Electronic Supplementary Material (ESI) for ChemComm.

Supplementary Materials for: Structure of the 5' untranslated region in SARS-CoV-2 genome and its specific recognition by innate immune system via the human oligoadenylate synthase 1

Emmanuelle Bignon,[*,†] Tom Miclot,[‡,†] Alessio Terenzi,[‡] Giampaolo Barone,[‡] and Antonio Monari[*,¶]

†*Université de Lorraine and CNRS, LPCT UMR 7019, F-54000 Nancy, France*

‡*Department of Biological, Chemical and Pharmaceutical Sciences. Universita degli Studi di Palermo, via delle Scienze 90126 Palermo, Italy.*

¶*Université de Paris, CNRS, ITODYS, F-75006, Paris, France*

E-mail: emmanuelle.bignon@univ-lorraine.fr; antonio.monari@u-paris.fr

Computational details

The starting structure for SL1 was generated using the RNAcomposer webserver,[1] and the OAS1 initial geometry was extracted from the crystal structure of the OAS1/ds-RNA complex (PDB ID 4IG8). The starting models for the OAS1/SL1 complex were generated by docking of representative structures from micro-second unbiased MD simulations of isolated SL1 and OAS1 using the HDock webserver.[2] The model chosen as starting structure for the MD simulations was the best ranked by the HDock algorithm. The control system was taken as the above-mentioned crystal structure showcasing OAS1 in interaction with a 18-bp RNA duplex. Protein residues were modeled using the ff14SB amber force field,[3] and bsc0+OL3 corrections were applied for RNA. The system was soaked in a cubic TIP3P water box with a 10Å buffer and potassium counter-ions were added to ensure a neutral total charge, resulting in systems from ∼22,000 to ∼60,000 atoms for the isolated SL1 and the complexes, respectively.

Protein/RNA Docking

The protein/RNA Docking was performed using the HDock online webserver with standard parameters (http://hdock.phys.hust.edu.cn/), and using OAS1 crystal structure and the prominent conformation of SL1 obtained by our MD simulations. HDock is based on a hybrid docking algorithm of template-based modeling and free docking, which provides ideal performances as extensively benchmarked elsewhere.[2] We also checked the HDock accuracy by performing a re-dock of the crystal structure of OAS1 with the RNA double strand (See Figure S4) obtaining a Root Mean Square Deviation (RMSD) of only 0.4 Å for the backbone atoms.

Molecular Dynamics simulations

MD simulations were carried out using NAMD3.[4] The Hydrogen Mass Repartitioning method (HMR)[5] was used to allow a 4 fs time step for the integration of the equations of motion. To prepare the system, 10,000 minimization steps were performed imposing positional constraints on the backbone atoms. 12 ns equilibration at 300K followed, during which the constraints were progressively released. The temperature was kept constant using the Langevin thermostat with a 1.0 ps^{-1} collision frequency, electrostatic interactions were treated using the Particle Mesh Ewald (PME) protocol.[6] After equilibration, the conformational ensemble was sampled along 1-μs production run and structures were dumped every 40 ps. The conformational sample was further explored using Gaussian Accelerated Molecular Dynamic (GAMD) simulations performed using the NAMD module. To this end an harmonic repulsive potential, following a Gaussian distribution, was added to the dihedral angles of the protein and the RNA. The potential was added considering an upper limit of the standard deviation of 6 kcal/mol every 10,000 steps. GAMD was run for a total time of 1.2 μs. The obtained biased sampling was reweighted to take into account the effect of the biasing potential, following the procedure described in ref,[7,8] and considering a free energy surface defined by the projection of the trajectory on top of the two main Principal Component Analysis (PCA) vectors describing the RNA dynamic. The two main PCA vectors are also reported in Figure S5.

Structural Analysis

The cpptraj module of AMBER18[9] was used to calculate distances and to perform the clustering analysis based on the of RMSD of the complex. Plots were generated using using the ggplot2 package of R[10] and representations of the molecules were rendered by VMD.[11]

Supplementary figures

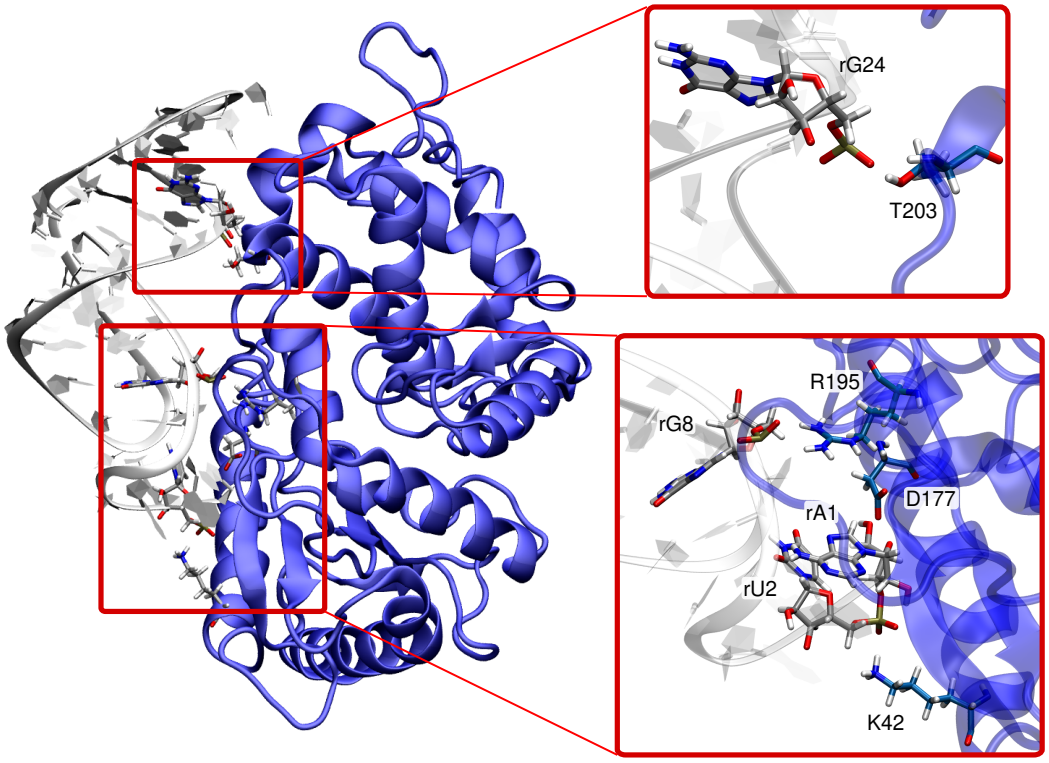

Figure S1 – Representative structure of the OAS1/SL1 complex highlighting the contact surface with the RNA minor grooves. Magnified sections: zoom onto the key interactions at the two contact surfaces: T203:HG1-rG24:OP1, R195:HH12-rG8:OP1, D177:OD2-rA1:HO5′, K42:HZ1-rU2:OP2. Nucleic and amino acids carbon atoms are depicted in grey and blue, respectively.

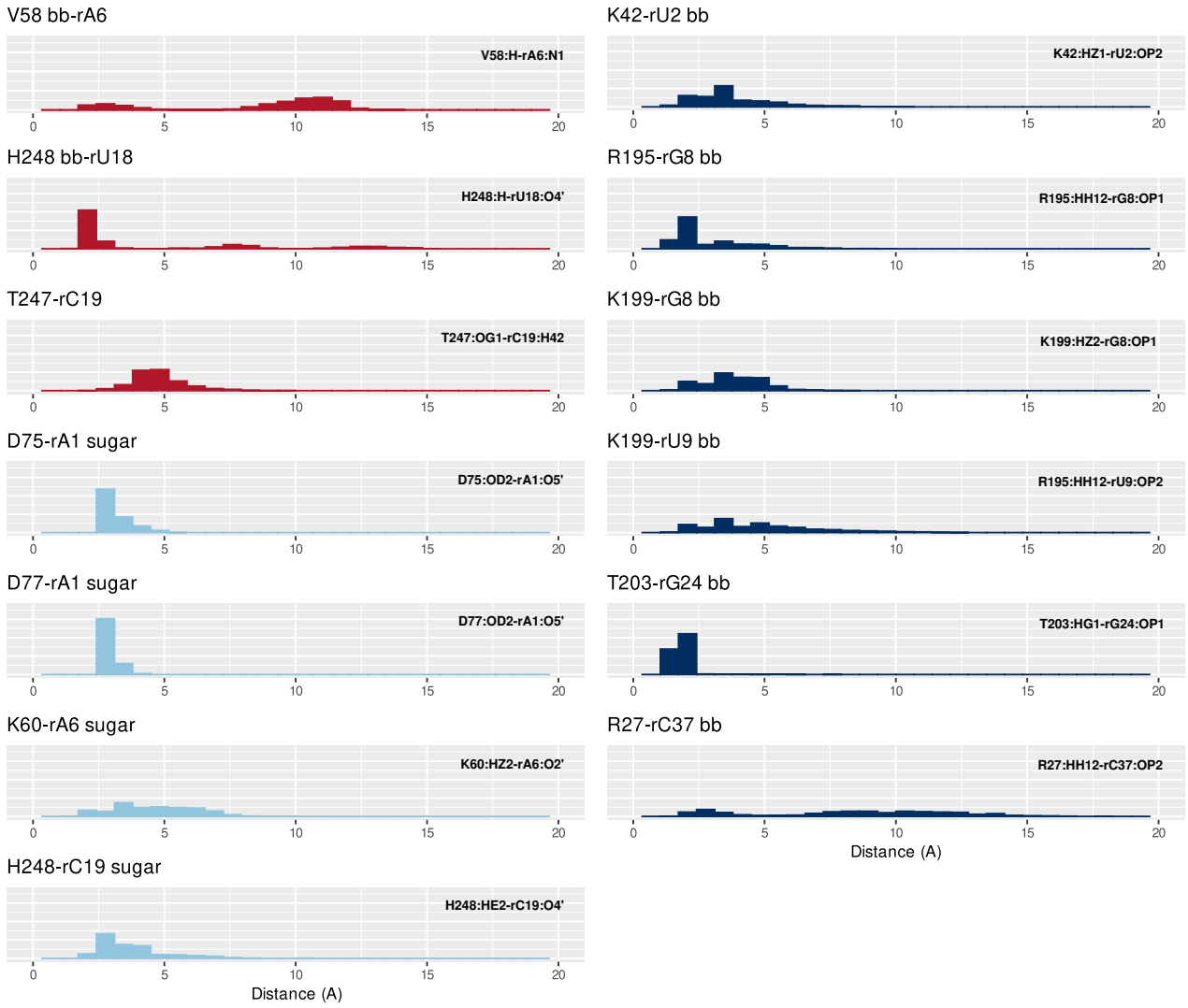

Figure S2 – Distribution of the key-distances (in Å) involved in the OAS1/SL1 interaction network. Contacts with the nucleobases appear in red, with the sugar in cyan and with the backbone (bb) in blue. A label 'bb' is added behind OAS1 residue names when the interaction involves the backbone atoms and not the side chain.

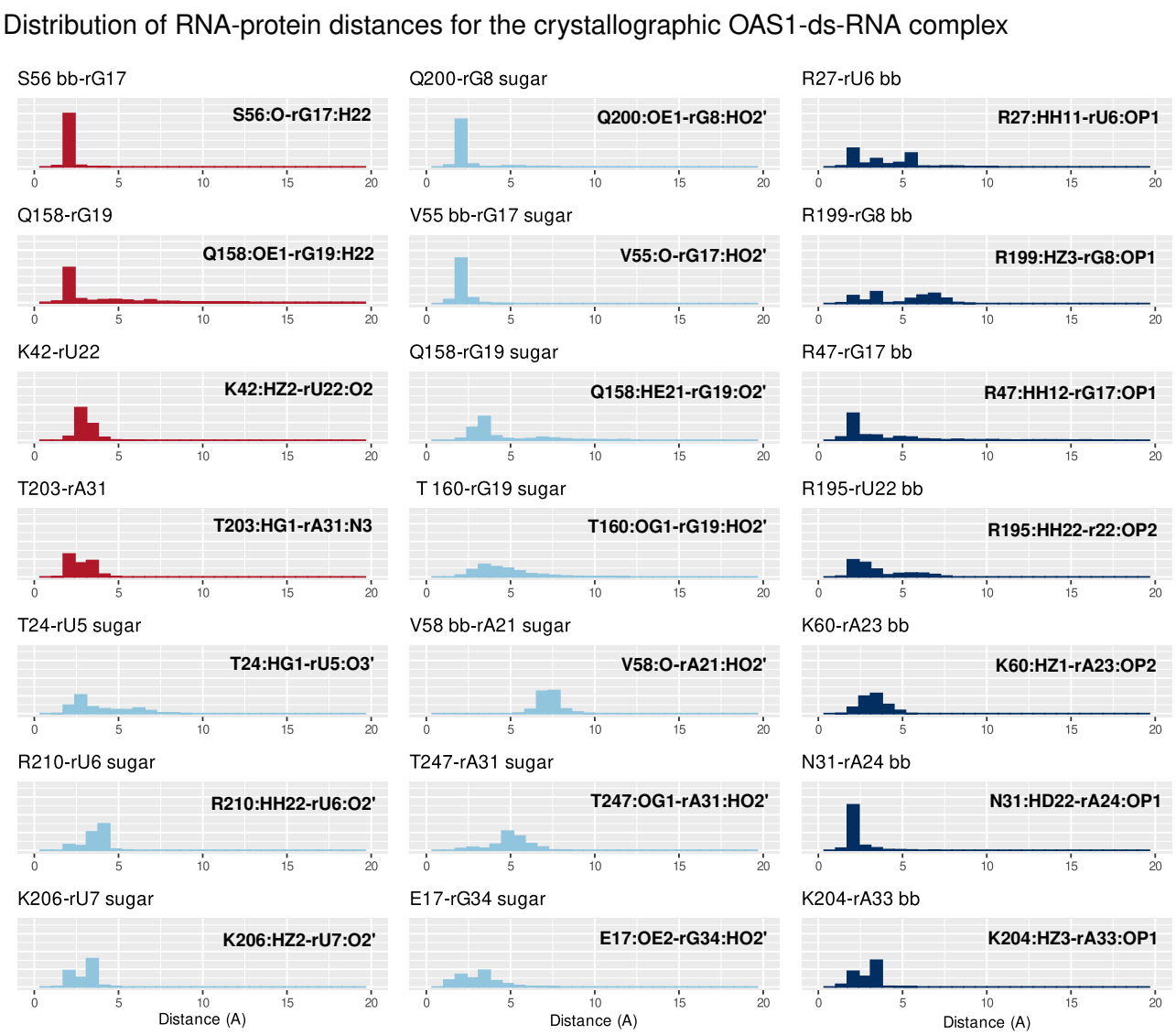

Figure S3 – Distribution of the key-distances (in Å) involved in the reference system (OAS1/ds-RNA crystallographic complex) interaction network. Contacts with the nucleobases appear in red, with the sugar in cyan and with the backbone (bb) in blue. A label 'bb' is added behind OAS1 residue names when the interaction involves the backbone atoms and not the side chain.

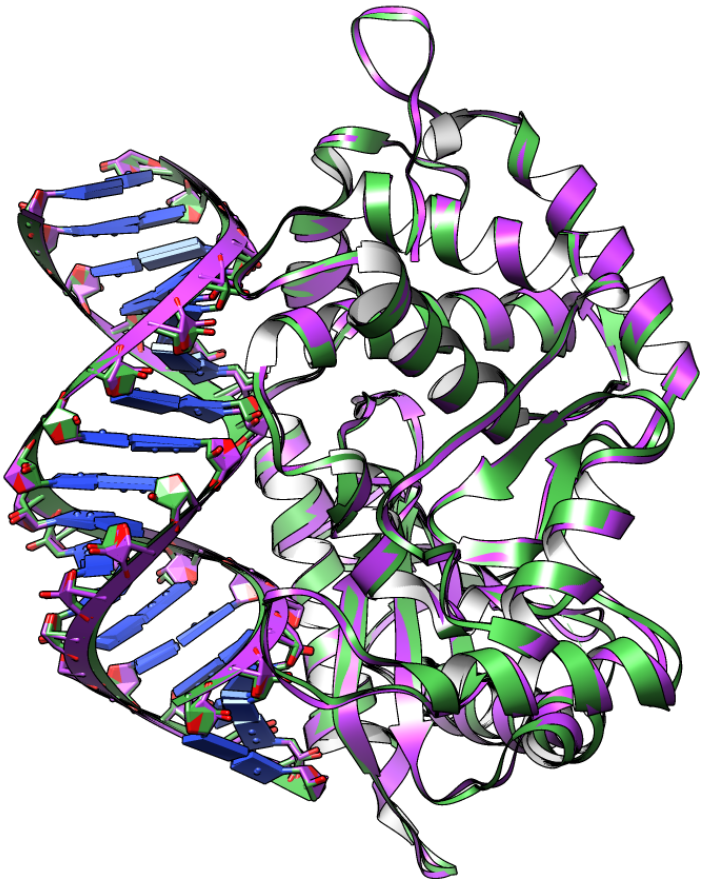

Figure S4 – Superimposition of the dsRNA-OAS1 reference crystal structure (green) and the prediction of this complex by HDock (magenta), showing the very good performance of the latter.

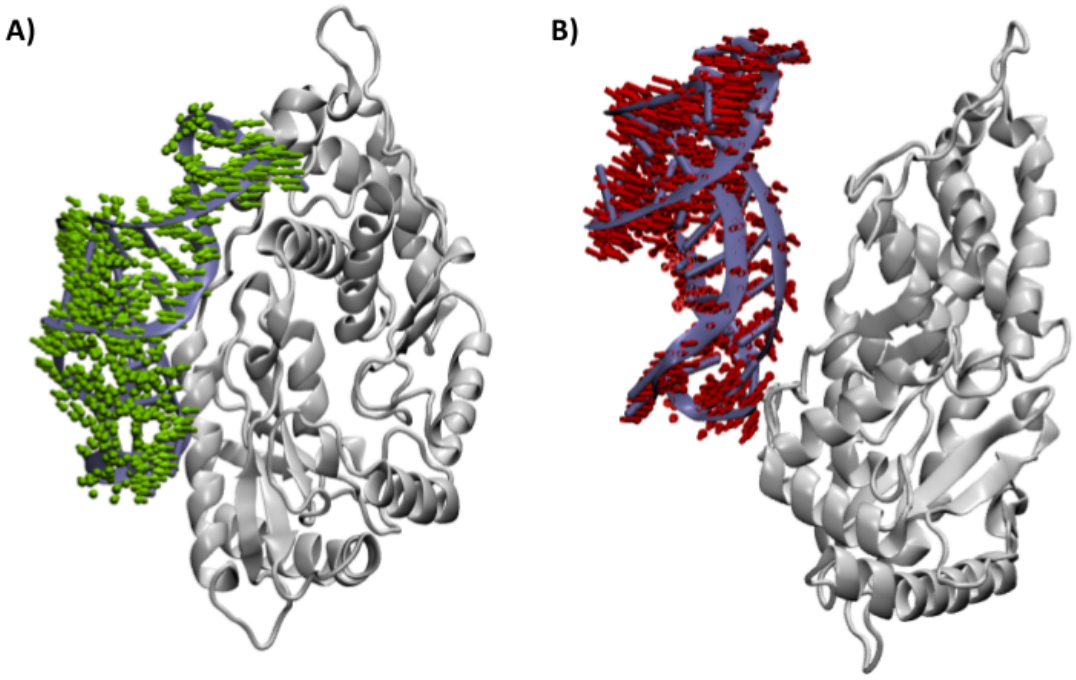

Figure S5 – Top: Representation of the first (PCA 1) and second (PCA 2) collective modes issued by the PCA analysis and used as collective variables for the GaMD reweight procedure. Bottom: global view of the conformations corresponding to minimum energy regions labeled B), C) and D).

References

(1) Popenda, M.; Szachniuk, M.; Antczak, M.; Purzycka, K. J.; Lukasiak, P.; Bartol, N.; Blazewicz, J.; Adamiak, R. W. Automated 3D structure composition for large RNAs. *Nucleic Acids Res.* **2012**, *40*, e112–e112.

(2) Yan, Y.; Zhang, D.; Zhou, P.; Li, B.; Huang, S.-Y. HDOCK: a web server for protein–protein and protein–DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Res.* **2017**, *45*, W365–W373.

(3) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.

(4) Phillips, J. C.; Hardy, D. J.; Maia, J. D.; Stone, J. E.; Ribeiro, J. V.; Bernardi, R. C.; Buch, R.; Fiorin, G.; Hénin, J.; Jiang, W., et al. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J. Chem. Phys.* **2020**, *153*, 044130.

(5) Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E. Long-time-step molecular dynamics through hydrogen mass repartitioning. *J. Chem. Theory Comput.* **2015**, *11*, 1864–1874.

(6) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(7) Sinko, W.; Miao, Y.; de Oliveira, C. A. F.; McCammon, J. A. Population Based Reweighting of Scaled Molecular Dynamics. *J. Phys. Chem. B* **2013**, *117*, 12759–12768, PMID: 23721224.

(8) Miao, Y.; Sinko, W.; Pierce, L.; Bucher, D.; Walker, R. C.; McCammon, J. A. Improved Reweighting of Accelerated Molecular Dynamics Simulations for Free Energy Calcu-

lation. *Journal of Chemical Theory and Computation* **2014**, *10*, 2677–2689, PMID: 25061441.

(9) Case, D.; Ben-Shalom, I.; Brozell, S.; Cerutti, D.; Cheatham III, T.; Cruzeiro, V.; Darden, T.; Duke, R.; Ghoreishi, D.; Gilson, M., et al. AMBER 2018: San Francisco. 2018.

(10) Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*; Springer-Verlag New York, 2009.

(11) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38.