Part B-1

1. Excellence

1.1 Quality and pertinence of the project's research and innovation objectives

Introduction - The 5' adenosine monophosphate-activated protein kinase (AMPK) is a master regulator of cell energy metabolism[1,2]. It provides stable levels of adenosine triphosphate (ATP), the universal and immediate energy carrier in cell chemistry. AMPK functions as a signaling hub through up- and downstream phosphorylation and dephosphorylation, therefore influencing multiple pathways of cellular energy budgeting. AMPK (dys)regulation has been connected to metabolomic diseases (e.g., obesity, inflammation, diabetes), heart conditions and viral infections[3]. Controlled pharmaceutical intervention can improve diseased states through (de)activation of AMPK, making it a drug target of interest [4].

AMPK is fine-tuned through its structural features, which are dependent on mRNA isoform splicing, protein complexation, post-translational modification (PTM) and allosteric activation[5]. The heterotrimeric AMPK complex (α-β-γ) exists in multiple isoforms, which are encoded by separate genes (α1-2, β-2, γ1-3), whose expression is highly dependent on tissue and environment. AMPK appears to be activated (x1000) by phosphorylation of α1 at Thr^{172} (P-Thr^{172}) and corresponding sites of α2. The other PTMs finetune the extent of P-Thr^{172}. Therefore, AMPK's diverse regulatory role is deeply connected to its post-translational modifications (43 known phosphorylation and 36 known other PTM sites). The result of this structural diversity is an extremely large combinatorial space of (1) unique covalently modified α-β-γ isoforms, the proteoforms, and (2) their unique non-covalent assemblies, the complex species. However, distinct proteoforms and complex species prevail, channeled by the functional association between protein PTMs, the so-called PTM-code. It is clear that AMPK (dys)regulation can only be understood in the context of its structural features. However, the diversity of AMPK is not well understood. The goal of this project is to study in detail the PTM-code and endogenous diversity of AMPK proteoforms and complex species.

State of the art - To date, bottom-up MS approaches have been widely used to study PTMs. For AMPK, signaling cascades (e.g. phospho-proteomics), protein-protein interaction (e.g. proximity labelling) and tissue or organism specific PTMs have been identified. Due to the peptide-to-protein inference problem, bottom-up approaches cannot derive isoform specificity, distribution, or coded interplay[6]. Also, enzymatic modifications of AMPK have been studied *in vitro* focusing on AMPK α- and/or β-, but not on the γ-subunits and/or on the trimeric complex. Moreover, inhibitor/activator, nucleotide binding, or site specific PTMs have been revealed through structural studies with X-ray crystallography (XRC) or cryo-electron microscopy (Cryo EM). However, these studies have been based on averaged electron maps and only show static structures that are blind to PTM subsets[7]. Despite AMPK being a highly researched protein, the current state-of-the-art lacks either the unique assignment of proteoforms and/or addressing of structural heterogeneity.

Analysis using top-down mass spectrometry (TDMS) will overcome these technological limitations[8]. Denatured top-down mass spectrometry (dTDMS) can decipher proteoforms and native mass spectroscopy (nMS) can identify complex species and their subunits. Recently, these methods have merged into native TDMS (nTDMS) offering a holistic revelation of complex, subunit and sequence information. nTDMS is driven by cutting-edge technology, such as Fourier-transform ion cyclotron resonance (FT-ICR) MS instruments, which allow isotopic resolution of medium- and large-sized proteins (~30-100 kDa)[9]. nTDMS is also driven by ion mobility (IM) technologies such as trapped ion mobility spectrometry Time-of-Flight (tims-ToF), which allow multidimensional gas phase separation[10]. Using these technologies to study the PTM-code and endogenous diversity of AMPK, this project will go beyond the state-of-the-art.

Specifically, insight into these proteoforms will be achieved with four objectives:

---

[1] S. Herzig and R.J. Shaw, "Ampk: Guardian of Metabolism and Mitochondrial Homeostasis," *Nature Reviews Molecular Cell Biology* 19, no. 2 (2018).

[2] A.J. Ovens et al., "Post-Translational Modifications of the Energy Guardian Amp-Activated Protein Kinase," *International Journal of Molecular Sciences* 22, no. 3 (2021).

[3] M.S. Bhutta, E.S. Gallo, and R. Borenstein, "Multifaceted Role of Ampk in Viral Infections," *Cells* 10, no. 5 (2021).

[4] G.R. Steinberg and D. Carling, "Amp-Activated Protein Kinase: The Current Landscape for Drug Development," *Nature Reviews Drug Discovery* 18, no. 7 (2019).

[5] Y. Yan et al., "Structure and Physiological Regulation of Ampk," *International journal of molecular sciences* 19, no. 11 (2018).

[6] F. Lermyte et al., "Top or Middle? Up or Down? Toward a Standard Lexicon for Protein Top-Down and Allied Mass Spectrometry Approaches," *J Am Soc Mass Spectrom* 30, no. 7 (2019).

[7] Y. Yan et al., "Structure of an Ampk Complex in an Inactive, Atp-Bound State," *Science* 373, no. 6553 (2021).

[8] J.A. Melby et al., "Novel Strategies to Address the Challenges in Top-Down Proteomics," *J Am Soc Mass Spectrom* 32, no. 6 (2021).

[9] Y. Ge et al., "Top-Down High-Resolution Mass Spectrometry of Cardiac Myosin Binding Protein C Revealed That Truncation Alters Protein Phosphorylation State," *Proc. Natl. Acad. Sci. U.S.A.* 106, no. 31 (2009).

[10] H. Li et al., "An Integrated Native Mass Spectrometry and Top-Down Proteomics Method That Connects Sequence to Structure and Function of Macromolecular Complexes," *Nature chemistry* 10, no. 2 (2018).

Objective 1 (O1): Establish model protein and MS approach. I will recombinantly produce and purify the AMPK complex (O1.1), which unmodified, is ideal for optimizing MS approaches and collecting data. This model protein will then be used to optimize sample delivery and acquisition parameters of dTDMS and nTDMS (O1.2). The collected data will serve as an efficient pipeline of data analysis (O1.3).

Objective 2 (O2): Study PTM-code using* in vitro *AMPK complex modifications. I will alter recombinant AMPK *in vitro* to gain insight into its PTM-code. Firstly, I will modify AMPK with upstream kinases to identify their individual AMPK proteoform footprints (O2.1). Secondly, I will modify AMPK with two or more enzymes to determine if non-expected proteoforms are produced (O2.2). I will also modify AMPK with crude cell lysates to reveal more physiological proteoforms (O2.3). The results will expose PTM-code and footprint maps and thereby, insights into how to design specific AMPK proteoforms. Additionally, I aim to use these results to design recombinant endogenous-like AMPK proteoforms as sample for the Non-academic phase (O2.4).

Objective 3 (O3): Establish a purification strategy for endogenous AMPK. I will access endogenous AMPK from heart tissue. Therefore, building on the Ge lab's expertise in isolating cardiac proteoforms and advances in TDMS sample preparation in denatured and native mode, such as new affinity and chromatography materials. Two generally different approaches, or in combination, will be tested: Enrichment of AMPK using affinity reagents (O3.1) and separation of AMPK using off- and online chromatography (O3.2). The results will reveal, for the first time, the endogenous diversity of AMPK in the heart and a strategy to study tissue or disease related states.

Objective 4 (O4): Map tissue-specific AMPK proteoforms. I will establish a sample prep pipeline at USIEGEN (O1.1) and collect tissue samples and their metadata (O4.2), including subtypes of diseased or virus-infected tissues. Then, established purification strategy will be used to characterize AMPK from the collected samples (O4.3). The results will capture the molecular makeup of AMPK in different tissue and/or disease related states, which will allow determining of archetypes for proteoform design.

Overall, deciphering and mapping AMPK proteoforms in their complexity and entirety through top-down mass spectrometry (TDMS) will further our understanding of genetic protein products in their entire diversity, including their function within their endogenous biological system[11]. This understanding is important for the identification of AMPK biomarkers and, ultimately, for the advancement of precision medicine, where treatment is tailored for the individual molecular setup of a patient subgroup such as in metabolomic or heart disease.

1.2 Soundness of the proposed methodology

Short overview - Outgoing phase (OP). .... ***Return phase (RP).*** ... ***Non-academic phase (NAP).*** Bruker Daltonics GmbH (M37-42): ...

Interdisciplinary aspect - Mass spectrometry in biomedical research is interdisciplinary by nature, through the combination of problem (medicine), sample preparation (chemistry) and the analysis (physics). The outlined objectives require understanding in or partial aspects of physical biology, biochemistry, material science, data science and medicine with the goal to build a molecular foundation for precision medicine.

Mass spectrometry - Two different approaches will be used, dTDMS and nTDMS (Figure 1)[12]. ***dTDMS.*** Denaturing complexes into single αβγ-subunit chains will result in isolated subunits, which provides detailed sequence information. For denatured sample delivery, online RPLC-MS with organic acidified solvent is employed. For proteoform determination using dTDMS, the general workflow is to first acquire, MS1 spectra, to have an overview of the proteoforms molecular weight, and then, MS2 spectra, to map the proteoforms sequence by dTDMS sequencing. ***nTDMS.*** The goal native analysis is to identify the interaction of specific proteoform subunits. The workflow encompasses spectral overview in MS1, subunit dissociation in MS2 and subunit (proteoform) sequencing in MS3. In MS1, a precursor of interest will be selected and isolated. In MS2, dissociation spectra will be acquired by applying increasing collisional energy. In MS3, fragment spectra will be acquired for unambiguous proteoform assignment directly from the complex. In general, nMS requires solutions with ionic strength and near physiological pH such as the volatile buffer-surrogate ammonium acetate (AmAc). In nMS, the common way is a manual buffer exchange of sample and direct injection into the MS via nano-electrospray ionization (nESI). Native LC-MS is still not widely used due to incompatibility of LC materials with nMS buffer. However, the Ge lab introduced hydrophobic interaction chromatography (HIC) solid phases compatible with AmAc[13]. This system will allow me to combine separation, purification, buffer exchange and sample injection. ***Top-down sequencing.*** Fragmentation will be based on gas phase collisions (CID/HCD) and electrons (ETD/ECD). Combining results will cover more sequence space than from a single fragmentation technique[8]. *In the return phase,* I can complement the sequence information with ultra-informative ultraviolet photodissociation (UVPD). ***MS instrumentation.*** In Madison and at Bruker, I will have access to a high-resolution Bruker SolariX FTICR-MS enabling me to resolve an expected heterogenous MS1

[11] J. Zhong et al., "Proteoform Characterization Based on Top-Down Mass Spectrometry," *Brief Bioinform* 22, no. 2 (2021).

[12] D.P. Donnelly et al., "Best Practices and Benchmarks for Intact Protein Analysis for Top-Down Mass Spectrometry," *Nat. Methods* 16, no. 7 (2019).

[13] B. Chen et al., "Online Hydrophobic Interaction Chromatography–Mass Spectrometry for Top-Down Proteomics," *Anal. Chem.* 88, no. 3 (2016).

signal pattern[14]. The Bruker MaxisII is the standard instrument for TDMS[12]. Moreover, I will use Bruker tims-ToF for nTDMS workflows, which facilitates for another dimension of separation through reduced ion mobility in the gas phase. The tims-TOF has been originally developed for bottom-up proteomics, but the Ge lab successfully used it for nTDMS applications[15]. At Bruker, I will further tune or connect *tims* devices in *tims-tims* mode, with the goal to separate proteoforms with similar molecular weight. *In the return phase,* the go to instrument for nMS is the Thermo Scientific Ultra-high mass resolution (UHMR)-Q-Exactive. Upon return, the UHMR will be modified with ion mobility and UVPD (ERC-2017-STG, SPOCkS MS), which will allow me to collect complimentary nTDMS datasets. Additionally, a tims-Tof is available and a new cutting-edge TDMS instrument will be purchased in the course of the Uetrecht lab's relocation to USIEGEN (in 2022).

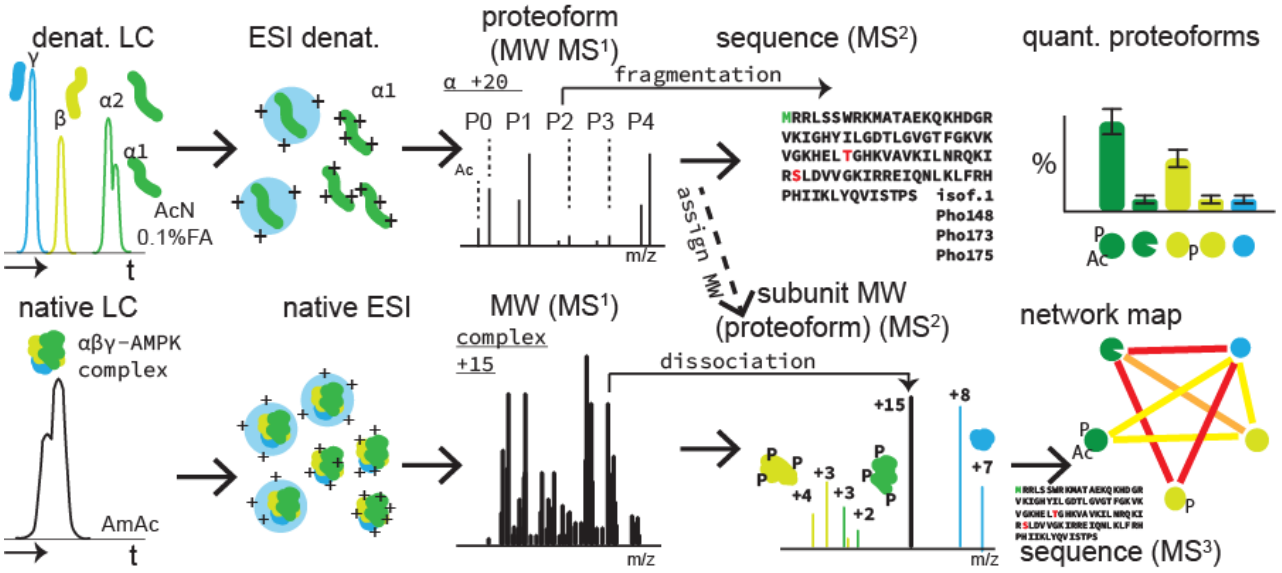

Figure 1: Workflow of dTDMS (top) and nTDMS (bottom) approach.

Recombinant AMPK - Production of recombinant AMPK complexes has been described in detail[16,17]. The gene construct is based on bacterial expression of αβγ subunits from separated open reading frames. The plasmid will be obtained from a collaboration partner from the Ge lab. The α-subunit is fused N-terminally to a polyhistidine affinity tag, while the β- and γ-subunits are not tagged. This allows for a purification based on affinity chromatography and size exclusion chromatography. Isoforms will be created using site-specific deletion mutations. I have extensive experience in protein production and many different groups have reproduced a similar protocol[7]. Thus, no major difficulties are expected. Acquiring mass spectra of recombinant AMPK is the first milestone in this project (Ms1).

In vitro *modification* - PTM decoration will be performed in levels of rising complexity and guided by existing knowledge from RNAseq about isoforms and Bottom-up MS about PTMs. For *in vitro* phosphorylation, I will incubate AMPK complex, ATP and commercially available kinases (e.g. LKB1, CaMKK2, PKA, DNA-PK, available at Merck Millopore GmbH or Promega Cooperation), identified as upstream modifiers of AMPK. For footprints, single kinases will be used in sub-stoichiometric concentration to prevent over-phosphorylation. For functional influence between PTMs two or more kinases will be incubated with AMPK. It will be considered that the enzymes sequential order and concentration can play a role. Functional influence is suggested, when other proteoforms appear than expected from their single congruent footprint results. The second milestone (Ms2) of this project is to collect data on AMPK incubated with crude lysates (HeLa cells and HEK cells) to receive a physiological picture of PTM decoration. During the return phase, modified sample will be produced for the Non-academic phase.

Purification of endogenous AMPK - Tissue samples. Work with and the documentation of human tissues holds up to the highest ethical standards, have ethics approvals from local authorities and is legally allowed to be performed in Germany (A4.3/Ethics self-assessment). In general, the dynamic concentration and complexity of the proteome makes analysis of low abundant endogenous proteoforms such as AMPK extremely difficult. However, cardiac tissue is an ideal sample to start with due to a relatively high AMPK muscle concentration[18]. I will use two generally

[14] T. Tucholski et al., "A Top-Down Proteomics Platform Coupling Serial Size Exclusion Chromatography and Fourier Transform Ion Cyclotron Resonance Mass Spectrometry," *Analytical chemistry* 91, no. 6 (2019).

[15] E.J. Larson et al., "High-Throughput Multi-Attribute Analysis of Antibody-Drug Conjugates Enabled by Trapped Ion Mobility Spectrometry and Top-Down Mass Spectrometry," *Anal. Chem.* 93, no. 29 (2021).

[16] D. Neumann et al., "Mammalian Amp-Activated Protein Kinase: Functional, Heterotrimeric Complexes by Co-Expression of Subunits in Escherichia Coli," *Protein Expression Purif.* 30, no. 2 (2003).

[17] F. Rajamohan et al., "Escherichia Coli Expression, Purification and Characterization of Functional Full-Length Recombinant A2β2γ3 Heterotrimeric Complex of Human Amp-Activated Protein Kinase," ibid.73 (2010).

[18] D. Garcia and R.J. Shaw, "Ampk: Mechanisms of Cellular Energy Sensing and Restoration of Metabolic Balance," *Molecular cell* 66, no. 6 (2017).

different approaches, or in combination. ***Enrichment.*** AMPK specific antibodies (Ab) will be coupled to magnetic beads via their A and G domain. A variety of well characterized Ab can be purchased commercially. Particularly important, is that the Ab, such as pan-α- and pan-β-AMPK-Ab, targets all isoforms of AMPK. For native AMPK elution, buffer conditions have to be carefully optimized, using low-pH high-salt buffer and/or an excess of Ab antigenic peptides. Alternatively, I will cooperate with a material chemist from the Ge lab, who currently customizes AMPK binding nanoparticles (NP)[19]. ***Chromatography.*** The Ge lab developed multiple LC methods that supported the advance of dTDMS sample delivery into an high throughput method, for example online serial size exclusion chromatography (sSEC) identifying in one single run hundreds of proteins and thousands of proteoforms from tissue lysates[14]. The separation of protein lysates for nTDMS is currently a highly researched topic. I will approach this by using the online native HIC/IEX systems, described above, and for a more efficient pre-purification offline multi-dimensional LC (MDLC) workflows combining two or more orthogonal separation principles (e.g. polarity, ionic strength, size). For benchmarking and troubleshooting of LC methods, the recombinant AMPK will be used. Nevertheless, purification of endogenous AMPK needs to be newly established, and my experience in protein purification and the access to a wealth of technology enable me to do so. The third milestone (Ms3) of this project is the data collection of endogenous AMPK proteoforms from the heart.

Map tissue-specific proteoforms - Tissue samples have to be collected within the established network of Charlotte Uetrecht (see 1.3). Ethic approvals have to be in place and metadata needs to be recorded in compliance with the highest standards (A4.3/Ethics self-assessment) (Milestone 4). AMPK tissue specificity will be estimated from its isoform expression[20]. The type of sample will be chosen depending on the previous results, feasibility and availability. The fifth milestone (Ms5) of this project is to generate an AMPK proteoform tissue atlas comprising of brain, liver, lung, muscles and blood. However, if the atlas not feasible, one type of tissue will be investigated exhibiting different conditions, such as virus infection or metabolomic disease. Some tissue types will not be complementary to the established purification strategy, and establishing another strategy will go beyond the scope of this project. One task of this objective is to transfer and implement the online HIC/IEC and sSEC technologies at USIEGEN.

Data processing, analysis and visualization - The data pipeline will be an important development of the project. Sequencing in nTDMS is less informative, therefore, subunits ejected in nTDMS need to be mass matched with proteoforms as identified in dTDMS (Figure 1). Furthermore, quantitation will be key to reveal PTM-code and preferred subunit assembly of specific proteoforms. For proteoform quantitation from dTDMS MS1 spectra, established pipelines will be applied. Information about qualitative and quantitative subunit interactivity is provided in nTDMS MS2 spectra. The goal is to visualize these interactions in a network-based map. I will work together with the Ge lab's bioinformatics team, which developed *MASH Suite Pro and MASH Explorer* including solutions for the above mentioned steps in data analysis[21]. Furthermore, corresponding training activities are planned (T4-6).

Open science - Data management will be guided by FAIR principles (Findable, Accessible, Interoperable, Reusable). Relevant data will be uploaded to the *Top Down Proteomics Repository* (CC BY 3.0 DE, open access)[22]. Vendor specific mass spectra will be converted the openXMSL format. Metadata will be published according to community guidelines, e.g., *Standard Proteoform Notation Pro Forma*. These measures empower others to mine and reveal hidden potential of my data. Manuscripts will be released on open preprints servers (e.g. BioRxiv). The work will be recorded with high integrity ensuring experiment reproducibility. Communication and dissemination measures will consider citizen science and public engagement. Open science training activities are planned (T1).

Consideration of the gender dimension - Gender difference in AMPK regulation has been suggested[23]. In principle, the planned methodology can be used to characterize specific differences in molecular AMPK regulation between patient subgroups, such as male and female. However, this question would be too specific for the scope of this project. As an outlook, a detailed characterization between AMPK regulation in male and female patients could be exploited for better healthcare in a precision medicine approach.

MSCA Green Charter - I aim to implement measures to minimize the environmental footprint of my activities. I will choose green and sustainable laboratory materials, such as green pipette tips, meetings will be largely held online, omitting unnecessary travel.

1.3 Quality of the supervision, training and of the two-way transfer of knowledge between the researcher and the host

Supervision - During the OP...

[19] T.N. Tiambeng et al., "Nanoproteomics Enables Proteoform-Resolved Analysis of Low-Abundance Proteins in Human Serum," *Nat. Commun.* 11, no. 1 (2020).

[20] Will be retireved from : https://www.proteinatlas.org/

[21] Z. Wu et al., "Mash Explorer: A Universal Software Environment for Top-Down Proteomics," *Journal of proteome research* 19, no. 9 (2020).

[22] http://repository.topdownproteomics.org/ (New launch announced)

[23] K.D. Brown et al., "Sex Differences in Cardiac Amp-Activated Protein Kinase Following Exhaustive Exercise," *Sports medicine international open* 4, no. 1 (2020).

During the RP...

During NAP, ...

Joint supervision – Each supervisor brings his or her unique focus of expertise...

Transfer of knowledge – ***Researcher will gain. ...Partners will gain.*** ...

Table 1: Planned training activities.

#	Course/Training	Month	Phase	Kind of training
T1				
T2				
T3				
T4				
T5				
T6				
T7				
T8				
T9				
T10				
T11				

1.4 Quality and appropriateness of the researcher's professional experience, competences and skills

Research.

Experience.

This project.

2. Impact

2.1 Credibility of the measures to enhance the career perspectives and employability of the researcher and contribution to his/her skills development

dTDMS/nTDMS is expected to contribute ...My long-term career goals...

Altogether, this MSCA will provide my growth from a scientist...into a highly skilled ...expert with a profile in .... The ... will make me highly employable for research institutes as well as in international cooperations.

2.2 Suitability and quality of the measures to maximise expected outcomes and impacts, as set out in the dissemination and exploitation plan, including communication activities

Data management **-** Upon starting the project, a data management plan will be delivered (D2), guided by open science practice and the FAIR principles (Findable, Accessible, Interoperable, Reusable) (B1.2). Upon release of pre-printed manuscripts, data will be published in an open database (D10, D12, D17, D18).

Commercial exploitation **-** ...

Dissemination **-** ...***The global MS community*** will be interested in the results for their ... ***The research communities*** in structural and molecular biology can use my results for ... ***Clinicians and stakeholders*** in European health politics will be interested in ... ***For pharma industry,*** the results could be used for ...

Communication **-** For general press releases, ...**One-way-communication** in media will target taxpayers to inform about local/European scientific endeavors... **Two-way communication** with the public will be approached during ...

Table 2: Planned dissemination (Dis) and communication (Com) activities. D=Deliverables *expected regular dates/not scheduled yet

C/D	Action	Month	Medium	Target comm.
Com				
Com				
Dis				
Dis				
Com				
Dis				
Com				
Com				
Dis				
Dis				
Dis				
Dis				
Com				
Dis				
Dis				
Com				

2.3 The magnitude and importance of the project's contribution to the expected scientific, societal and economic impacts

Host institute **-** ...

Societal **-** ...

3. Quality and Efficiency of the Implementation

3.1 Quality and effectiveness of the work plan, assessment of risks and appropriateness of the effort assigned to work packages

Following here, the work plan is summarized (Table 3), visualized (Table 4) and its risks are assessed (Table 5).

Table 3: Overview of work plan. D = Deliverables, Ms = Milestones M=Project month, PM=Estimated person months

Table 4: Gantt chart according to work plan.

Table 5: Risk assessment.

Risks	Mitigating measures

3.2 Quality and capacity of the host institutions and participating organisations, including hosting arrangements.

At the UW-Madison, ...

In the RP, ...

In the NAP, ...

---------- End of page count (max 10 pages) ----------