

JCTC Letter

Dear reviewers, dear colleagues,

Please find a revised and condensed version of the manuscript which is cut down by 7 pages; in summary:

- the introduction has been fully re-written according to the comments provided by the reviewer 2 (1a, 1b, 1c, 6, 7);
- the results & discussion section has been shortened to retain only the main messages and removing the more technical details or complementary analyses which are now found in the methods and the supplementary material, respectively (reviewer 2: 1);
- the methods include some initially missing information (reviewer 1: 1, 2), and the references to the detailed analyses on the benchmark and MCSS performances are provided in the supplementary material (reviewer 2: 1);
- two figures have been modified (Figures 4 becoming Figure 1, and Figure 11), two figures have been moved to the supplemental information material (Figures 1 and 2), one supplementary figure (Supplementary Fig. 4) has been modified and included in the main body of the manuscript.

General statements:

The current study is a preliminary work towards the improvement of a fragment-based approach based on MCSS focusing on the major limiting factor of these approaches (screening power) in an attempt to apply it to single-stranded RNAs. The abstract was modified to emphasize the motivation for this work. The current state of the art is limited with only a few publications on the application of an FBD approach to the modeling of single-stranded RNAs which are actually oligonucleotides (de Beauchene et al., 2016a-b, introduction: page 4, last paragraph). As commented in the introduction, a method for linking nucleotidic fragments already exists. However, the method is not robust enough to model oligonucleotides without any sequence constraint due to the lack of accuracy of the scoring function (introduction, page 4, last paragraph).

On the other hand, MCSS is not strictly speaking a docking method or not only a docking method. It is developed in the framework of fragment-based approaches and includes fragment libraries that can be easily created and extended using the universal CHARMM force field (introduction, page 2, last paragraph to page 3, 2nd paragraph). We report for the first time the evaluation of the screening power for nucleotide fragments which is the more challenging step of the FBD process.

The issues listed by the reviewers are addressed in detail below.

Reviewer 1:

1. MCSS allows a partial treatment of the protein flexibility which is restricted to the side-chains. However, this functionality has not been used extensively because it results in computational expensive without any clear gain of performance in the predictions. The protein targets of the benchmark were

actually treated as rigid bodies as indicated clearly now in methods (section MCSS, page 10, 2nd paragraph).

2. Generalized Born models have been used before to rescore MCSS minima by including solvation terms as indicated in the introduction (page 3, 2nd paragraph: Haider et al., 2011; Haider et al., 2013). The results on GB models were not included initially with the results on the alternative scoring functions (Supplementary Figure 4) because of their low performance. The results with two GB models have now been included in the results and discussion (Figure 7, page 25) and described in methods (section Docking and screening powers, page 14, 2nd paragraph). The two GB models are based on the GBSW and GBMV implementations in CHARMM; they were tested using a rescoring scheme as the one used for the other four alternative scoring functions (Figure 7). Other implicit solvent models such as ACE and FACTS are currently implemented in MCSS and can be used as native scoring functions during the MCSS sampling and scoring iterations, i.e., with no rescoring scheme. ACE has been tested previously on protein-nucleotide complexes but not on the full benchmark. FACTS has not been tested extensively either.

3. The erroneous or missing units have been fixed (methods, section MCSS: page 9, 1st paragraph).

Reviewer 2:

1. The meaning of the MCSS acronym was added in the introduction (page 2, last line).

1a. The detailed information on the benchmark has been moved to the supporting information material which also includes all the raw data. Supplementary figure 1 was giving a graphical description of the benchmark, it is now a figure included in the main manuscript, (Figure 2 in methods, section Protein-nucleotide benchmark, page 8). The motivation for using a series of molecular features is basically for the sake of further optimization of the method, it is briefly mentioned at the end of the introduction (page 5, 2nd paragraph) and more in details in the beginning of the results and discussion (page 17, 1st paragraph), and at the end of the conclusions (last paragraph).

1b. MCSS is described on page 2 of the introduction (last paragraph) with mentions of its specificity in terms of sampling (page 3, 1st paragraph), its application within the FBD framework (page 3, 2nd paragraph), and its more specific applications on biopolymers (page 4, 1st paragraph).

1c. Figure 4 was modified and included in the introduction (page 5) to give a more visible view of the data generated in the study. The figures have been properly renumbered.

1d. Docking programs usually perform better when the ligand establishes many contacts with the target. The analysis of the X-ray structures gives a baseline about the intrinsic features that are more frequent and more likely to drive the interaction. The correlation between these features and the success rates in the predictions allows us to identify ways to optimize the method (see Conclusions, page 42, 2nd paragraph and page 44, last paragraph).

The minimization of the protein target is mandatory for scoring functions based on molecular mechanics. The disadvantage resides in the possible distortion of the protein structure. Some features have been specifically evaluated on the minimized structure to take into account the possible biases introduced by the distortion of the binding site.

Figure 11 (page 33) and other related supplementary figures (Supplementary Fig. 11-12 & 14) correspond to Upset plots which are alternative representations of Venn diagrams. These representations have become popular for large sets of intersecting data. This Figure has been modified to include only the plots showing the impacting features; the other part of the figure has been moved to the supplemental information material (Supplementary Fig. 11).

As pointed out by the reviewer, the dataset is relatively small. Only a qualitative analysis of the features is relevant. The choices of the threshold values are detailed in methods (section Molecular features, page 15). The significance criterion is also described.

2. The comparison between the different scoring functions is only qualitative for the same reasons given above. It is true that the differences between the scoring functions in terms of docking power are rather small and have only a qualitative value. However, the differences between the MCSS-SCAL and MCSS-STDW models are significant. The results and discussion explains in detail the reasons for the improved performance (page 24, 1st paragraph) which is summarized in the conclusions (page 43, 2nd paragraph).

4. The relationships between the current MCSS study and the FBD approaches is addressed in the introduction about the existing FBD approach applied to oligonucleotides (page 4, last paragraph to page 5, 1st paragraph). Designing oligonucleotides (without any sequence constraint) requires primarily to be able to predict the binding selectivity of different nucleotides which can be evaluated by the screening power. The more challenging step is to identify the more selective nucleotides along the binding site, i.e., to optimize the screening power. The abstract and conclusions have been modified in that sense.

5. The comparison with other scoring functions is provided in the manuscript with complementary data in the supporting information material. For more details, see the response to reviewer 1, second item.

6. The part on FBD approaches in the introduction has been reduced to its most minimal way. However, a few sentences on the assembling strategies have been conserved because it indicates precisely the specificities of FBD approaches that apply to biopolymers and thus applicable to oligonucleotides.

7. A paragraph has been added in the introduction (page 4, 2nd paragraph) about RNA therapeutics and the possible perspectives of applications in the field provided by FBD approaches applied to the design of oligonucleotides.

Editor:

The references and the supporting information available have been reformatted according to the ACS guidelines. The graphical abstract has been rescaled to fit into the box.