Supplementary Information

Evaluating DCA-based method performances for RNA contact prediction by a well-curated dataset

F. Pucci*, M. Zerihun*, E. K. Peter and A. Schug,
John von Neumann Institute for Computing, Forschungszentrum Jülich

- **Table 1**. Dataset of all RNA structures with their annotations (length, chain, PDB code, RFAM family, BIT-score, host organism)
- **Table 2**. Prediction results for all entries of the dataset and all methods tested (mean-field pydca, EVFold, Boltzmann learning, Gremlin, CCMPRED, PSICOV)
- **Table 3-5**. Impact of the search (Table 3), alignment (Table 4) and trim (Table 5) procedures on the mean-field DCA prediction performances for all entries of the dataset.
- **Table 6**. Impact of the contact definition on the performances of mean-field DCA for all entries in the dataset.

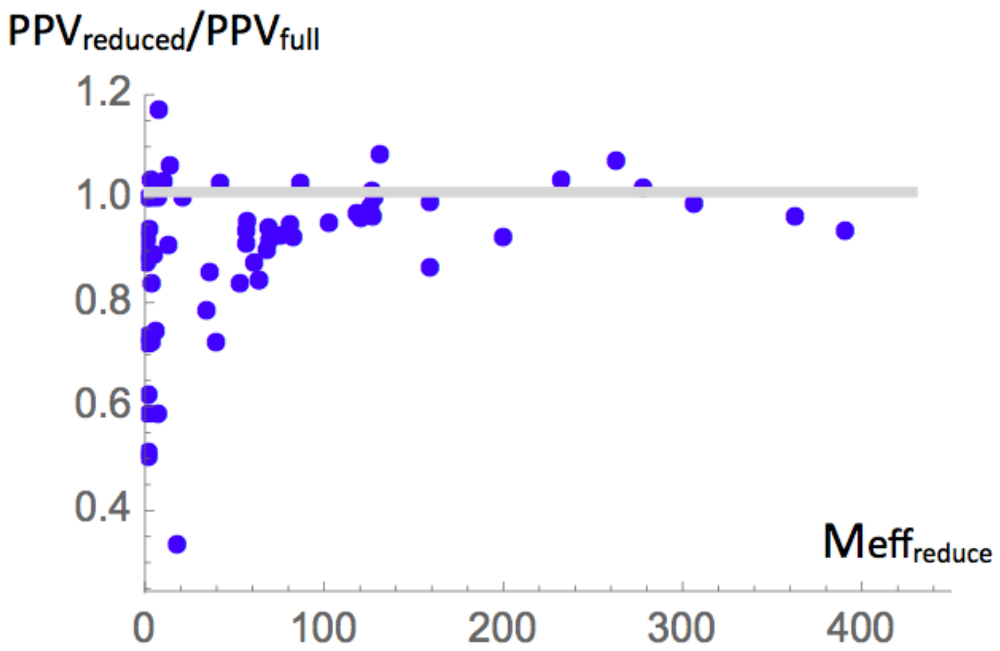

Figure 1: Ratio of the pydca PPVs computed for the full RFAM families (PPV_{full}) and for randomly chosen subsets composed of 1/3 of their entries ($\mathrm{PPV}_{reduced}$) as a function of the number of effective sequences in the reduced subsets.