The use of interatomic contact areas to quantify discrepancies between RNA 3D models and reference structures

SUPPLEMENTARY TABLE AND FIGURES

Kliment Olechnovič[1,2] & Česlovas Venclovas[1]

[1]*Institute of Biotechnology, Vilnius University, Graičiūno 8, Vilnius LT-02241, Lithuania*

[2]*Faculty of Mathematics and Informatics, Vilnius University, Naugarduko 24, Vilnius LT-03225, Lithuania*

Corresponding author:

Česlovas Venclovas
E-mail: venclovas@ibt.lt

Table S1. Correlation between major CAD-score variants and other scores on RNA puzzles and the *randstr* decoy set data

CAD-score variants[1]	INF[2]		DI[3]		RMSD[4]	
	Pearson[5]	Spearman[6]	Pearson[5]	Spearman[6]	Pearson[5]	Spearman[6]
	RNA puzzles (5 reference structures, 104 models)					
A-A	0.94	0.94	-0.90	-0.86	-0.89	-0.84
A-S	0.95	0.94	-0.89	-0.84	-0.87	-0.82
S-S	0.95	0.95	-0.88	-0.83	-0.86	-0.80
S-S stacking	0.96	0.95	-0.84	-0.82	-0.82	-0.79
S-S non-stacking	0.88	0.87	-0.85	-0.76	-0.83	-0.73
	***randstr* decoy set (67 reference structures, 33343 models)**					
A-A	0.83	0.84	-0.70	-0.82	-0.67	-0.71
A-S	0.88	0.89	-0.75	-0.87	-0.71	-0.75
S-S	0.88	0.89	-0.74	-0.87	-0.70	-0.74
S-S stacking	0.90	0.91	-0.74	-0.87	-0.70	-0.74
S-S non-stacking	0.80	0.80	-0.62	-0.75	-0.56	-0.62

[1]CAD-score variants defined by contacts of either all residue atoms or side chains (bases): 'A', all atoms, 'S', side chain (base). S-S (base-base) contacts are further subdivided into stacking and non-stacking ones.
[2]Interaction network fidelity
[3]Deformation index
[4]Root-mean-square deviation
[5]Pearson's correlation coefficient
[6]Spearman's ranking correlation coefficient

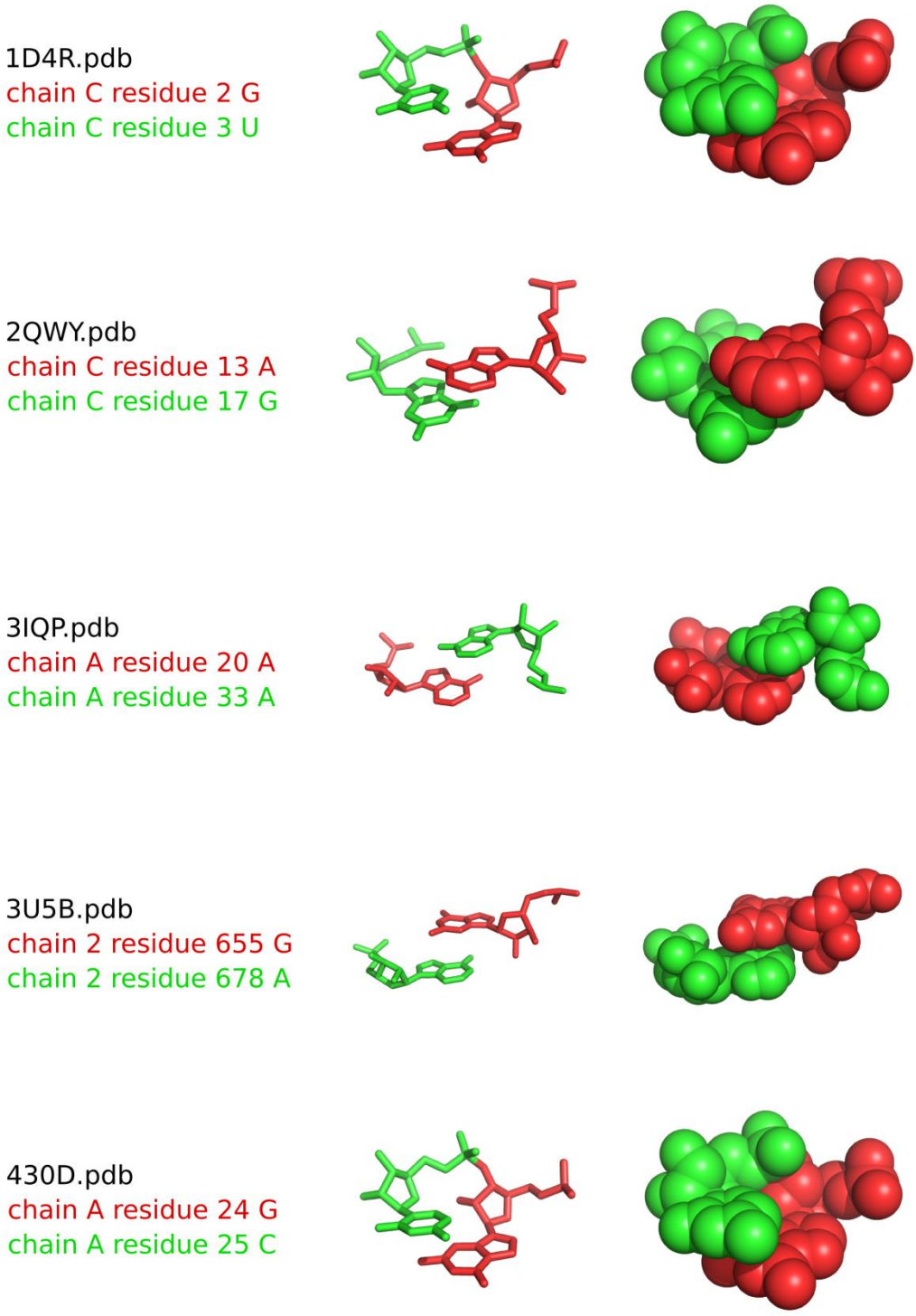

Figure S1. Examples of extensive base-base overlaps not detected by MC-Annotate as the base stacking interactions. For each pair of interacting bases the PDB code and the identities of both nucleotides (left) as well as two structural representations, sticks (middle) and space-filling (right), are shown.

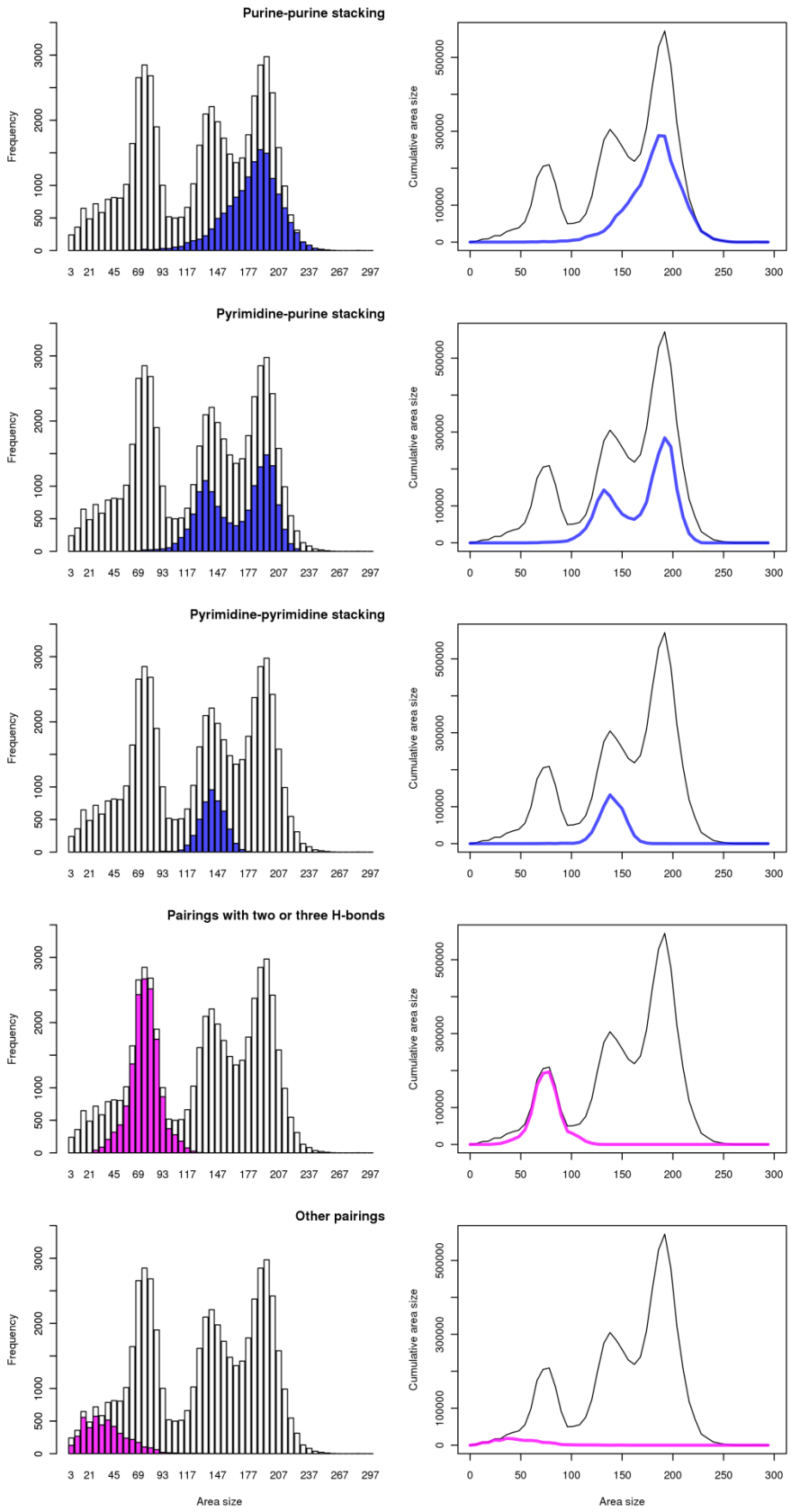

Figure S2. Partitioning of the base-base interactions annotated by MC-Annotate into distinct components according to the base-base contact frequency (left panels) and the cumulative contact areas (right panels). Grey bars and lines correspond to all the annotated base-base interactions, blue and magenta indicate their distinct base stacking and base pairing components respectively.

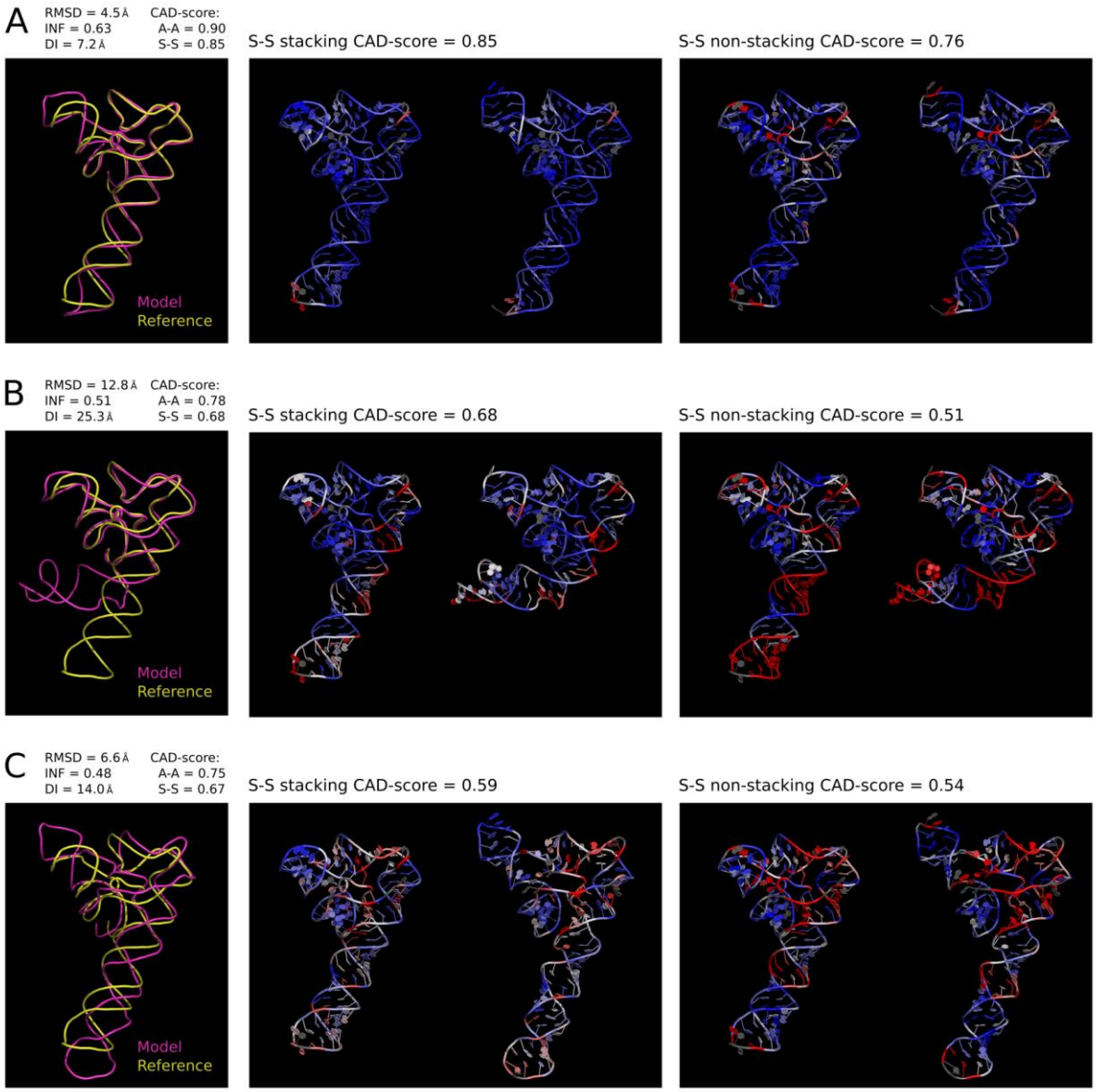

Figure S3. Illustration of the ability of base stacking and non-stacking (pairing) CAD-score variants to point out the location of errors in modeled RNA structures. The figure shows three models (A-C) from Challenge 4 of RNA-puzzles. Left panels show models (magenta) superimposed with the reference x-ray structure (yellow). Middle panels show the reference and a model colored according to the base stacking local CAD-score, right panels show the same structures colored according to the non-stacking local CAD-score. Blue-white-red color gradient represents the accuracy of reproduced contacts (blue – accurate, red –inaccurate). One of the models (A) represents a fairly accurate prediction, while the other two (B, C) represent poor predictions. In one of the poor models (B) incorrect base pairs are largely confined to two regions, while in the second one (C) the errors are dispersed throughout the structure. Although overall both models are of comparable poor quality, RMSD values give a misleading impression that one (C) is significantly more accurate than the other (B).

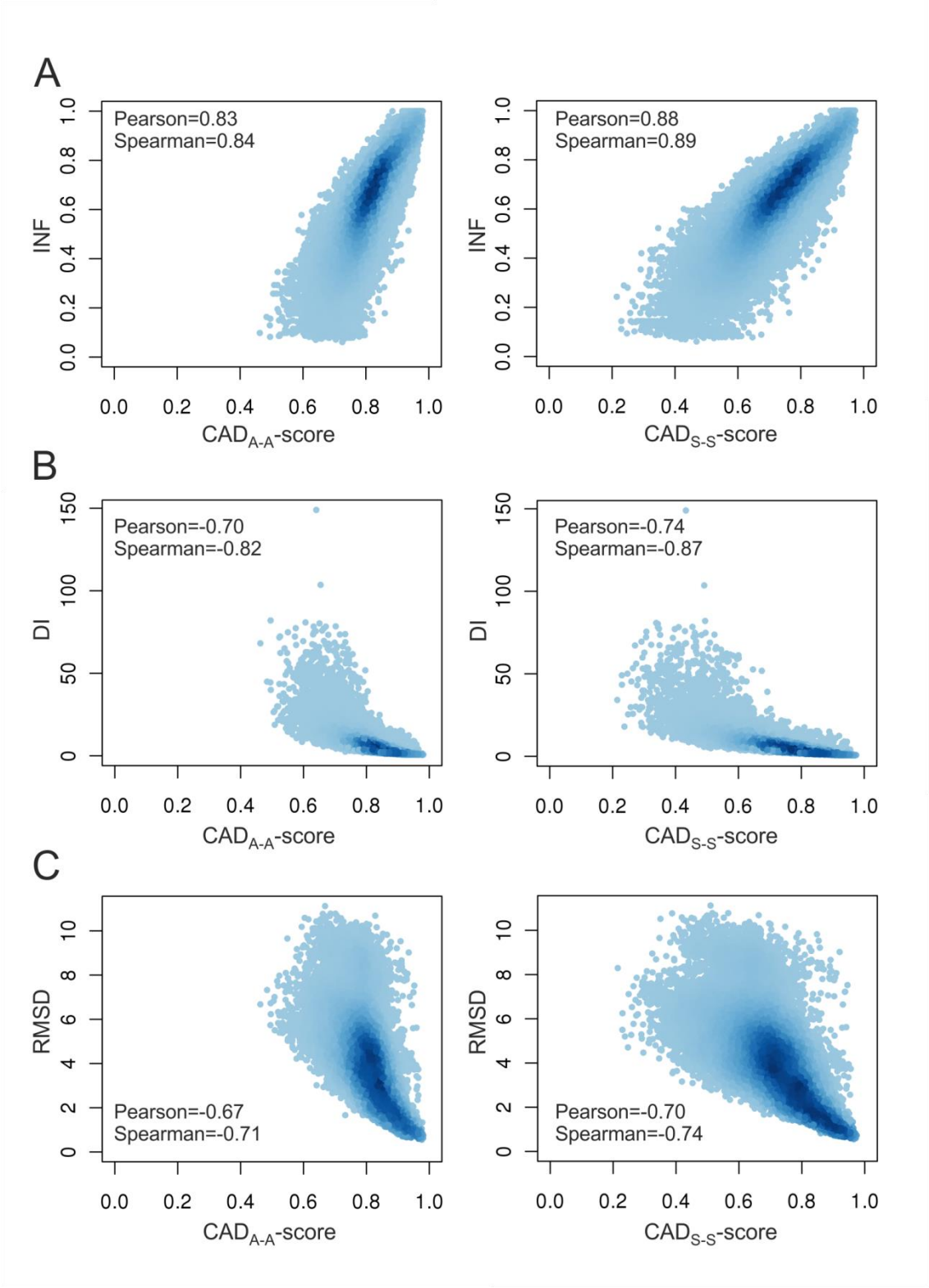

Figure S4. Relationship between CAD-score and the other three scores on the *randstr* decoy set. $CAD_{A\text{-}A}$-score (left) and $CAD_{S\text{-}S}$-score (right) are correlated with INF (A), DI (B) and RMSD (C). Pearson's correlation coefficients and Spearman's rank correlation coefficients are indicated for each plot. Higher color intensity reflects higher density of points.