Opponent's assessment of bachelor thesis

First name and surname of student: *Petra Gajdošová*

Thesis name: *Structural identification of protein-DNA interactions using machine learning*

A. Point assessment of individual aspects of the work (choose **only** one option)

1. Scale of BT and its structure	
✓	A - excellent
	B – very good - it is however not balanced – scale of some parts does not reflect their importance
	C – good, but some parts are not covered sufficiently
	N - unacceptable

2. Language level	
	A - excellent
✓	B – very good, only a few minor errors or typos found
	C - good, minor errors are common
	N – unacceptable, many serious mistakes and errors

3. Work with literature	
✓	A – excellent – citations are relevant and appear at relevant places
	B – very good – minor objection to a number or a placement of citations
	C – good, wrong format or missing citations do appear
	N – unacceptable – a lot of missing citations

4. Clarity of the work	
✓	A – excellent – the thoughts are well-formulated, it is a pleasure to read
	B- very good – there are a few clumsy expressions or unclear formulations
	C - good, clumsy expressions or unclear formulations are more common
	N – unacceptable, the the work is difficult to understand

5. Formal and graphical level of the work	
✓	A - excellent
	B – very good, but individual missing links to figures, shortcuts etc. appear,
	C - good, bigger errors like missing pages appear
	N – unacceptable, many serious errors

Please add any comment to the above (if you feel like doing so):
In this work, the author conducted a literature overview in the area of in-silico, protein-DNA

interaction prediction and concentrated on the analysis of 3 existing tools. Additionally, the author focused on a software application that was primarily designed for protein-ligand binding site prediction (P2Rank), and explored the efficiency of a modified version of the program, meant to predict DNA binding-sites on protein structures, by comparing it with the original version, as well as against the (3) aforementioned predictors. The author also developed scripts that may help expand the current set of descriptors of P2Rank to include electrostatic potential and dipole moment features.

The work is concise and has very good structure and flow throughout. The experiments are well justified and have the potential of adding value to the topic.

There are slight inconsistencies with the format of some citations (some names are full and others use initials, i.e. citation [38], or have inverted first and last names, i.e. citation [66]), but they are well chosen overall. A few repetitions of text can be found (i.e. section 2.1, page 3, section 3.2.2, page 13) and some syntax, spelling, and phrasing errors.

If the work contains results of the authors (it is not mandatory) please comment the following:
Are the aims of the work clearly stated? - *Yes.*
Is the amount of the experimental work equivalent to the aims? - *Yes.*
Are the results well documented? - *Yes.*
Are the results discussed with existing literature? - *Yes.*

Please add any other comment to the work of student:

B. Defense

Your questions to the author (please have at least one)
Taking into account the large number of programs developed towards the prediction of protein-DNA interactions, were there any particular reasons for choosing these three methods for analysis and comparison with P2Rank (DISPLAR, iDBS, DBSI)? *With respect to the (future) implementation of electrostatic potential descriptors into P2Rank to improve its prediction performance, which approach would you consider more practical and suitable between DelPhi and APBS? Are there any reasons you would choose one over the other?* *In the same context, how well would the dipole moment calculation compare to an electrostatic potential descriptor (in terms of the expected improvement on prediction quality)?* *(- i.e. Would the expected results be better, worse or similar?)*

C. Final assessment

I do recommend this work to be accepted: [✓ **Yes**] / **NO**

Suggested classification: A - *excellent*

Date: *11.9.2020*

Name and signature: *Christos Feidakis*

Instruction for filling and sending this form:

- Please use this form for your assessment.
- Please send the filled form to the following address: marian@natur.cuni.cz. We also need an original of signed document – either send it to the following address: sekretariát Katedry buněčné biologie PřF UK (p. Růžičková), Viničná 7, 128 44 Praha 2, or being it to the defence.
- Student shall get your assessment at least three days before the defence - you can send it yourself or we can do it for you.