Supporting Information

Deciphering off-target effects in CRISPR-Cas9 through accelerated molecular dynamics

Clarisse G. Ricci, Janice S. Chen, Yinglong Miao, Martin Jinek,

Jennifer A. Doudna, J. Andrew McCammon and Giulia Palermo

Supplementary Figures

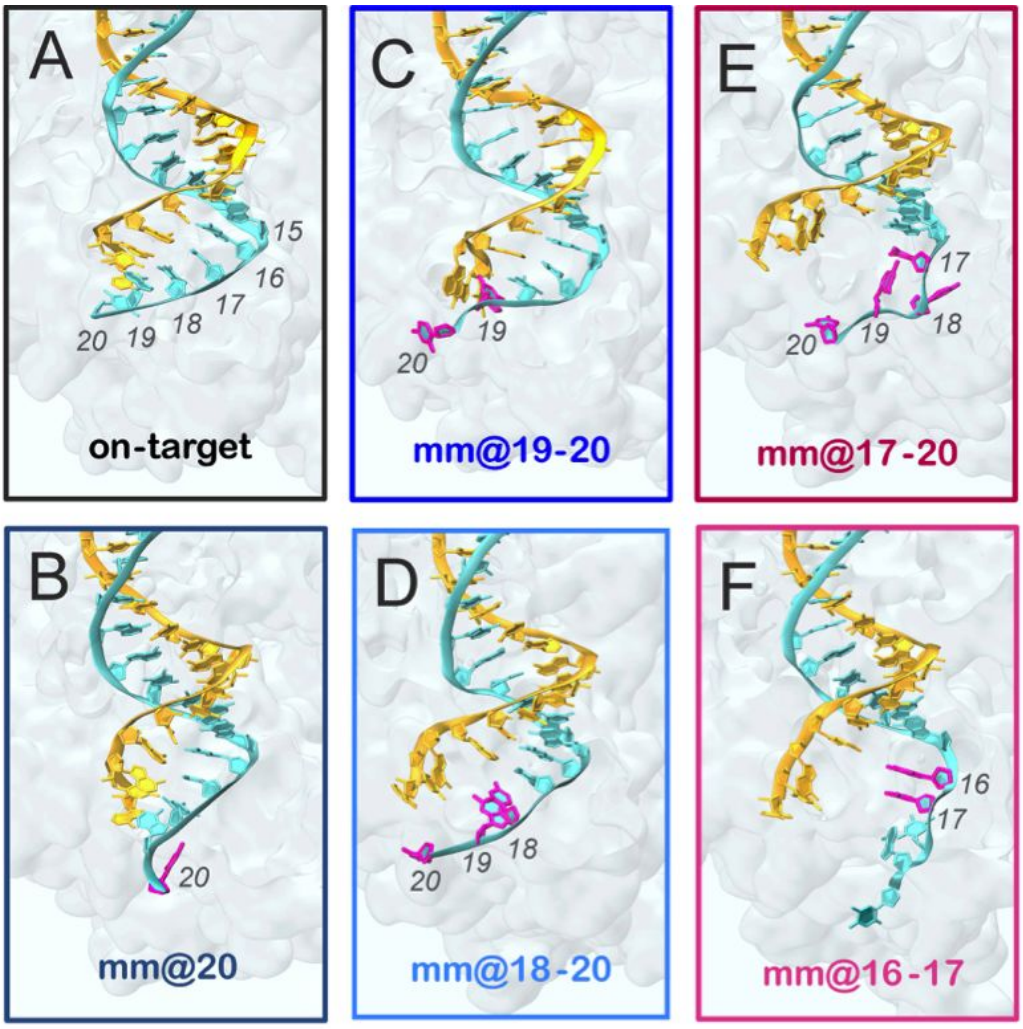

Figure S1. Conformations adopted by the RNA:DNA hybrid along GaMD simulations. Representative snapshots extracted from GaMD simulations of the CRISPR-Cas9 system, including the on-target DNA **(A)** and base pair mismatches (mm) at different positions of the hybrid: mm@20 **(B)**, mm@19-20 **(C)**, mm@18-20 **(D)**, mm@17-20 **(E)** and mm@16-17 **(F)**. The RNA (orange) and the target DNA (TS, cyan) are shown as ribbons. Mismatched bases on the TS are highlighted in magenta. The protein environment is shown as a molecular surface. These configurations are representative of the conformational changes detailed in Figures S2-5 and in Figure 3 of the main text.

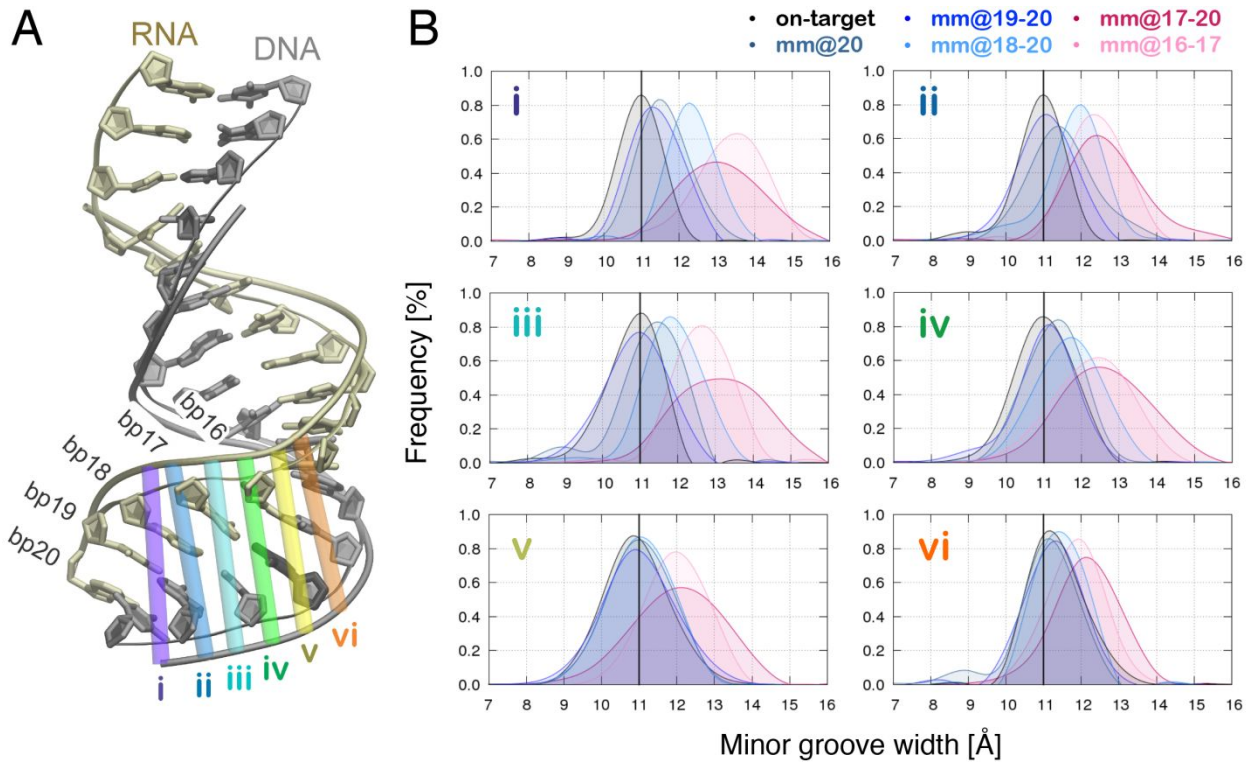

Figure S2. RNA:DNA minor groove width. **(A)** The RNA:DNA minor groove width has been computed at PAM distal sites at six different levels (*i-vi*) perpendicularly to the global helical axis, which are schematically shown on the 3D structure of the RNA:DNA hybrid. **(B)** Probability distribution of the RNA:DNA minor groove widths, computed at the (*i-vi*) levels perpendicularly to the global helical axis for six model systems of CRISPR-Cas9: including the on-target DNA (on-target system) and base pair (bp) mismatches (mm) at different positions of the RNA:DNA hybrid (mm@20, mm@19-20, mm@18-20, mm@17-20 and mm@16-17). A vertical bar indicates the experimental minor groove width (i.e., ~11 Å from X-ray crystallography, enlarged by ~1 Å if NMR data are considered. Experimental data are taken from Perez A.; Lankas, F.; Luque, F. J.; Orozco, M. *Nucleic Acids Res.* **2008**, *36*, 2379)

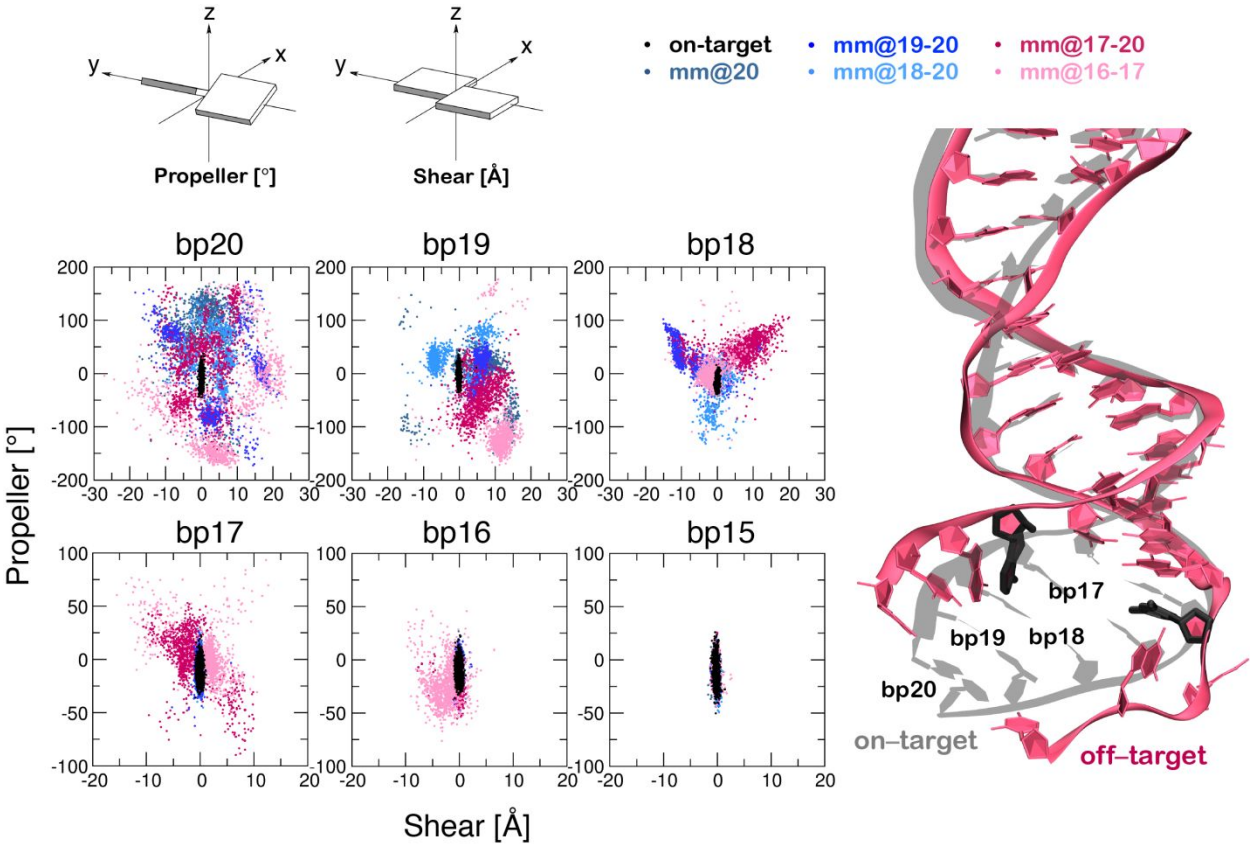

Figure S3. Scatter plots of the *Propeller vs. Shear* parameters (shown on top of the graphs), defining the base pair complementarity of the RNA:DNA hybrid. Data are computed for the base pairs (bp) at positions 20 to 15 (bp20–bp15) of the hybrid during production GaMD runs for each of the simulated systems of CRISPR-Cas9: including the on-target DNA and bp mismatches (mm) at different positions of the hybrid (mm@20, mm@19-20, mm@18-20, mm@17-20 and mm@16-17). The RNA:DNA hybrid is shown on the right for the on-target and the mm@17-20 systems. In the on-target system, the DNA:RNA hybrid stably maintains its Watson-Crick base pairing, as shown from the confined distribution of the *Propeller vs. Shear* (black dots). In all systems including off-target sequences, we observe a loss of complementarity at positions 20 to 18 of the hybrid, as shown by a broad distribution of the scatter plots (top graphs). At bp17–bp16, only the mm@17-20 and mm@16-17 systems show remarkable loss of the base pairing, whereas at bp15, all systems keep their Watson-Crick base pairing.

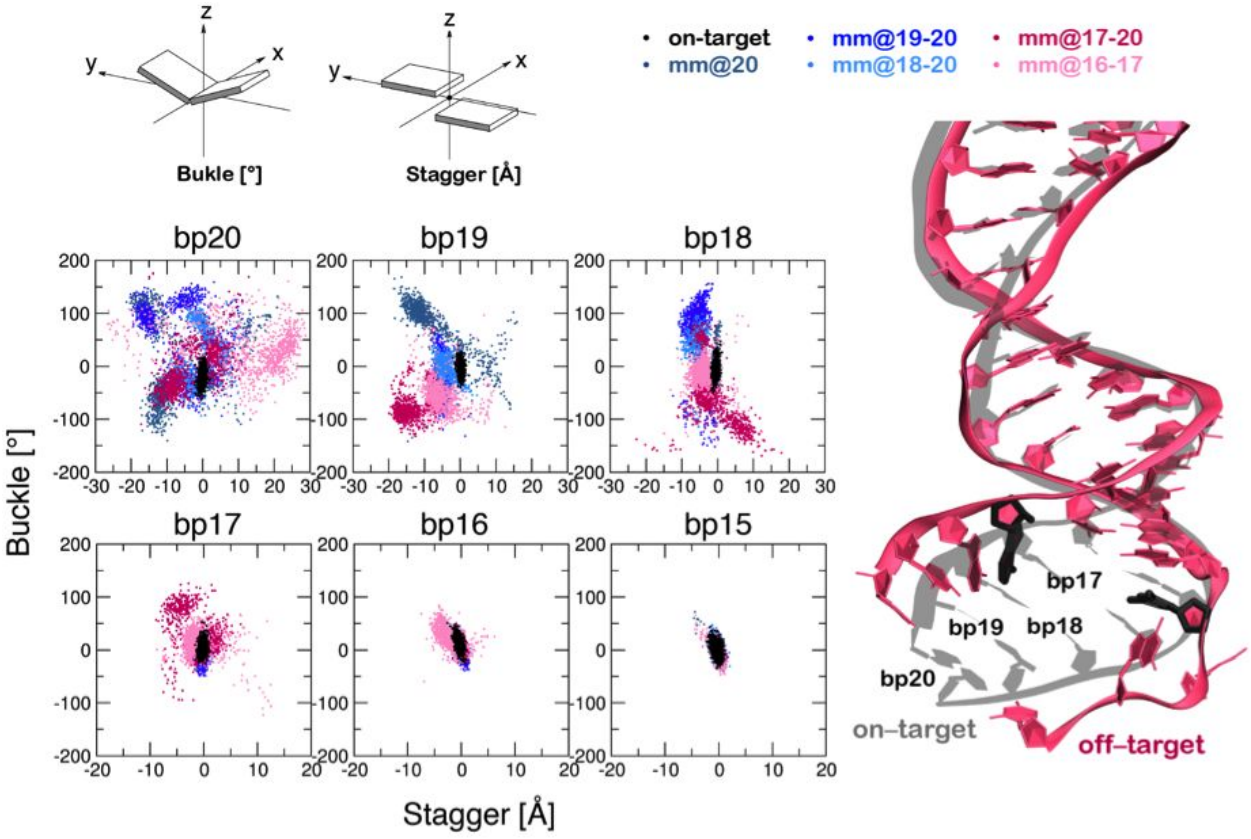

Figure S4. Scatter plots of the *Buckle vs. Stagger* parameters (shown on top of the graphs), defining the base pair complementarity of the RNA:DNA hybrid. Data are computed for the base pairs (bp) at positions 20 to 15 (bp20–bp15) of the hybrid during production GaMD runs for each of the simulated systems of CRISPR-Cas9: including the on-target DNA and bp mismatches (mm) at different positions of the hybrid (mm@20, mm@19-20, mm@18-20, mm@17-20 and mm@16-17). The RNA:DNA hybrid is shown on the right for the on-target and the mm@17-20 systems. In the on-target system, the DNA:RNA hybrid stably maintains its Watson-Crick base pairing, as shown from the confined distribution of the *Buckle vs. Stagger* (black dots). In all systems including off-target sequences, we observe a loss of complementarity at positions 20 to 18 of the hybrid, as shown by a broad distribution of the scatter plots (top graphs). At bp17–bp16, only the mm@17-20 and mm@16-17 systems show remarkable loss of the base pairing, whereas at bp15, all systems keep their Watson-Crick base pairing.

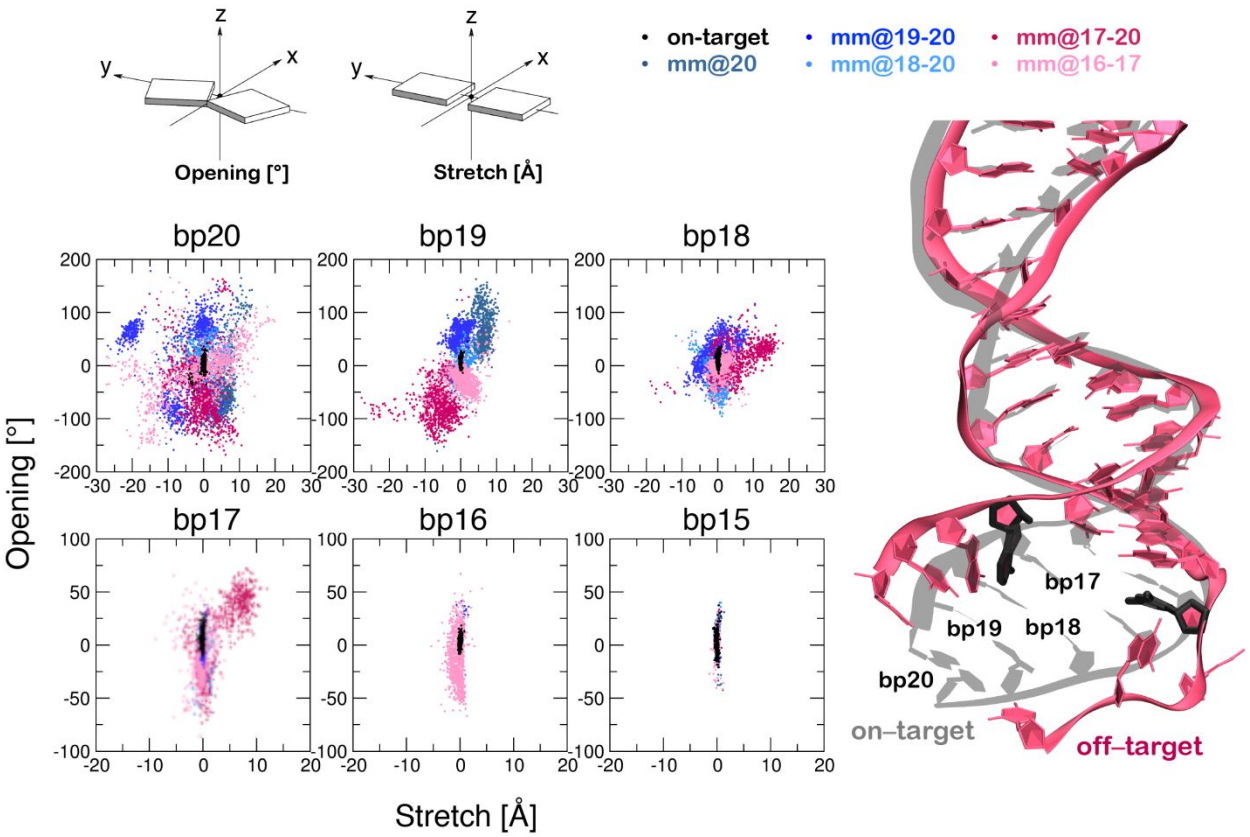

Figure S5. Scatter plots of the *Opening vs. Stretch* parameters (shown on top of the graphs), defining the base pair complementarity of the RNA:DNA hybrid. Data are computed for the base pairs (bp) at positions 20 to 15 (bp20–bp15) of the hybrid during production GaMD runs for each of the simulated systems of CRISPR-Cas9: including the on-target DNA and bp mismatches (mm) at different positions of the hybrid (mm@20, mm@19-20, mm@18-20, mm@17-20 and mm@16-17). The RNA:DNA hybrid is shown on the right for the on-target and the mm@17-20 systems. In the on-target system, the DNA:RNA hybrid stably maintains its Watson-Crick base pairing, as shown from the confined distribution of the *Opening vs. Stretch* (black dots). In all systems including off-target sequences, we observe a loss of complementarity at positions 20 to 18 of the hybrid, as shown by a broad distribution of the scatter plots (top graphs). At bp17–bp16, only the mm@17-20 and mm@16-17 systems show remarkable loss of the base pairing, whereas at bp15, all systems keep their Watson-Crick base pairing.

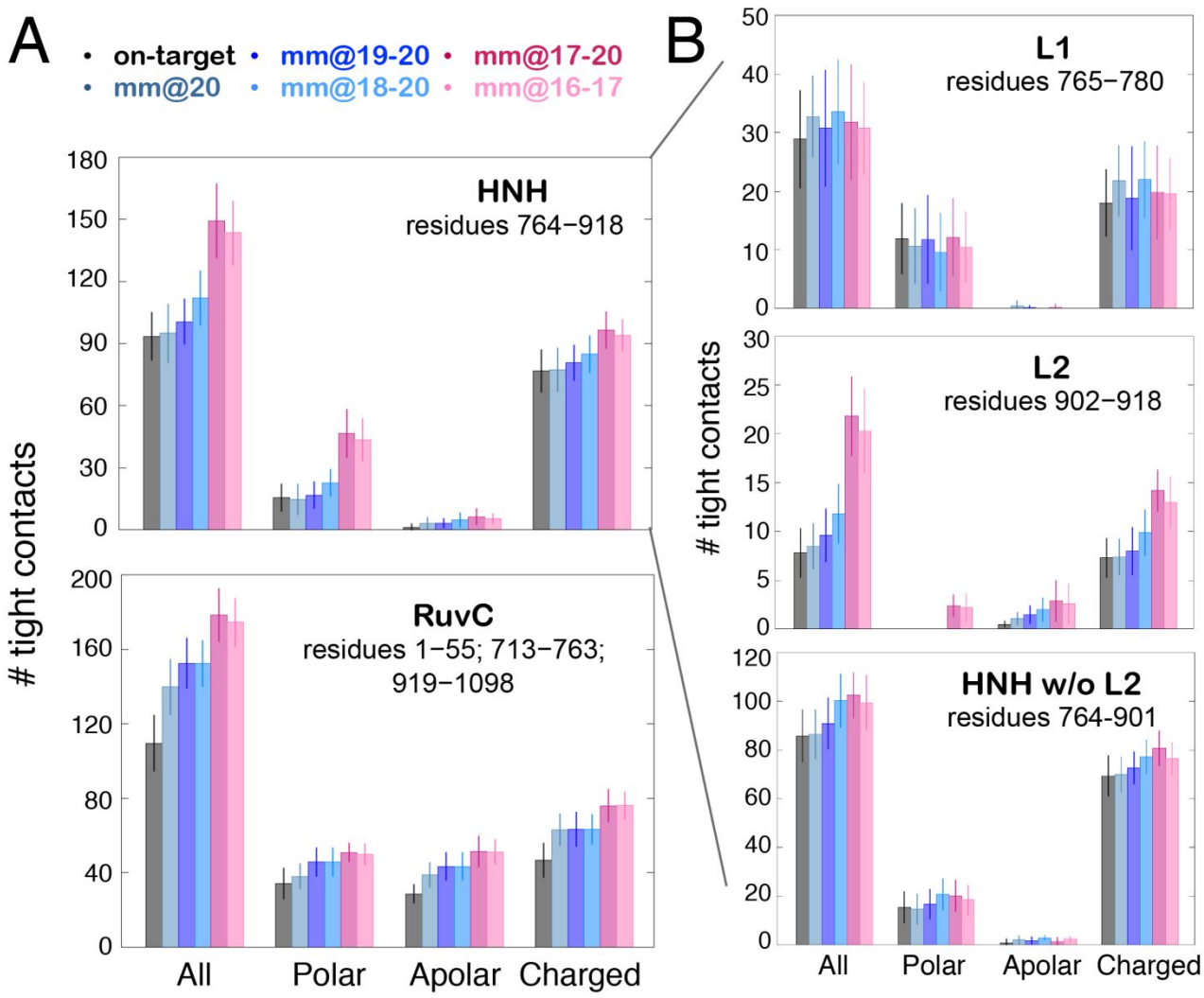

Figure S6. Number of tight contacts between the RNA:DNA hybrid and the neighboring residues. **(A)** Number of tight contacts (i.e., within 4 Å radius) established along the dynamics by the RNA:DNA hybrid and the neighboring residues of the HNH (top graph) and RuvC (bottom graph) domains. **(B)** The HNH domain includes two flexible loops, L1 and L2, which connect HNH to RuvC. The number of tight contacts established by the RNA:DNA hybrid and the L1 (top graph) and L2 (central graph) loops, as well as with the HNH domain excluding the residues constituting the L2 loop (i.e., HNH w/o L2; bottom graph) are reported. The number of tight contacts is computed considering all domains residues, the polar, apolar and charged residues. Data are reported for all simulated systems (on-target, mm@20, mm@19-20, mm@18-20,

mm@17-20 and mm@16-17). Data are computed and averaged over the last ~400 ns of GaMD simulations, allowing us to take into account well established interactions (details are in the Methods section of the main text). The associated statistical errors are reported. In the mm@17-20 and mm@16-17 systems, the number of tight contacts computed between the hybrid and HNH excluding the residues belonging to L2 (panel **B**, bottom graph), reveals a decrease in interactions for the with respect to HNH including L2 (panel **A**, top graph). Moreover, in these systems, we measure an increase in the number of tight contacts between the hybrid and the L2 loop of the HNH domain (panel B, central graph). This indicates that the interactions established by HNH and the hybrid occurs at the level of L2 and mainly involves polar and charged residues.

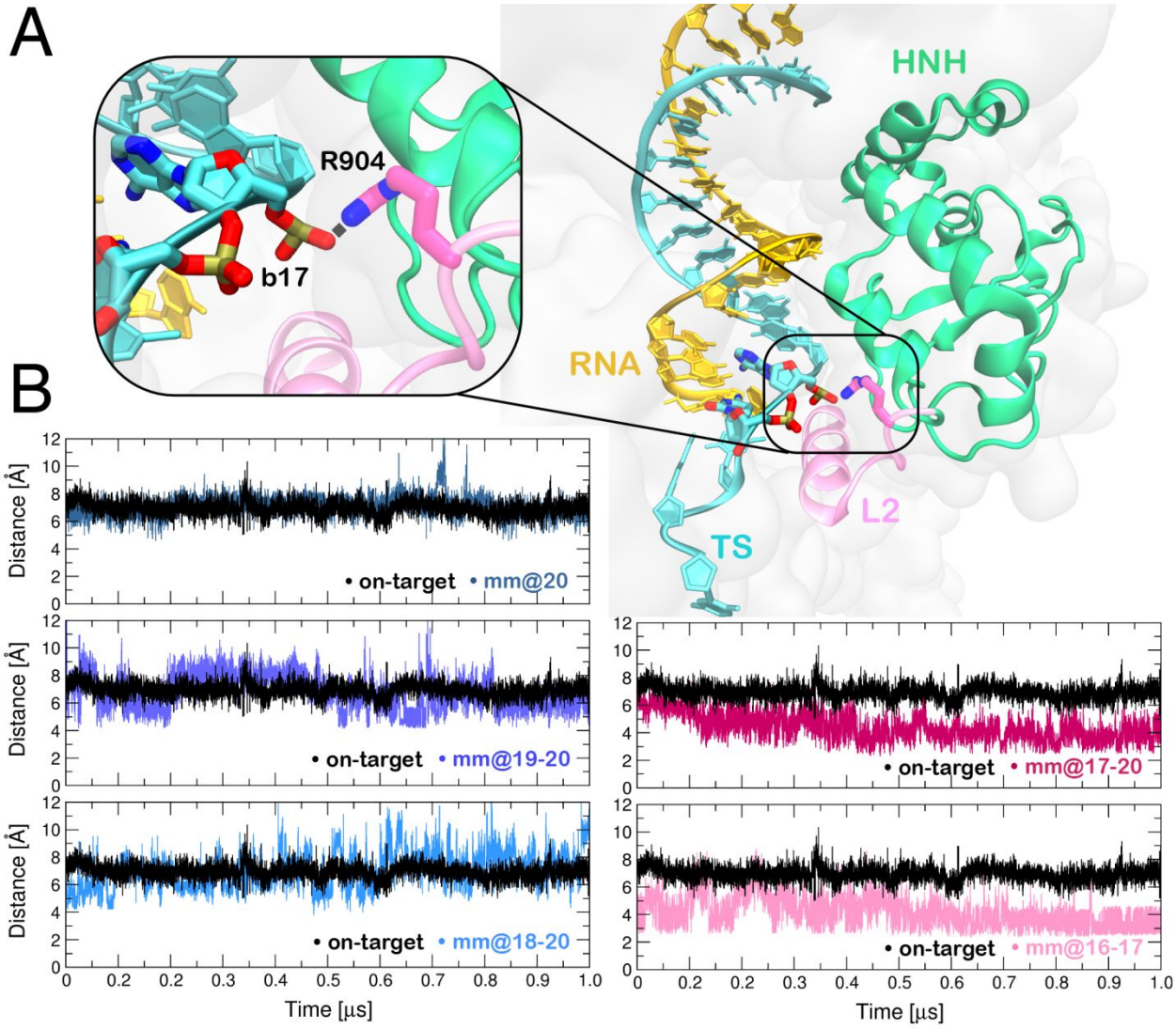

Figure S7. (A) Interaction between the R904 residue (belonging to the L2 loop, which connects HNH to RuvC) and the DNA TS at position 17 (R904–b17) Representative snapshot from GaMD simulations of the mm@16-17 system. **(B)** Time evolution of the R904–b17 distance (computed between the R904 terminal nitrogen and the oxygen of the TS backbone at position 17), along GaMD of the investigated systems. Each graphs compares the R904–b17 distance for the on-target system (black) with the mm@20, mm@19-20, mm@18-20, mm@17-20 and mm16-17 systems. Remarkably, the R904–b17 interaction does not form in the "productive" systems (i.e., on-target, mm@20, mm@19-20 and mm@18-20).

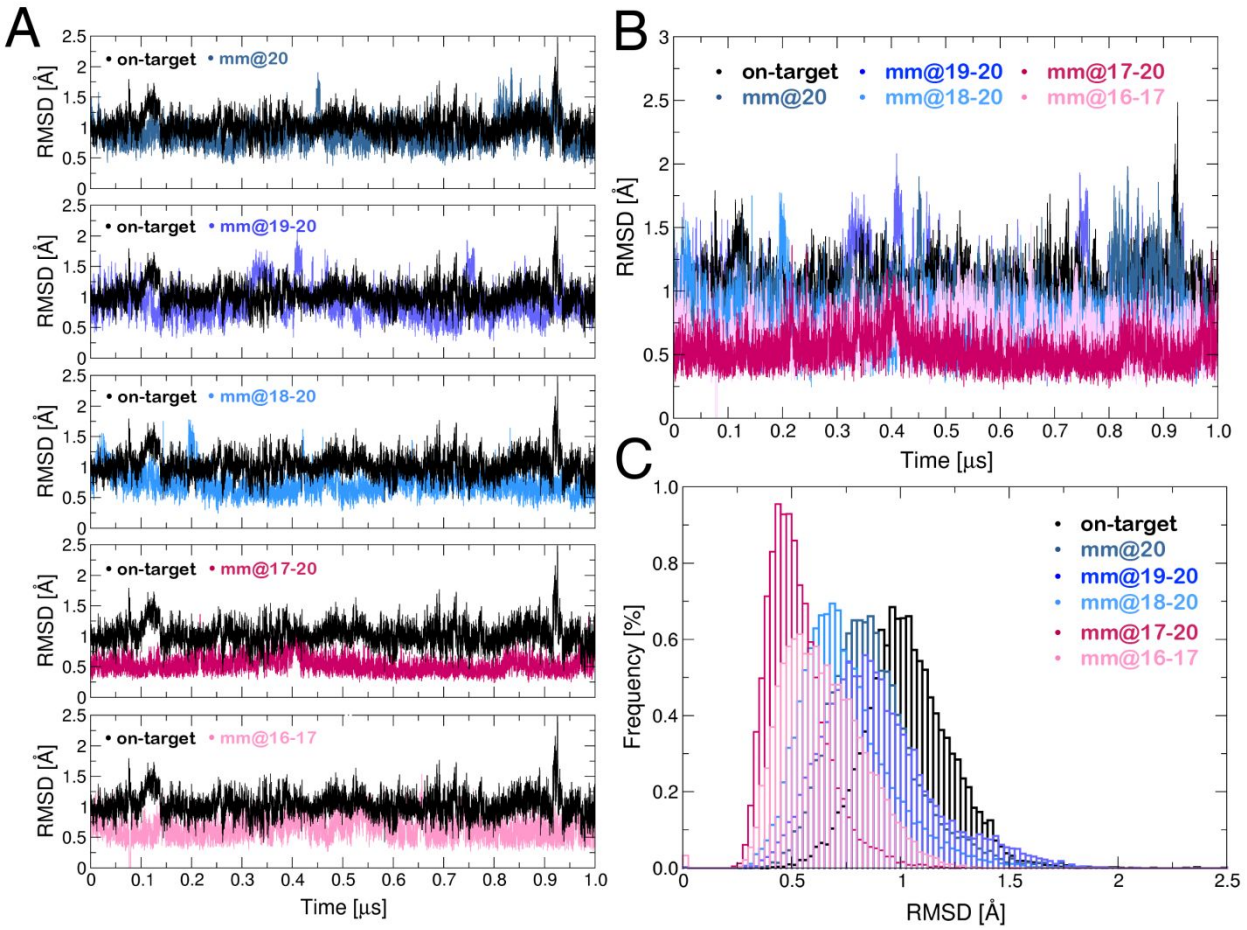

Figure S8. (A) Time evolution of the Root Mean Square Deviation (RMSD) of the Cα atoms of the L2 loop connecting HNH to RuvC (residues 902–918), along GaMD of the investigated systems. Each graphs compares the RMSD for the on-target system (black) with the mm@20, mm@19-20, mm@18-20, mm@17-20 and mm16-17 systems. **(B-C)** Data computed for all systems are superposed **(B)** and plotted as probability distribution **(C)**. The RMSD is computed with respect to the equilibrated structure, extracted as an average from the GaMD simulations. This allowed us to determine the stability the L2 loop as an effect of the conformational changes occurring during GaMD (i.e., the opening of the hybrid and its interaction with the L2 loop). In the on-target Cas9 (black lines), L2 displays higher values of the RMSD, indicating an increased conformational flexibility. In the "productive" off-target systems

(i.e., mm@20, mm@19-20, mm@18-20), the RMSD superposes the on-target Cas9, indicating a similar stability. In the “unproductive” off-target system (mm@17-20), as well as in the system including base pair mismatches at positions 16-17 (mm@16-17), the RMSD of L2 shows lower values, which indicate increased stability of the L2 loop.

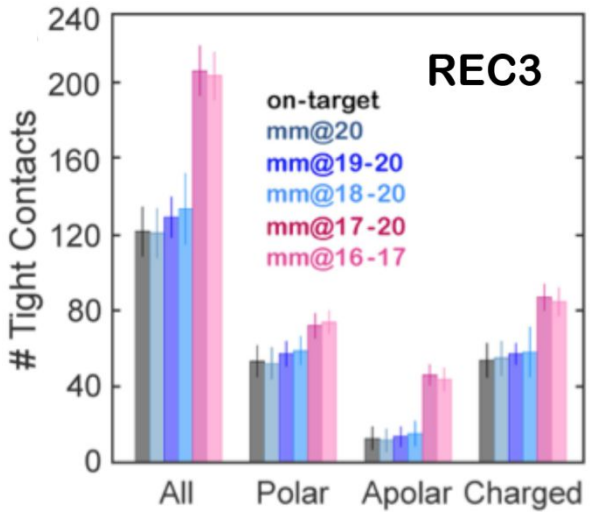

Figure S9. Number of tight contacts (i.e., within 4 Å radius) established along the dynamics by the RNA:DNA hybrid and the neighboring residues of the REC3 region computed for all the simulated systems: on-target, mm@20, mm@19-20, mm@18-20, mm@17-20 and mm@16-17. The polar, apolar and charged groups of residues have been considered.

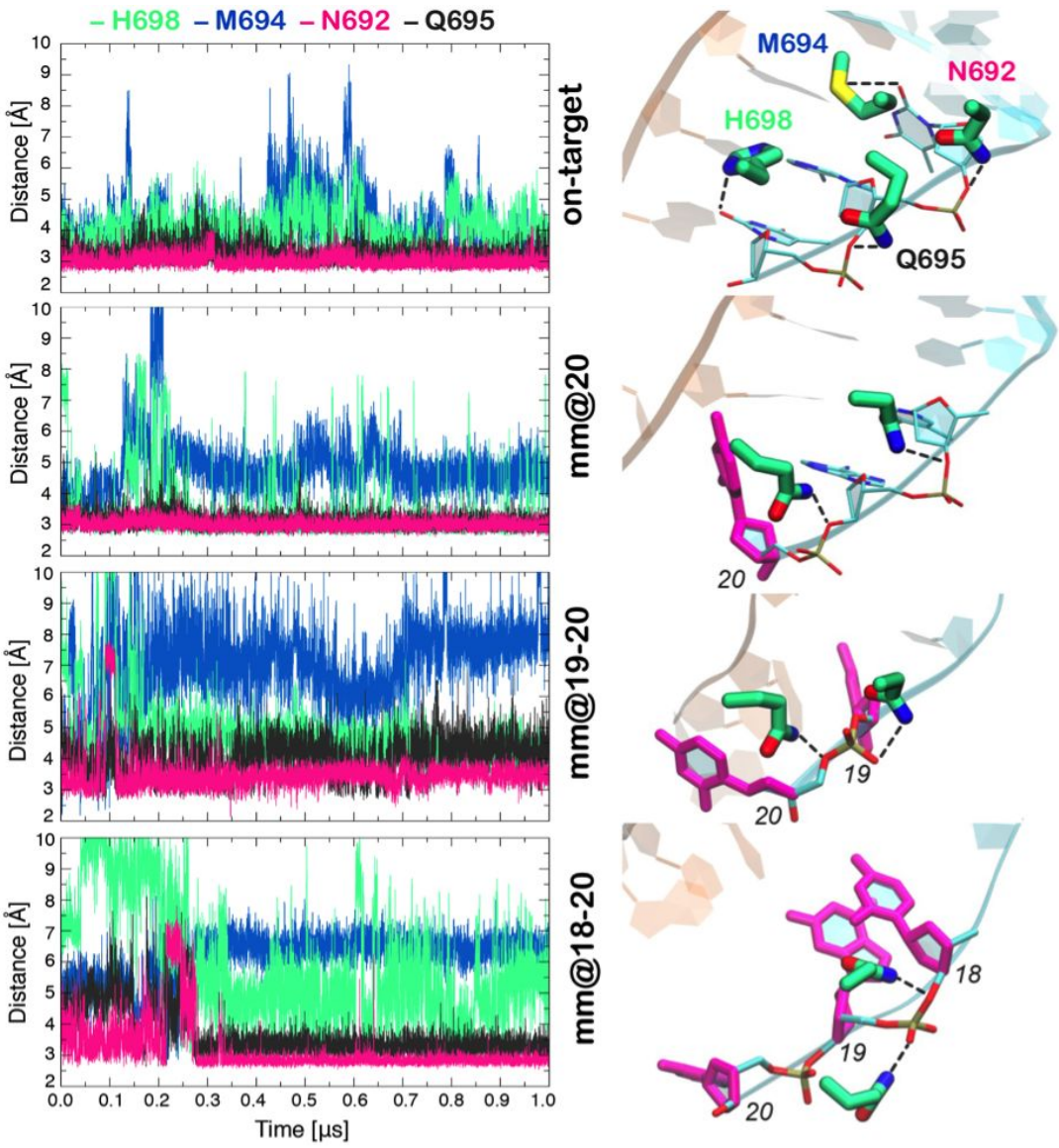

Figure S10. Time evolution of the interactions established by the H698, M694, N692 and Q695 residues with the DNA TS along GaMD of the on-target CRISPR-Cas9 system and in the mm@20, mm@19-20 and mm@18-20 systems (i.e., "productive"). The data show that in the "productive" systems, N692 and Q695 establish conserved interactions at different positions of the TS (displayed on the right). The RNA (orange) and TS (cyan) are shown as ribbons. Protein residues (green) and mismatched bases (magenta) are shown as sticks.

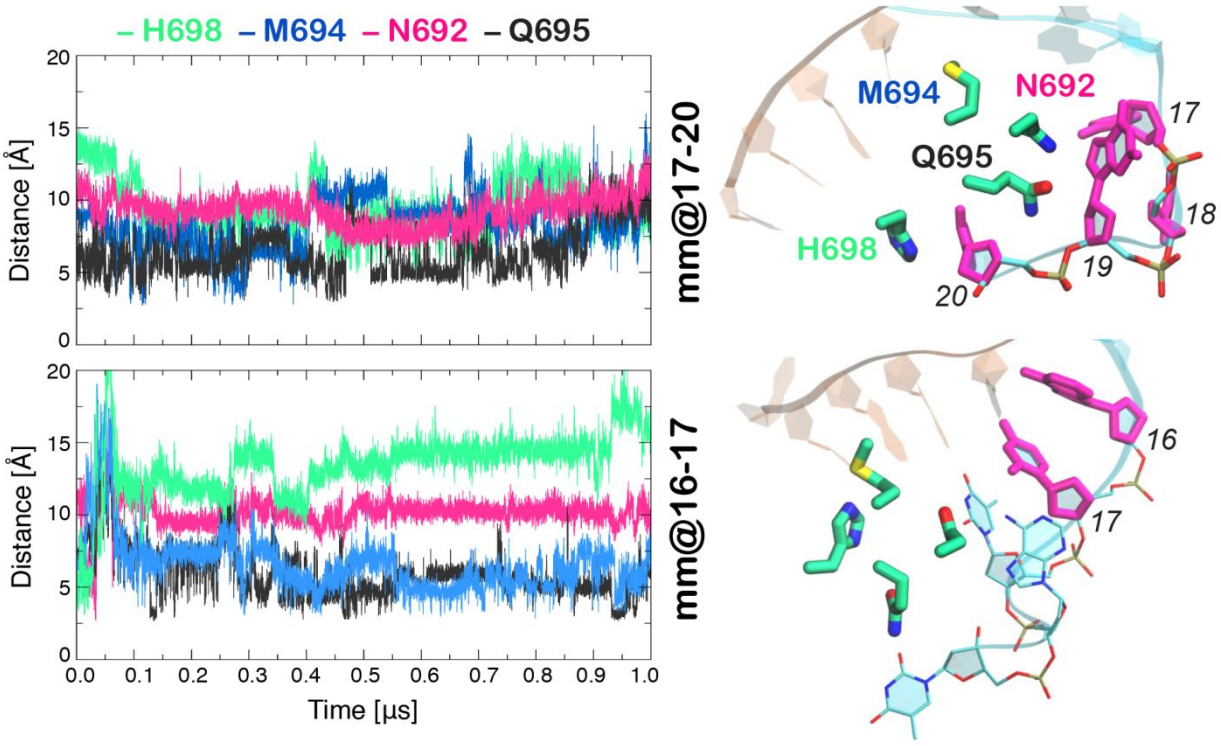

Figure S11. Time evolution of the interactions established by the H698, M694, N692 and Q695 residues with the DNA TS along GaMD of the "unproductive" system mm@17-20 and of the mm@16-17 system. The data reveal the loss of interactions between the protein residues and the TS backbone, as opposite to the "productive" systems (Figure S9). The RNA (orange) and TS (cyan) are shown as ribbons. Protein residues (green) and mismatched bases (magenta) are shown as sticks.