HIPPO: HIstogram-based Pseudo-POTential for scoring ssRNA-protein fragment-based docking poses

Anna Kravchenko, Sjoerd Jacob de Vries, Malika Smail-Tabbone, Isaure Chauvot de Beauchene

To cite this version:

Anna Kravchenko, Sjoerd Jacob de Vries, Malika Smail-Tabbone, Isaure Chauvot de Beauchene. HIPPO: HIstogram-based Pseudo-POTential for scoring ssRNA-protein fragment-based docking poses. The 31st Annual Intelligent Systems For Molecular Biology (ISMB) and the 22nd Annual European Conference on Computational Biology (ECCB), Jul 2023, Lyon, France. hal-04168414

HAL Id: hal-04168414
https://hal.science/hal-04168414
Submitted on 21 Jul 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
1 MOTIVATION

Single-stranded (ss) RNA-protein complexes are very challenging to model:
1. The absence of an unbound ssRNA structure due to its disorder prevents the use of classical docking methods,
2. ssRNA flexibility does not allow for exhaustive conformational sampling for long chains,
3. The relatively small number of experimental ssRNA-protein structures prevents the use of end-to-end deep learning for this problem.

4 PERFORMANCE

HIPPO application entails independent scoring of all docking poses with each of four 3C, followed by pooling together the top 5% of each scoring list.

<table>
<thead>
<tr>
<th>ASF</th>
<th>HIPPO</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of cases with 60% of more correct poses selected</td>
<td>26</td>
</tr>
<tr>
<td>% of cases with 80% of more correct poses selected</td>
<td>7</td>
</tr>
<tr>
<td>% of correct poses selected, averaged over all test cases</td>
<td>43</td>
</tr>
<tr>
<td>Nb of cases with 80% or more correct poses selected</td>
<td>15</td>
</tr>
<tr>
<td>Highest % of correct poses selected among the cases of a complex, averaged over all test cases</td>
<td>60</td>
</tr>
<tr>
<td>Nb of complexes with 60% or more correct poses selected for at least one fragment</td>
<td>54</td>
</tr>
<tr>
<td>Nb of complexes with 80% or more correct poses selected for at least one fragment</td>
<td>9</td>
</tr>
<tr>
<td>Nb of complexes with higher % of correct poses selected for the best-scored fragment</td>
<td>18</td>
</tr>
</tbody>
</table>

5 PERSPECTIVES

- Use in incremental modelling approach
- Build classifier to select best 3C per case
- Use as a sampling function
- Apply to all-atom models
- Apply to general protein-ssRNA benchmark
- Apply to protein-ssDNA benchmark

6 PIPELINE

1. dock the case (the same procedure for each case)
   - (1) dock the case
   - (2) zoom to on GC contacts
   - (3) count distance occurrences (a) derive histogram

2. {H}ype
d

3. {D}ata
c

4. {A}nalys
d

5. {R}esults
c
d

6. {S}uccess
c
d

**REFERENCES**


**ACKNOWLEDGEMENTS**

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 813289.