the Berendsen NPT ensemble; (4) Finally, a 24 ps simulation at 300 K using the Berendsen NPT ensemble without restraints was used.

After the refinements and equilibration steps, MD simulations were performed for all the molecular systems using the OPLS 2005 force field (Jorgensen, Swenson et al. 1996). The pressure was kept constant at one bar and the temperature was kept at 300 K, using the Nose-Hoover chain (Martyna, Klein et al. 1992) and Martyna-Tobias-Klein methods (Martyna, Tobias et al. 1994). The short-range and long-range Columbic interactions were calculated with a cut off radius of 9 Å and with the Smooth particle mesh method (Essmann, Perera et al. 1995) (Ewald tolerance: $1.0 \times 10^{-9}$). The M-SHAKE algorithm was used to constrain bonds containing hydrogen atoms. All the simulations used a multistep RESPA integrator (Humphreys, Friesner et al. 1994) with 2.0 fs time step for bonded interaction and short range non-bonded interactions, and 6.0 fs time step for the long-range non-bonded interactions. During MD, the energies were recorded every 1.2 ps, while the coordinates were recorded every 4.8 ps.

The simulations were analyzed using the Maestro program. The Root Mean Square Deviations (RMSD) from the initial structure and the root mean square fluctuations (RMSF) were calculated during all MD simulations. All atoms were included in the calculations of the RMSF. Atomic distances important for ligand binding and structural changes in MK5 were calculated for all the simulations.

**Kinase assay**

Although substrate specificity differs among the kinases, they all utilize ATP as a phosphate donor substrate. Kinases catalyze the transfer of the γ-phosphate from ATP onto a serine, threonine or tyrosine amino acids on an acceptor molecule. This enzyme catalysis can be monitored by measuring either ADP or the phosphorylated substrate, or by quantifying the ATP consumption. There are several different techniques for measuring this activity, but the most common can be grouped in three different categories (Jia, Quinn et al. 2008): (1) radiometric assays, (2) phospho-antibody-dependent-fluorescence/luminescence assays, and (3) phospho-antibody-independent fluorescence/luminescence assays.

Radiometric assays are the earliest kinase assay technology developed (Jia, Quinn et al. 2008). These assays utilize ATP which is radiolabeled on its γ-phosphate (normally $^{32}$P or $^{33}$P). In the kinase reaction the radiolabeled isotope is transferred from ATP to the acceptor molecule and the rate of product
formation can be quantified by measuring the extent of isotope incorporation. There are various variants of the radiometric assay, but the one we used in this work is the filtration version. This assay uses phospho-cellulose membranes to separate phosphorylated substrates from radiolabeled ATP based on ionic interaction. An acid quenched kinase reaction mixture is transferred to the negatively charged filter and the positively charged phosphorylated substrate will be captured while the negatively charged unreacted radiolabeled ATP passes through. Washing steps are then used to remove the residual labeled ATP, and the remaining signal is attributed to the phosphorylated substrate. Radioactivity is then measured by adding a scintillator cocktail and using a scintillation counter. Radiometric assays are best suited for small scale screening since it produces a vast amount of radioactive waste which has to be handled the correct way.

Phospho-antibody-dependent-fluorescence/luminescence assays use phosphor-specific antibodies together with various fluorescence/luminescence detection technologies to monitor kinase reaction products.

Phospho-antibody-independent fluorescence/luminescence assays rely neither on radioactivity or phosphor-specific antibodies. It is a collection on methods that has the potential to being universal and that measures kinase activity by ADP accumulation, phosphorylated product or ATP.
Aims of study

The X-ray crystallographic three-dimensional structure of MK5 has not been resolved yet. The goal of the present study was to gain insight into the structure, function and inhibition of MK5. Through theoretical molecular modeling methods we would study the interactions of MK5 with inhibitors and partners of the MAPK pathway. Using a combination of molecular modeling studies and experimental verification we were also aiming to discover new putative MK5 inhibitors that can be used in the elucidation of its overall function. MK5 inhibitors may also be of therapeutically value.

The specific aims were:

a. Construct homology models of MK5.

b. Study the molecular mechanism involved in the interaction between MK5 and various interaction partners by performing molecular dynamic simulations on a MK5/p38α complex, a MK5/inhibitor complex, and on MK5 alone.

c. Perform a virtual screening of commercial available databases to unveil putative new MK5 inhibitors.

d. Test putative MK5 inhibitors in vitro in kinase assays and in cell culture.
Summary of papers

Paper I


The exact structure of MK5 is not known. A reliable MK5 working model is of great importance in the development of chemical probes to help elucidate the role of MK5. In this paper, homology models of MK5 were constructed based upon known homologs, and tested against known inhibitors. Three different models were constructed, based on three different crystal structures of the close homologs MK2 and MK3. Two of the crystal structures (PDB: 3FHR and PDB: 3M2W) were kinases in active conformation in complex with an inhibitor in the ATP binding site, while the last crystal structure (PDB: 2OZA) was a kinase in complex with an interaction partner. The resulting three models were quite similar except in the ATP binding pocket. The models based on the crystal structures with inhibitor had slightly larger, and more accessible, ATP-pockets. The models were tested against previously known inhibitors and decoys by semi-flexible docking, and the results were analysed using ROC curves and visual inspection of the docked complexes. The results showed that the model based on the MK3 X-ray structure in complex with a ligand (PDB: 3FHR) had the best ability to discriminate between known inhibitors and decoys (88% chance). The inhibitors bound nicely into the ATP pocket of the model and blocked the access of ATP to the known ATP binding residues. A molecular dynamic (MD) of simulation of 50 ns was performed, which showed the model to be stable for this period of time. These results indicated that this model might be a good candidate for designing further experimental studies and for structure aided drug design.
Paper II

Inger Lindin¹, Ymingjiang Wuxiuer¹, Aina Westrheim Ravna¹, Ugo Moens² and Ingebrigt Sylte¹.*


In this study, the previously constructed homology models of MK5 were used for molecular dynamics (MD) simulations of: (1) MK5 alone; (2) MK5 in complex with an inhibitor; and (3) MK5 in complex with the interaction partner p38α. MD calculations with MK5 alone showed a stable structure over 100 ns with a nice H-bonding network in the proposed ATP binding pocket. When the inhibitor was introduced it occupied the active site and disrupted the intramolecular network of amino acids while the structure itself remained stable. However, intramolecular interactions consistent with an inactive protein kinase fold were not formed. MD with p38α also showed a stable complex. Not only the p38 docking region, but also amino acids in the activation segment, αH helix, P-loop, regulatory phosphorylation region and the C-terminal of MK5 might be involved in forming a very stable MK5-p38α complex, and that p38α binding decreases the residual fluctuation of the MK5 model. Electrostatic Potential Surface (EPS) calculations of MK5 and p38α showed that electrostatic interactions are important for recognition and binding.

Paper III


Structure based Virtual Ligand Screening (VLS) was used to identify new possible MK5 inhibitors that interact in the ATP pocket of MK5. Sub-structures were generated from known inhibitors (paper I) and used in a screen of commercially available databases for compounds similar in structure. The retrieved compounds were then docked by semi-flexible docking into the best MK5 homology model from paper I. A selection of the best scoring compounds from VLS were ordered and tested by in vitro kinase assay. Several compounds were identified as promising MK5 inhibitors, and one of them with an IC₅₀ value of 22.7 µM in vitro. The identified compounds might help elucidating the precise function of MK5 and might be used as a lead in the development of high affinity MK5 inhibitors.
Discussion

Protein kinases have become central in the efforts to understand the nature of various diseases, and a lot is invested into creating effective therapeutic strategies and finding effective and selective protein kinase inhibitors. In order to succeed it is also important to focus on the structure of the kinases, their exact biological role, and how they interacts and cooperates in the signaling. Extensive work has already been done for many of the MAPK kinases, but a large amount still remains for several. MK5 is one of these kinases. There is to date no known X-ray crystal structure of MK5, and selective inhibitors are yet to be identified. Even though some of its biological roles are starting to emerge a lot needs to be done, including searching for selective inhibitors, and analyzing its structure and how it interacts with its substrates and upstream MAPK’s.

Homology models of MK5

The absence of an available X-ray crystal structure of MK5 prompted us to create a homology model of the structure. Initially three human MK5 homology models were constructed based on the structures of its close homologues MK2 (43% sequence identity) and MK3 (41% sequence identity). Three different templates were used; MK2 in inactive conformation [PDB: 2OZA] (White, Pargellis et al. 2007), MK2 in active conformation with a ligand bound [PDB: 3M2W] (Revesz, Schlapbach et al. 2010) and MK3 in active conformation with ligand bound [PDB: 3FHR] (Cheng, Felicetti et al. 2010).

The typical bilobal protein kinase fold consists of a small N-terminal lobe and large C-terminal lobe connected by a structurally flexible linker (hinge) region. Our three MK5 models showed that this fold is also conserved in MK5 (Figure 9). In MK5, the hinge region corresponds to the segment Glu103-Met104-Met105. The residue N-terminal of the hinge region is termed the gatekeeper residue (Met102 in MK5). The X-ray structure of ERK2 shows that the gatekeeper residue confers selectivity for binding nucleotides and small molecular inhibitors, but also seems to control auto-phosphorylation through intramolecular interactions (Emrick, Lee et al. 2006). The N-terminal lobe of ERK1 and ERK2 includes a five-stranded antiparallel β-sheet (β1–β5) and a highly conserved αC-helix. The C-terminal lobe is mainly α-helical and consists of six conserved α-helixes (αD–αI) and four small β-strands (Roskoski 2012). Figure 9 show that the N- and C-terminal lobes of the MK5 model have an overall organization similar to that of ERK1 and ERK2. X-ray structures of protein kinases show that the active site is located in between the two lobes and constitutes the binding of ATP and two magnesium ions. The active site is shielded by a loop named the ATP-phosphate binding loop (also
called the P-loop). The P-loop is located between \( \beta_1 \) and \( \beta_2 \) in the N-lobe (Figure 9). The loop contains a glycine-rich segment (GxGxG), and in ERK1 and ERK2 this loop is suggested to be important for positioning the ATP \( \beta \)- and \( \gamma \)-phosphates for phosphate transfer to the substrate (Roskoski 2012). The GxGxG domain is also conserved in MK5 and corresponds to the Gly29-Ala30-Gly31-Ile32-Ser33-Gly34. The GAGISG domain is also conserved in MK5 sequences from other species (Kostenko, Dumitriu et al. 2011). X-ray structures of ERK1 and ERK2 have shown that the \( \alpha_C \)-helix of the N-lobe may occur in an activated and inactivated orientation. Structures of inactive and active ERK1 and ERK2 have shown that in the activated structure a glutamic acid of \( \alpha_C \) forms a salt bridge with the lysine of an AXK sequence in the \( \beta_3 \) strand. In MK5, the \( \alpha_C \) helix consists of residues Pro57-Ala71, and the glutamic acid corresponds to Glu62. The \( \beta_3 \) strand is constituted by Arg47-Leu54 in MK5 and the lysine corresponds to the Lys51 of the Ala49-Leu50-Lys51 sequence. The active conformation is often named the \( \alpha_C \)-in conformation, while the inactive conformation is named the \( \alpha_C \)-out conformation of protein kinases. Three amino acids in ERK1 and ERK2 are defining a catalytically important K/D/D motif (Roskoski 2012). These amino acids are also conserved in MK5. In ERK1 and ERK2, the lysine of this motif resides in the \( \beta_3 \) strand and corresponds to Lys51 in MK5. In addition to forming a salt bridge with the glutamic acid in \( \alpha_C \) in the activated state, the lysine also binds the \( \alpha \)- and \( \beta \)-phosphorus atoms of ATP. The aspartic acids of the K/D/D motif are located in the C-terminal lobe as part of the catalytic loop and the activation segment, respectively (Figure 9). In MK5, the first aspartic acid of this conserved K/D/D motif is part of a HRDLK segment in the catalytic loop, located in the C-terminal of the \( \beta_6 \)-strand, and corresponding to His146-Arg147-Asp148-Leu149-Lys150-Pro151 in MK5. In both ERK1 and ERK2 this aspartic acid is suggested to position the substrate hydroxyl group, and extract the proton of the OH-group and facilitate a nucleophile attack of the substrate oxygen on the \( \gamma \)-phosphorus atom of ATP. The aspartic residue thereby functions as a base in the catalytic reaction (Roskoski 2012). The ERK1 and ERK2 structures also show that the lysine (corresponding to Lys150 in MK5) of this segment binds the oxygen of the substrate hydroxyl group and the \( \gamma \)-phosphorous atom of ATP. The second aspartic acid in the K/D/D motif of ERK1 and ERK2 is the first amino acid in the activation segment that begins with the DFG motif corresponding to Asp169-Phe170-Gly171 in MK5. The DFG motif is followed by the activation loop that contains the regulatory phosphorylation site (Thr182 in MK5). In an active protein kinase, the aspartic acid is pointing into the active site and interacts with two Mg\(^{2+} \) ions that interact with the \( \alpha \)-, \( \beta \)- and \( \delta \) phosphates of ATP. The activation loop is fundamental both for substrate binding and catalysis.

The most striking differences between the three constructed MK5 models were found in the shape of the proposed ATP binding pocket and in the flexible loop structures of the activation segment in the
C-terminal lobe. The 2OZA-based model seemed to have a somewhat narrower ATP binding pocket than the other two models, which may be due to the fact that the crystal structure 2OZA is in the apo- (unbound) form, whilst the other two templates have a ligand bound in the ATP pocket. The position of the P-loop was different in the three models. In the 2OZA-based model the gap between Ile32 in the middle of the P-loop and the Asp169 of the DFG motif was 8.68 Å. The same distance in the 3M2W- and 3FHR-based MK5 models respectively were of 8.55 Å and 6.89 Å.

In order to evaluate the models’ ability to distinguish between real inhibitors and decoy substances, the three different models were docked both with the 90 decoy ligands and 10 inhibitors in a semi-flexible mode. Common for all of the models was that strong inhibitors had a score value better than -31 (the more negative, the better is the scoring) and important amino acids for binding of ligands into the ATP pocket were Lys51, Met105 and Asp169. These are some of the amino acids that correspond to amino acids also of importance for positioning the ATP for phosphate transfer in ERK1/2 (Roskoski 2012). The best scoring ligand (figure 11) was predicted to bind closely to the previously mentioned amino acids via hydrogen bonds. The ligand was located in the ATP binding cleft between the N- and C-terminal lobes of MK5. The ligand has a planar, boomerang shaped architecture with a flexible end, mimicking ATP binding with the pyrimidine ring occupying the same position as seen for the adenine moiety of ATP in MK2 [PDB: 1NY3]. A partial stable hydrogen bond to the backbone amide of Met 105 in the hinge region via the nitrogen of the pyrimidine ring was formed. In addition, the ligand formed a limited number of hydrogen bonds to the active site residues of MK5. The oxygen of the pyrrole-pyrimidin ring interacted with the $\epsilon$-amino group of Lys 51 (1.88 Å), and the nitrogen (N5) of the pyrrole-pyrimidin ring interacted with the carboxyl group of Asp 169 (1.51 Å). The active site residues also formed a number of intermolecular hydrogen bonds contributing the structural stability of the complex. Lys51 formed a hydrogen bond with the carboxyl group of both Asp 169 (1.72 Å) and Glu62 (1.6 Å), while Glu62 formed an additional hydrogen bond with the nitrogen of Gly71 (1.68 Å). The gatekeeper residue Met102 of MK5 did not seem to make contact with this ligand. The Lysine residue 51 is positioned at the back wall of the ATP-binding niche and is highly conserved amongst mammalian protein kinases. It is believed to be involved in the orientation of the $\alpha$- and $\beta$-phosphate groups of ATP (Hillig, Eberspaecher et al. 2007). Mutation of Lys51 to Methionine or Glutamic acid in the ATP pocket of MK5 completely removed the kinase activity (Seternes, Johansen et al. 2002, New, Jiang et al. 2003). Consistent with this we observed that many of the known inhibitors bound to Lys51 in the model or blocked the access of ATP to Lys51.
Comparing the 3FHR- and 3M2W-based models with the active ADP bound crystal structure of MK2 [PDB: 1NY3], revealed that the models appeared to have adopted an active kinase conformation (Huse and Kuriyan 2002), with the catalytic residues Lys51 and Glu62 close enough to make an ion pair interaction, and Asp169 in DFG-in conformation, the same conformation as that seen in the active MK2/ADP structure [PDB: 1NY3]. In that conformation the catalytic residues can interact with ATP for phosphate transfer. ADP was docked into the ATP binding sites of the 3FHR- and 3M2W-based models, and it appears that since the 3FHR- and 3M2W-based models originated from complexes with inhibitors, the ATP-binding site was too narrow to accommodate ADP. This was confirmed by comparing the 3FHR- and 3M2W-based homology models with the X-ray crystal structure of MK2 in complex with ADP [PDB: 1NY3] which showed that the ATP binding site of the MK2 – ADP complex was slightly wider. The distance between the carboxyl group of Asp169 and Gly31 in the glycine-rich loop in the 3FHR-based model was 5.3 Å. In contrast, the distance between the carboxyl group of Asp207 and Gly73 in the glycine-rich loop of the MK2 – ADP crystal structure complex [PDB: 1NY3] was 8.63 Å. ADP in the 3FHR based model yielded the same orientation as ADP in the MK2 complex [PDB: 1NY3], but was localized approximately 1 Å further out from DFG motif.
The narrowing of the ATP binding site induced by a ligand has also been shown in the crystal structure of MK3 in complex with the high-affinity pharmaceutical lead compound (2-(2-quinolin-3-ylpyridin-4-yl)-1,5,6,7-tetrahydro-4H-pyrrolo[3,2-c]pyridin-4-one) (P4O) [PDB: 3FHR] (Cheng, Felicetti et al. 2010). In this crystal structure complex, the flexible glycine-rich loop in the binding site occupies the position of the β-phosphate of ADP, thus binding of P4O induces a conformation of a deep and narrow ATP binding pocket in MK3.

The docking results were evaluated both by score values and ROC curves, which describes the tradeoff between “specificity” and “sensitivity”. “Specificity” is the ability to avoid false positives, and “sensitivity” is the ability of the classifier to detect true positives. The area under an ROC curve indicates the quality of enrichment, and while the ROC value of a random classifier is 0.5, the ROC value of an excellent classifier is greater than 0.9 (Swets, Dawes et al. 2000, Fawcett 2006). The 3FHR-based model gave a near ideal ROC curve, with an area under the curve calculated to be 0.88, indicating this model performs very well for predicting ligand binding to the ATP site. The 2OZA- and 3M2W-based models, however, gave a graph close to the diagonal line (with an area under curve calculated to 0.67), indicating that these models gave a somewhat random result and are less suited for predicting ligand binding than the 3FHR-based model. Both 3FHR (MK3) and 2M2W (MK2) are crystal structures of kinases with a ligand in the ATP binding site. 2OZA on the other hand is in apo-(unbound) form bound to the upstream kinase p38. The models built reflect this difference in template by having a much narrower ATP binding pocket for the 2OZA-based model than the others. 2OZA (MK2) is in apo- (unbound) form, which probably explains why 2OZA is an unsuitable template for constructing an active MK5 model that can be used to identify chemical probes that bind the ATP pocket. This was also obvious when docking known inhibitors into this model. The pocket is too narrow, and poor scoring values were obtained. The ROC curves also indicate that the model was not able to distinguish between known inhibitors and decoys. However, the 2OZA model may be suitable for identification of possible type II kinase inhibitors that typically bind to kinases in the inactive state (Liu and Gray 2006).

In order to assess the stability of the MK5 model, molecular dynamic simulations were run for 50 ns with the 3FHR based model with the best ligand docked. The run indicated a very stable complex and it did not deviate significantly from the starting complex. Most of the flexibility was observed in the loop regions, which was expected.
At the moment no selective MK5 inhibitors are known, and the selected inhibitors in the present docking study are also able to inhibit MK2 and MK3. This, however, was not important at the present stage (paper I) since the model only was meant to be able to identify binders versus non binders. The 3FHR-based MK5 model can be used to elucidate the uniqueness of the MK5 ATP-binding site and may serve as a working tool for developing chemical probes using interactive drug design. Figures 12A, B and C show the EPS of the ATP binding sites of the 3FHR-based MK5 model, the 3FHR MK3 crystal structure and the 3M2W MK2 crystal structure. The EPS of the ATP binding site of the 3FHR-based MK5 model appears to be electronegative in its two deeper cavities and electropositive in its outer areas. The EPS of the ATP binding site of the 3FHR MK3 crystal structure is entirely electronegative, and the EPS of the ATP binding site of the 3M2W MK2 crystal structure is partly electropositive and partly electronegative. The sizes of the cavities of the 3FHR-based MK5 model, the 3FHR MK3 crystal structure and the 3M2W MK2 crystal structure, measured using ICM Pocket Finder, were 420 Å³, 384 Å³ and 331 Å³, respectively. Amino acids Asp169, Lys51 and Glu62 (MK5 numbering), which are important for the binding and phosphate transfer from ATP (Huse and Kuriyan 2002), are located in the similar positions in all three kinases. There are, however, some differences, the most important being Cys168 in MK5. In the corresponding position, MK2 and MK3 both have a threonine. A cysteine in the ATP binding site may represent a unique opportunity for the construction of irreversible covalent inhibitors to the nucleophilic cysteine residue (Zhang, Yang et al. 2009). Other amino acids in the ATP binding site of MK5 that are different from the corresponding positions in both MK2 and MK3 are Ala30 (leucine in MK2 and MK3), Arg47 (lysine in MK2 and MK3), and Met104 (cysteine in MK2 and MK3). All of which could be used in future lead optimization.
Figure 12: EPS of ATP binding sites. A: EPS of the ATP binding site of the 3FHR-based MK5 model. B: EPS of the ATP binding site of the 3FHR MK3 crystal structure. C: EPS of the ATP binding site of the 3M2W MK2 crystal structure. Color coding: negative (-5 kcal/mol) red, via white, to positive (+5 kcal/mol): blue. Adapted from Paper 1 (Lindin, Xu, et al. 2013).
MK5 and interaction partners

High quality X-ray structures are first choice as starting structures for MD simulations and other predictions of molecular mechanisms of action using molecular modelling. However, in the absence of such structures, simulation with high quality homology models may contribute with important structural and mechanistic insight into different classes of proteins and their interactions. Proteins and protein mechanisms that recently have been successfully studied by combining homology modelling and MD simulations include among others: the E7 protein from high- and low-risk types of human papillomavirus (Nicolau and Giuliatti 2013), the CHeW coupling protein of the chemo taxis signaling complex (Cashman, Ortega et al. 2013), lipases and their interactions with known inhibitors (Wang, Li et al. 2013), molecular mechanism of the serotonin transporter (Gabrielsen, Ravna et al. 2012), conformational changes underlying ion channel opening (Rahman, Cui et al. 2013) and molecular mechanism of G-protein-coupled receptors (Yuan, Vogel et al. 2013, Yuan, Wu et al. 2013, Arumugam, Crouzy et al. 2014).

In Paper II, the homology models of MK5 in Paper I were used for MD simulations. The model constructed by using the X-ray structure of MK3 in complex with an inhibitor [3FHR] (Cheng, Felicetti et al. 2010) as template, showed highest accuracy in selecting known MK5 binders in front of decoys during docking. This model was therefore used for 100 ns MD simulation of MK5 alone and for two simulations (100 ns and 200 ns) of MK5 in complex with a pyrrolopyrimidone-based inhibitor. The MK5 model used for MD with p38α was based on the X-ray crystal structure of MK2 in complex with p38α (White, Pargellis et al. 2007).

The structurally most flexible regions during 100 ns MD of the native MK5 model were as expected the loop regions, and particularly the region around residues 190–210. This region is at the C-terminal end of the activation loop. In most protein kinases the activation segment starts with the DFG motif (Asp169-Phe170-Gly171 in MK5) and ends with an APE motif. The APE motif is present in both MK2 and MK3, but in MK5 this motif has been replaced with an APQ motif. The APQ motif seems to be conserved between MK5 from different species. Further, the MK3 structure (PDB: 3FHR) that was the template for constructing the MK5 model is disordered in this part of the activation loop, which may indicate that this part of the activation loop is also structurally flexible in MK3. In addition, the MK5 has an insert between the APQ motif and the αF-helix compared with MK2 resulting in a longer activation loop than in MK2 and MK3. The lacking part of this loop was
constructed by homology with available loop structures in the PDB database (see Paper I), and is present at the surface of the MK5 model.

The starting MK5 model was in an active kinase conformation with interactions between Lys51 and Glu62, while Asp169 of the DFG motif in the activation segment interacts with amino acids in the P-loop (see figure 13). The hydrogen bonding network between the amino acids at the active site stabilized after about 30 ns of MD simulation and was preserved throughout the MD.

Figure 13: The network of interacting amino acids at the active site of the starting structure for MD with MK5 alone and in complex with the inhibitor. The model was based on an X-ray structure of MK3 in an active kinase conformation, and the MK5 model is therefore also in active kinase conformation. Color coding of atoms: red; oxygen, blue; nitrogen, green; carbon atoms of the inhibitor, grey; carbon atoms of the amino acids, white; hydrogen atoms. Dotted lines are showing observed interactions between amino acids during MD of the free MK5 and the MK5-inhibitor complex. Adapted from Paper II (Lindin, Wuxiuer et al. 2014).

Incorrect ligand protonation may influence ligand-protein interactions during MD. However, predicting the pKa value is not straightforward, especially when several ionisable groups exist in the ligand. According to CHEMBEL database (https://www.ebi.ac.uk/chembl/), the pKa of the pyrrolopyrimidine-based inhibitor is 6.58. pKa predictions using the Jaguar program of the Schrödinger suite of programs, and the ICM-software indicated a pKa value in the range of 6.5–6.6. It is therefore highly possible that the majority of the inhibitor molecules are in neutral form when interacting with MK5. However, it is still quite possible that some MK5 molecules may interact in protonated form. The most probable group for being protonated is the nitrogen atom of the
morpholine ring. Docking of both the neutral and protonated forms of the inhibitor showed that the morpholine ring did not participate in direct interactions with MK5, but instead interacts with water molecules on the surface of MK5 (Figure 14). Based on this, we decided to perform MD simulations with a neutral form of the inhibitor.

Figure 14: 2D plot of the stable inhibitor binding pose observed during the 100 ns MD. Amino acids and water molecules within 4 Å from the inhibitor are included in the figure. Adapted from paper II (Lindin, Wuxiuer et al. 2014).

The MD simulation showed that the loops were also the most flexible regions during 100 ns MD with the inhibitor, and the region around residues 190–210 showed the largest fluctuation. Another very flexible segment was Pro90-Arg95, which was even more flexible than during MD with the free MK5 model. This is a part of the loop between β4 and β5 (Figure 9). In MK5, this loop has an insert of 6 amino acids compared to MK2 and MK3. Further, the region around residue 240–260 was also more flexible during MD with the inhibitor than during MDs with the free model, and with p38α. This loop structure is part of the activation segment and has insertions in MK5 compared to MK2 and MK3.

The pyrrollopyrimidone-based inhibitor (Figure 14) was first discovered as a MK2 inhibitor, but was later shown to inhibit MK5 and other protein kinases (Schlapbach, Feifel et al. 2008). The MD simulation indicated that binding of the inhibitor strongly affected the hydrogen bonding network and the dynamics of the active site. The interactions between Asp169 (DFG motif) and Ile32 and Ser33 of the P-loop (Figure 14), that were observed for the free enzyme, were not present during 100
ns MD with inhibitor. During the MD with the free enzyme, Lys51 formed stable interactions both with Glu62 and with Ser33 in the P-loop (Figure 14). The Lys51-Glu62 interactions were also observed for most of the 100 ns simulation with the inhibitor, however, close binding interactions between Lys51 and Ser33 were not observed.

The 100 ns MD with the inhibitor indicated that a very stable MK5-inhibitor complex was formed and that the inhibitor was in a stable binding mode throughout the MD. During the 100 ns, the pyrimidine ring system interacted within a charged and polar part of the MK5 binding pocket, and the CO-group of the pyrimidine ring interacted with Lys51 in the αC helix (Figure 14). Water molecules also seem very important for binding the inhibitor (Figure 14). During MD, water molecules were forming hydrogen bonds with both the oxygen and the nitrogen atom of the morpholine ring, the NH-group of the pyrrolo ring and the NH-group of the pyrimidine ring (Figure 14).

In order to further test the stability of the inhibitor binding mode, a new MD simulation of the MK5-inhibitor complex was run, but now for 200 ns. The 200 ns MD was started from the same equilibrated complex as the 100 ns MD. During the 200 ns, the inhibitor binding mode was similar to the binding mode during the 100 ns. The atomic distances between central amino acids at the active site were also similar to those observed during the 100 ns MD. Taken together, the simulations with the inhibitor indicated that the inhibitor binds to an active protein kinase fold of MK5 and functions by occupying the ATP binding site. Upon inhibitor binding, the intramolecular network and dynamical motions of amino acids at the active site are affected. The intramolecular interactions involving Phe170 (DFG motif) were changed, but the interactions between Lys51 and Glu62 were maintained, indicating that a fully inactive kinase fold was not formed during the MD.

The MK5 protein includes functional binding motifs for the kinases ERK3, ERK4 and p38. Further, experimental studies have indicated that p38 can phosphorylate and activate MK5 and regulate the subcellular location of MK5 (New, Jiang et al. 1998, Ni, Wang et al. 1998, Seternes, Johansen et al. 2002, New, Jiang et al. 2003). The interactions have been demonstrated in vitro after overexpression of MK5 and p38, but the in vivo value of these observations is still under some debate, as previously explained, and is not completely resolved (Shiryaev and Moens 2010). In resting cells, MK5 is predominantly located in the nucleus but is able to shuttle between the nucleus and the cytoplasm. MK5 contains functional nuclear export signals (NES) and nuclear localization signals (NLS), and the opposite action of these motifs may explain the nucleocytoplasmic shuttling of MK5. Activation of
the p38 pathway was shown to induce nuclear export of MK5. Both p38α and p38β were reported to control the distinct subcellular localization of MK5 (Seternes, Johansen et al. 2002, New, Jiang et al. 2003, Li, Zhang et al. 2008). Another goal of Paper II was therefore to study the molecular interactions between MK5 and p38α.

The MK2 structure in the MK2-p38α X-ray complex (PDB id: 2OZA) was used as a template for constructing the MK5 model of the MK5-p38α complex. The MK5 sequence consists of 473 amino acid residues, while MK2 consists of 400 amino acids, such that based on the template, we could not predict the entire MK5 structure. The MK5 model is therefore C-terminally truncated and consists of the amino acids Met7-Gly367. Previous experimental studies have shown that the region Asn356 to Ser373 (Gly373 in mouse) of MK5 is important for interacting with p38 and was termed the p38 docking site (Seternes, Johansen et al. 2002). This indicates that the model did not contain the entire experimentally verified p38 docking site. However, experimental studies have also shown that a C-terminal truncated form of MK5 consisting of residues 1 to 368 also binds p38α, although not as strong as the full length MK5 (Seternes, Johansen et al. 2002). The docking of p38α into MK5 was guided by the X-ray structure complex of MK2 and p38α. The MK5-p38α complex is depicted in Figure 15. The docked complex showed that not only Asn356-Gly367 in the C-terminal of the model is involved in p38α binding, but also other amino acids in the C-terminal region are also required. In addition, the complex showed that amino acids in the P-loop, activation segment, αH helix, and the regulatory phosphorylation region are also important for binding to p38α (figure 16). The complex showed that MK5 and p38α interact such that the MK5 N-lobe interacts with the p38α N-lobe, and the MK5 C-lobe interacts with p38α C-lobe. The docked complex indicates that the C-terminal segment which contains most of the experimentally verified p38α docking site and the regulatory phosphorylation region of MK5 interacts (as two arms) at opposite sides of the p38α C-lobe.
Figure 15: The model of MK5 (green) in complex with p38α (yellow), outlining the important structural elements of MK5 interacting with p38α. The Cα-traces of the proteins are shown. Adapted from Paper II (Lindin, Wuxiuer et al. 2014).
Figure 16: A close up of interacting amino acids forming the MK5-p38α interface. MK5 is in grey ribbon, while p38α is in yellow ribbon. Dashed lines indicate interactions between amino acids in MK5 and p38α. The percentages indicate the percentages of MD frames with the particular interacting atomic distance <3.5 Å. (A) the activation segment of MK5; (B) the C-terminal end of MK5; (C) the P-loop of MK5; (D) the αH helix of MK5; (E) the regulatory phosphorylation region of MK5. Adapted from Paper II (Lindin, Wuxiuer et al. 2014).

Calculations of the Electrostatic Potential Surfaces (EPS) indicated that the C-terminal segment with the p38 docking site and the regulatory phosphorylation region of MK5 are strongly electropositive and the other regions of MK5 facing p38α are mainly electropositive. However, the ATP binding pocket of MK5 located between the N-lobe and C-lobe is electronegative. The regions of MK5 facing the solvent are partially electronegative and partially electropositive. p38α has a solvent-exposed side that is mainly electronegative, while the side facing MK5 is mixed electronegative and electropositive with highly electronegative areas in the regions interacting with MK5, the P-loop, αH helix, regulatory phosphorylation site, and p38 docking domain. This observation suggest that the charge distribution on the surface of MK5 and p38α is important for orienting the kinases relative to each other for a proper binding and that the binding is initiated by electrostatic forces.

MD simulation with MK5 in complex with p38α revealed that the fluctuations of the flexible loops (including the P-loop) during MD were decreased compared with the simulations with MK5 alone and the inhibitor complexes. However, monitoring the atomic distance between Lys51 and Glu62 at the active site showed that this distance varied a great deal during the MD. A similar trend was also seen for the following atomic distances: Glu62-Phe170, Lys51-Ser33, Asp169-Ser33 and Asp169-Ile32. These atomic distances were very stable during the MD with the free MK5 model. Therefore, it
seems like p38α binding not only reduces the structural motions of flexible MK5 loops at the interaction surface, but also destabilizes intramolecular interactions between amino acids in the region of the active site, which may be important for MK5 activation and the binding of substrates to MK5.

In general, the MK5-p38α complex was structurally very stable throughout the 100 ns MD. During the entire MD, amino acids in the P-loop, activation segment, αH helix, the regulatory phosphorylation region and the C-terminal were forming stable interactions with p38α.

A combined molecular modeling and site-directed mutagenesis study suggested that Leu156 of p38α is important for directing the MK5-p38α complex to subcellular compartments (Li, Zhang et al. 2008). When Leu156 was mutated to Val, the complex was mainly localized in the cytoplasm, and it was suggested that Leu156 interacts with the NLS. Our model did not show any direct interaction between Leu156 of p38α and MK5. However, our model showed that mutating Leu156 into Val will have an effect on the 3D structure of p38α, such that the observed interaction of Glu160/Asp161 of p38α with the region around Lys350 of MK5 is disturbed. Lys350 is a part of the NES, and compared to the wild type p38α, the Leu156Val mutant may interact differently with the NES.

**MK5 inhibitors**

Virtual ligand screening (VLS) has become an important contributor in the identification of bioactive small molecules and in the hit to lead process, and several papers are describing such approaches (Schapira, Abagyan et al. 2003, Katritch, Byrd et al. 2007, Khan, Fuskevag et al. 2009, Sager, Orvoll et al. 2012, Gabrielsen, Kurczab et al. 2014). The lack of selective MK5 inhibitors inspired us to apply VLS in the identification of new promising compounds for MK5 inhibition (paper III).

VLS of bioactive small molecules by docking and scoring can only be carried out if a three-dimensional structural model is available for the target protein. The “golden standard” for running structure based VLS is normally to use a three dimensional x-ray crystal structures as target for the screening (Klebe 2006). Most often, the screening is performed against a single conformation of a rigid x-ray crystal structure, and that may represent a problem since ligand-protein binding is a dynamic process. In addition not all proteins have been resolved by X-ray or NMR, and the gap between resolved structures and known amino acid sequences is rapidly growing. In later years, homology models have had a rapidly increasing quality and have come forward as an alternative to crystal structures in the search for new compounds. The main reason for that is the growing number of available templates for modeling. Several studies have also proven that a homology model often

It is often a concern that it is difficult to achieve selectivity when targeting an inhibitor to the ATP pocket since this pocket is highly structurally conserved between the 518 kinases. It could however be achieved by utilizing the naturally occurring structural idiosyncrasies in the ATP-binding pocket (unique MKS amino acids are described under the discussion section homology models). In paper III we chose to look for possible ATP competitive inhibitors since the structure we based the homology model on is crystalized with an ATP competitive inhibitor. This made us more confident that our model had the right conformation for this type of search. A structure based approach for searching MK5 inhibitors other than type I must await the MK5 X-ray crystal structure or a very detailed homology model.

VLS of multiple ligands to the same protein site uses docking and scoring to generate a ranked list of compounds (Neves, Totrov et al. 2012). In paper III several promising compounds (see paper III for complete list) were identified by VLS, which all bound in the proposed ATP binding pocket of MK5. They all had hydrogen bonds to one or more of Glu103, Met105, and Glu109 of the hinge region, Gln26 and Ala30 in the P-loop, Lys 51 in β3 sheet, and Asp169 of the DGF motif. All these amino acids are important in the binding of ATP, and might give an indication on how the inhibitors block ATP interaction with the pocket. Some of the compounds bound mainly in the outer area (hinge region) of the ATP-binding pocket while others also utilized the inner area closer to Lys51 and Asp169. An example of the best binding pose is given in figure 17, and shows compound 3 interacting both with the hinge region and Lys51/Asp169.
Figure 17: Binding mode of compound 3 from VLS screen. This inhibitor docked nicely into the proposed ATP pocket and interacted with MK5 mainly through Met105 in the hinge region, Asp169 in the DGF motif and Lys 51 in β3-sheet. Adapted from Paper III.

Because of an incomplete understanding of the physics of the binding process the scoring function in VLS will never be completely accurate (Klebe 2006) and there is no guarantee that a high score will give high inhibition in an experimental approach. It will however work well enough for an initial selection of possible candidates for further experimental testing.

Based on availability and cost, ten of the identified compounds with VLS score higher than -31 were ordered and tested for potential inhibitor activity. The first screenings of the potential inhibitors at concentrations of 100 µM showed that all compounds decreased MK5 kinase activity. Compound 3, 4 and 9 however showed good potential inhibition with only 9-11% kinase activity remaining. Comparing the binding mode from VLS with ability to inhibit it seemed like compounds predicted to bind deeply into the ATP pocket gave a better inhibition than the other compounds. Compound 7 and 6 for example were predicted only to bind in the outer hinge region and had an inhibition at 100 µM of only 29% and 11%, respectively. Compound 3, 4, 5, and 9 which all protruded into the ATP pocket had an inhibition more than 70%. Comparing the score value with actual inhibition we also observed that the best scoring compound actually was the compound with the lowest inhibition (compound 6). This demonstrates that although score can be used to predict binding it doesn’t tell us how well the binding performs in the actual in vitro experiment.

One of the best compound (compound 3) was subjected to a dose-response study and the IC₅₀ value was determined be 22.7 µM. This was slightly better than the previously presented MK5 inhibitor,
Noroxoaconitine, from our group (Kostenko, Khan et al. 2011) and the inhibitor SFV785 from Anwar et al. (Anwar, Hosoya et al. 2011) which have an IC\textsubscript{50} of 37.5 \mu M and 50 \mu M respectively. Compared to other published MK5 inhibitors, like Epigallocatechin gallate (EGCG), Flavokavain A and B, Immidazopyrazine derivatives, and GLPG0259, which all have an IC\textsubscript{50} values in the nM range (Bain, McLauchlan et al. 2003, Folmer, Blasius et al. 2006, Andrews, Clase et al. 2012, Namour, Vanhoutte et al. 2012, Westhovens, Keyser et al. 2013) our compound performed poorly. However all of the above mentioned inhibitors are tested in large kinase screening assays and none of them has been reported to be exclusively selective for MK5. Additional studies are required to determine the selectivity of compound 3 towards other protein kinases, as well as the stability, activity and toxicity in cell culture of this compound. However, this inhibitor might be a good starting point for synthesizing modified analogues with improved potency and specificity and may contribute to further unraveling the functions of MK5.
Concluding remarks

Even though the biological function of MK5 still remains elusive, it will be interesting to see how its role as a possible drug target may be elucidated in the near future.

The 3FHR-based MK5 homology model presented in Paper I was effective in discriminating between known inhibitors and decoys. This indicates that the homology model may be used as a working tool for further experimental studies and possibly structure aided drug design.

Paper II contributes with increased insight into the structure of MK5 and its interactions with molecular partners. The information can be used in the design of selective compounds interfering with MK5 activity, which would be useful in the elucidation of the exact biological role of MK5.

In Paper III the MK5 homology model from Paper I was used in a virtual ligand screening in a search for new possible MK5 inhibitors and several promising compounds were identified. Even though these compounds were not stronger binders than previously published inhibitors, they may represent a unique opportunity for further studies synthesizing structurally modified analogues with improved potency and specificity and may contribute to further unraveling the functions of MK5.

In the future it will be interesting to further test compound 3 from Paper III for selectivity against other kinases. If the crystal structure of ERK3 or ERK4 is solved it would also be interesting to use molecular dynamic simulations to study the interaction between MK5 and ERK3/4 in the same way as done in Paper II for p38α. The ultimate goal would of course be to solve the crystal structure of MK5 itself.
References


Raviv, Z., E. Kalie and R. Seger (2004). "MEK5 and ERK5 are localized in the nuclei of resting as well as stimulated cells, while MEKK2 translocates from the cytosol to the nucleus upon stimulation." Journal of cell science 117(Pt 9): 1773-1784.


Paper I
Paper II
Paper III