Opponent's review of bachelor thesis

Author: **David Pech**

Title: **Determination of amino acid sequence of hemelipoglyco protein from tick *Dermacentor marginatus* by mass spectrometry**

The submitted bachelor thesis studies the primary sequence of a hemelipoglycoprotein from the tick *Dermacentor marginatus* using tandem mass spectrometry with electrospray ionization and compares it to the primary sequence of hemelipoglycoprotein and vitellogenin protein from *Dermacentor variabilis*.

I have the following comments/questions to the work:

- p. 4, l. 20: mass spectrometry identifies the analyzed compounds based on their mass only in seldom cases, because the mass of a compound is often not sufficient information for the identification
- p. 6, l. 5: not "according to their masses", but according to their m/z ratios
- p. 8, l. 4-5: in the reflectron, not the faster ions penetrate deeper, but those with higher kinetic energy
- p. 14, l. 4: what was the concentration of ammonium bicarbonate buffer?
- p. 14, l. 13-14: what is understood under "both washes" if elution was done 10 x 20μl?
- p. 15, l. 7: what is the version of SwissProt database used here?
- p. 16, l. 3-4: was the de-novo sequencing done based solely on the MSMS data or was the similarity to HLGP from *D. variabilis* also used for the sequence interpretation?
- p. 16, l. 7-10: how do you know the number of amino acids in HLGP 1 subunit (1546) and HLGP 2 subunit (1492) in *D.marginatus*, especially when these subunits in *D. variabilis* consist of about 880 amino acids (Fig. 3.1 and 3.2)
- chapter "Results" and "Appendix A.1", Fig. A.1-A.12: the interpreted sequences (with the exception of sequence GALEQPFAAE, resp. HLNGALEQPFAAEVYQ) are shown with their preceding and following amino acids, which cannot be interpreted from MSMS data and is not known information from not sequenced HLGO from *D. marginatus*. Even if it is known, the preceding and following amino acids should be denoted, e.g. (R)YVLPLWETNPR(F)
- in all interpreted sequences, every interpretation of "leucine" amino acid residue is ambiguous, because low energy CID analysis is not generally able to distinguish between leucine and isoleucine amino acid residue
- p. 20, l. 22: interpreted sequence "RLV" might not be correct, because arginin residue was probably ambiguously derived only from trypsin cleavage specificity and might not be correct (e.g. lysine residue could be there located as well)
- p. 22, l. 4-6 and p. 24, l. 30-35: interpreted sequences GALEQPFAAE, resp. the whole interpreted sequence HLNGALEQPFAAEVYQ does not seem to be correct. Especially, the last four amino acid residue sequence EVYQ is in mass just 0.32 Da different from the sequence KFDE that is present in the HLGP 2 subunit of *D. variabilis* following the alignment of GALEQPFAA sequence. The KFDE interpretation is more likely also because of the V8 enzyme specificity (cleavage after glutamic acid). As it is concluded in the last paragraph in the page 24, the mutation K → E is not likely and both interpreted and here suggested sequence would need additional analysis for reliable confirmation.
- p. 23, l. 22 and next: what is the "theoretical maximum coverage" and how was it computed?

- p. 24, l. 7 and next: was the isolated protein (i.e. its possible disulfide bridges) reduced and alkylated prior to the enzymatic digestion?
- Appendix A.1, Figures: is it common to detect z-ions in CID mass spectra of peptides measured using ESI-Q-TOF
- Fig. A.8: uncertainty in sequence determination – NL or LN
- Fig. A.11: uncertainty in sequence determination – SG or GS
- Fig. A.12: uncertainty in sequence determination – GL or LG
- Fig. A.13: uncertainty in sequence determination – HL or LH and in the part EVYQ
- Fig. A.14-A.20: source organism for proteins analyzed/described is missing
- Formal aspects of the work (many unnecessary typos can be avoided using spellchecker!):
  - in the title and elsewhere: "hemelipoglyco protein" → "hemelipoglycoprotein"
  - p. 3, l. 17: "Britis" → "British"
  - p. 3, l. 24: "ot" → "of"
  - p. 4, l. 29: "eg." → "e.g.", "Electro Spray" → "Electrospray"
  - p. 5, l. 15: "acetonitrille" → "acetonitrile"
  - p. 5, l. 16: "durin" → "during"
  - p. 8, l. 4: "flypath" → "flight path"
  - p. 8, l. 13: "co" → "to"
  - p. 9, l. 6: "paralell" → "parallel"
  - p. 10, l. 10: "Desorbtion" → "Desorption"
  - p. 10, l. 12: "lenght" → "length"
  - p. 11, l. 4: "case." → "case,"
  - p. 12, l. 2: "fragment" → "fragment."
  - p. 13, l. 10-11: better description of "V8" compound and "MiliQ H_2O" (should read "MilliQ")
  - p. , l. : "identidy" → "identity"
  - p. , l. : "subunit" → "subunits"
- References:
  - formatting of author full names is not uniform, first names start with surname, second and other names start with an abbreviation of author's name
  - Dupejová, J.(2008a), Dupejová, J.(2008b): "frep" → "FREP"
  - Maya-Monteiro, C.M. (2004), Maya-Monteiro, C.M. (2000): "Help" → "HeLp"
  - Paul, W (1953): "Ein neues massenspektrometer ohne magnetfeld" → "Ein neues Massenspektrometer ohne Magnetfeld"

The work fulfills the aims given by the topic of the thesis. Both content and formal point of the work is done properly (with respect to the above mentioned comments). Author showed an active approach to the work proved also by the results presented in the work and that is why **I recommend the thesis to be accepted for its defence** and I suggest the classification **B** (**= Excellent minus**).

Hradec Králové, 11th of June, 2010

Pavel Řehulka ……………………………………

RNDr. Pavel Řehulka, Ph.D.
Institute of Molecular Pathology FMHS UD
Třebešská 1575
CZ-500 01 Hradec Králové