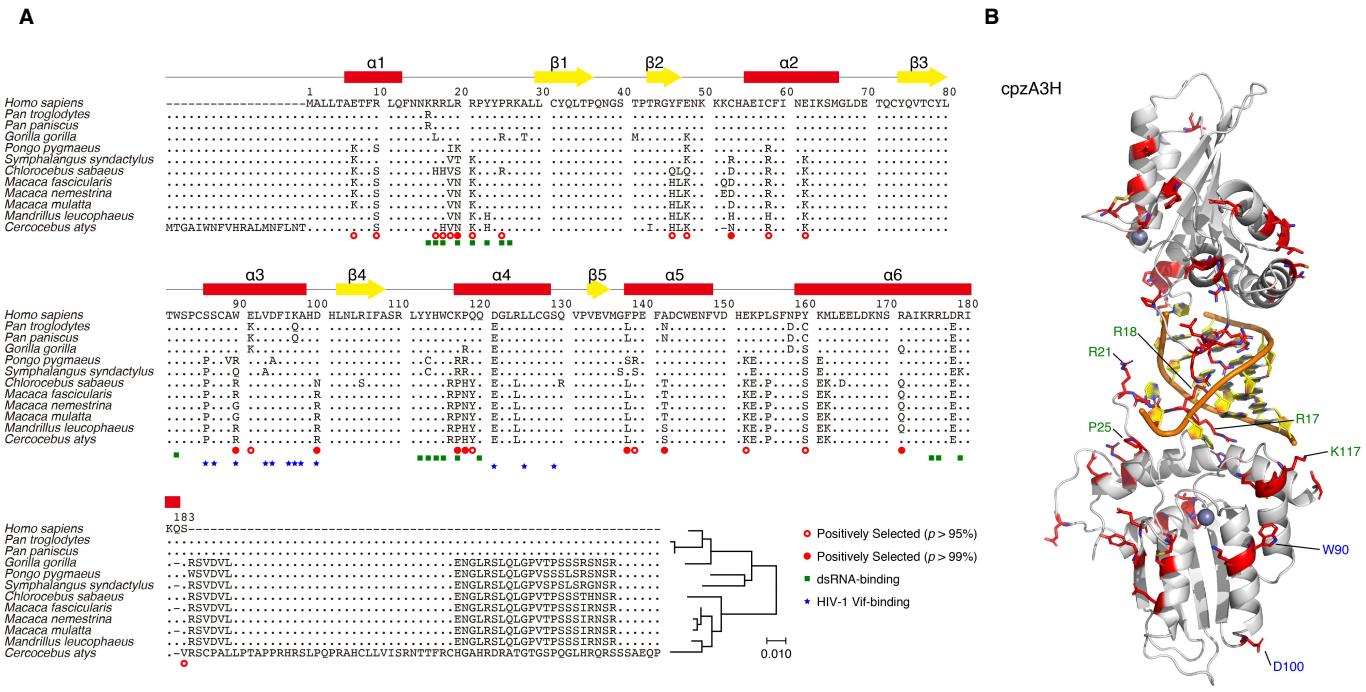
Supplementary Data

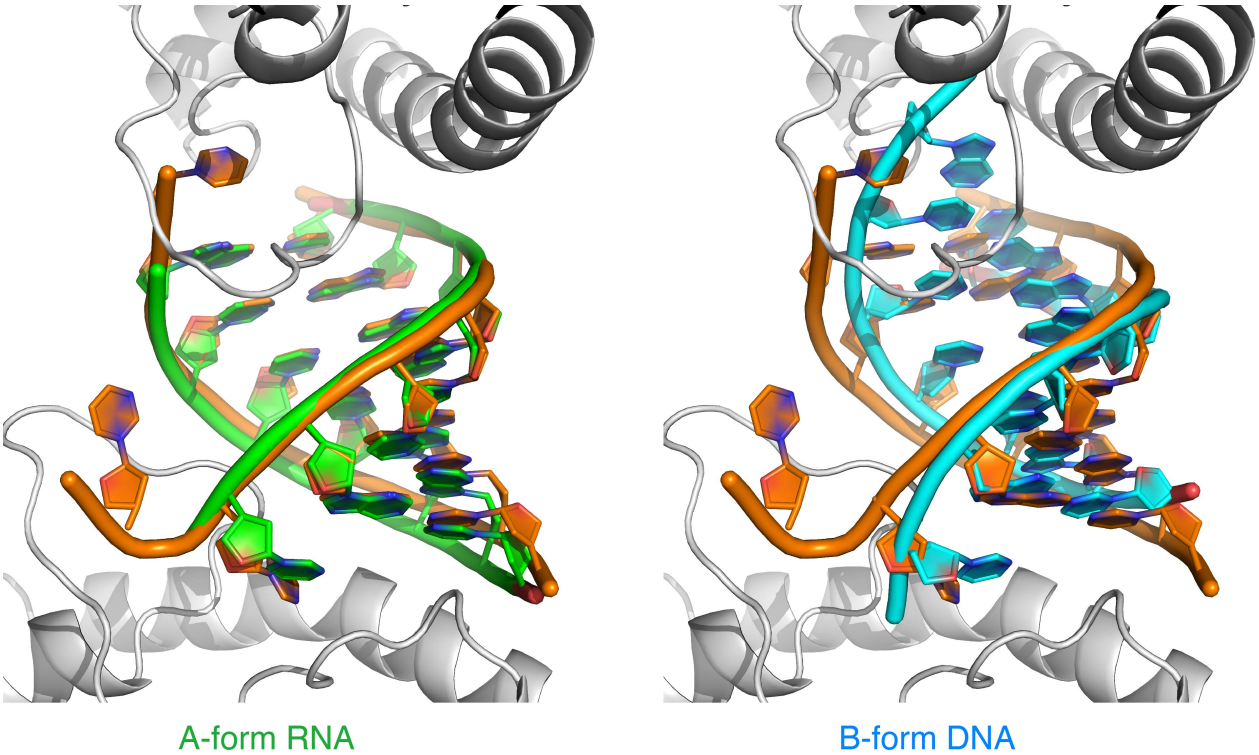
Structural basis of chimpanzee APOBEC3H dimerization stabilized by double-stranded RNA

Tatsuya Matsuoka, Takayuki Nagae, Hirotaka Ode, Hiroaki Awazu, Teppei Kurosawa, Akiko Hamano, Kazuhiro Matsuoka, Atsuko Hachiya, Mayumi Imahashi, Yoshiyuki Yokomaku, Nobuhisa Watanabe, and Yasumasa Iwatani

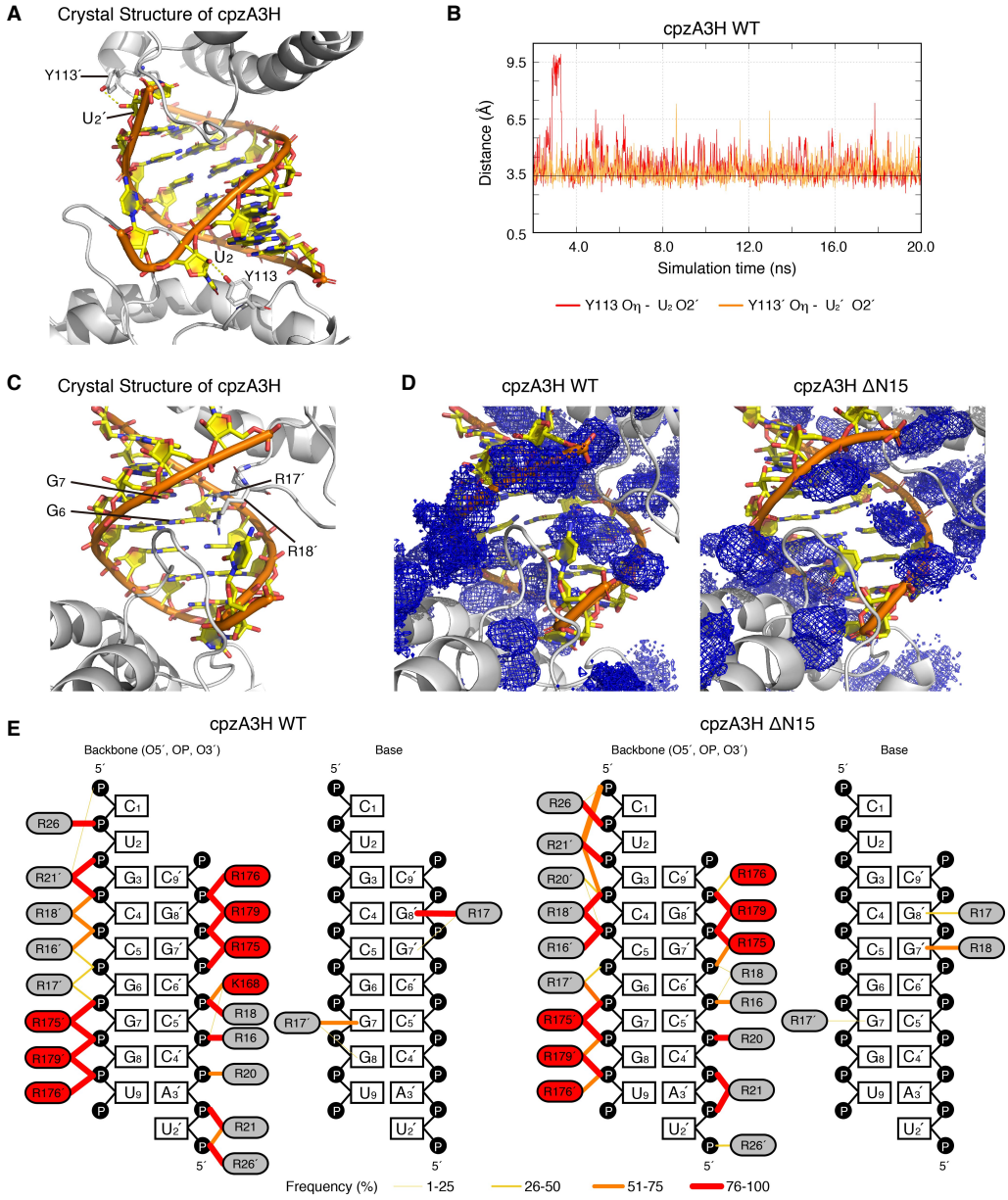
Overview



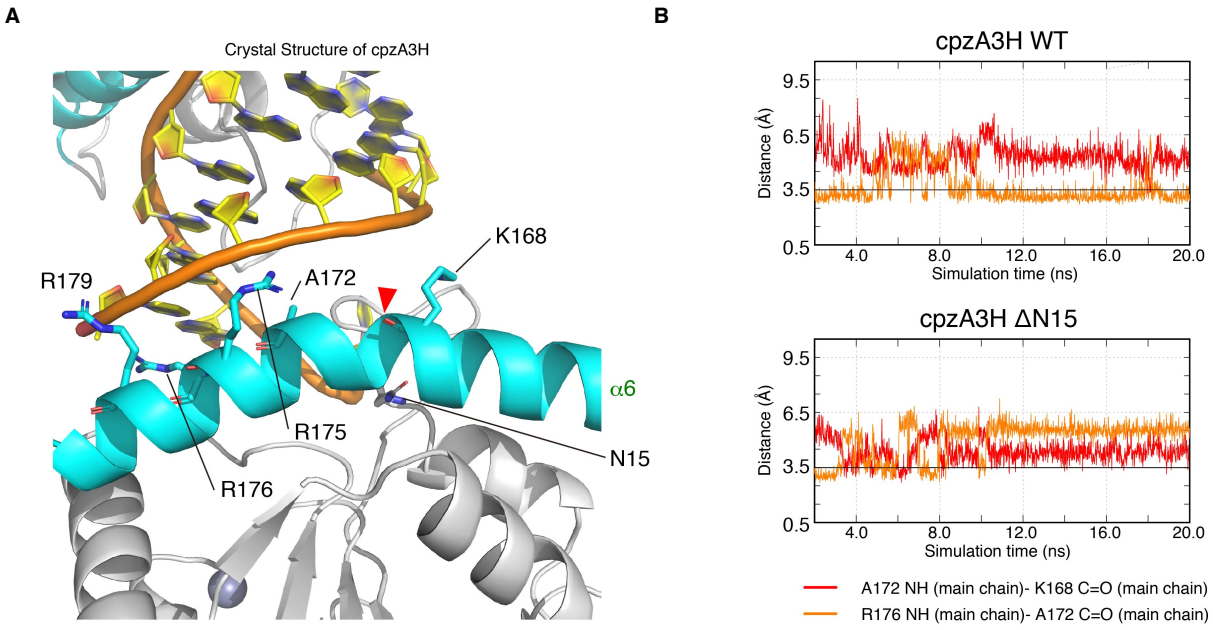
Supplementary Figure S1. Comparisons between the A3H residues critical for dsRNA or Vif interactions and positions under positive selection in primates. (**A**) The amino acid sequence alignments of twelve primate A3Hs with the phylogeny of A3H are shown. The cpzA3H α helices (α1-α6) and ß strands (ß1-ß5) are colored in red and yellow, respectively. Open and closed red circles depict sites under positive selection with a posterior probability of > 0.95 (p > 95%) and >0.99 (p > 99%), respectively. Green squares and blue stars indicate residues responsible for dsRNA binding and Vif binding, respectively. (**B**) Positively selected sites (p > 95%) are mapped and shown using red sticks on the cpzA3H structure. Of these, the residues responsible for dsRNA binding and Vif binding are indicated in green and blue, respectively.



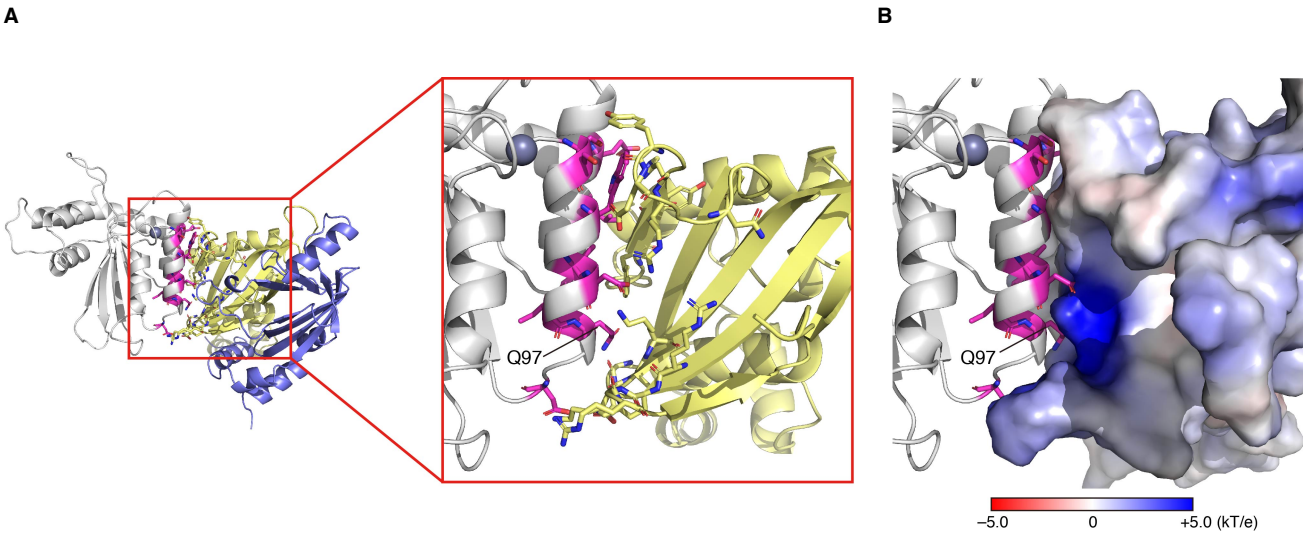
Supplementary Figure S2. cpzA3H dimerization mediated by the A-form RNA duplex. The dsRNA (orange) bound to cpzA3H in our crystal structure was compared with the A-form nucleotide duplex (green) in the left panel and the B-form duplex (light blue) in the right panel. The expected A-form and B-form duplexes were generated with AmberTools17 NAB module (http://ambermd.org/) and were superimposed onto the cpzA3H crystal structure. The main chains of the cpzA3H structure are represented in gray ribbons. The dsRNA root-mean-square deviations (RMSDs) for A-form RNA and B-form DNA were 0.57 Å and 2.40 Å, respectively.



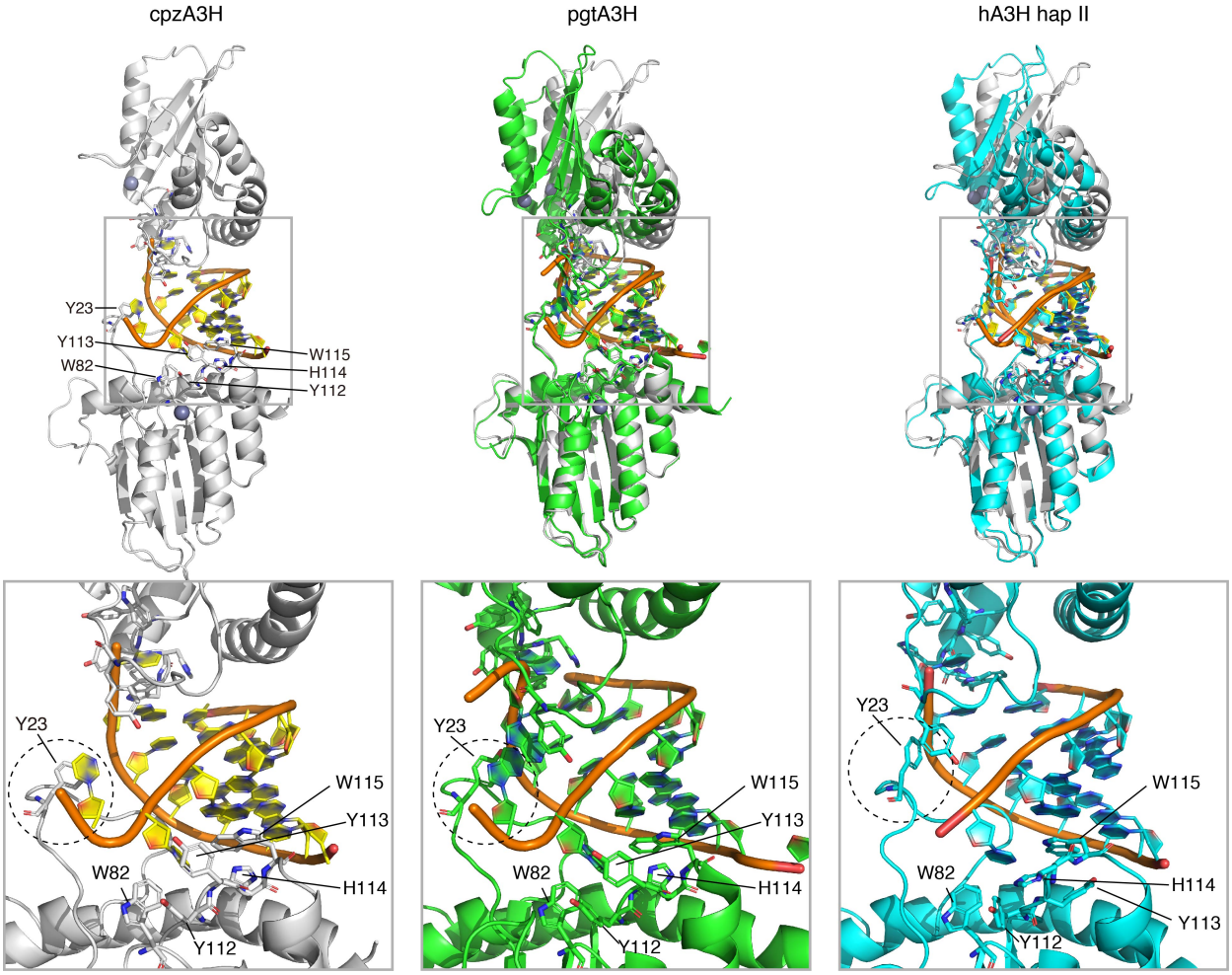
Supplementary Figure S3. cpzA3H residues critical for the dsRNA interaction. (**A**) Hydrogen bonds between the Y113 Oη and the RNA ribose U2 O2′ in the cpzA3H-dsRNA crystal structure. The phosphate backbones of dsRNA are traced with orange lines. The C, O, and N atoms of the dsRNA are colored in yellow, red and blue, respectively. The distances of Y113 Oη -U2 O2′ or Y113′ Oη -U2′ O2′ are both 3.1 Å. (**B**) Time-course analyses of the distances between the Y113 Oη and U2 O2′ atoms by MD simulations. The MD simulation performed for the WT cpzA3H dimer is shown. The maximum distance for hydrogen bonds is highlighted at 3.5 Å with black solid horizontal lines. (**C**) Possible hydrogen bonds formed by R17′ and R18 with G6 and R17′ with G7 in the cpzA3H-dsRNA crystal structure. Of note, the R17′ and R18′ side chains were refined based on the relatively low electron densities for their atoms. The distances of the R17-G6 N7 bond, the R18′-G6 N7 bond and the R17′-G7 N7 bond are close enough to form hydrogen bonds. (**D**) Grid densities of the arginine guanidino group atoms and the lysine Nζ atoms in the cpzA3H-dsRNA complexes. MDs for the WT cpzA3H dimer or the cpzA3H ΔN15 dimer were performed. The predicted densities are represented with blue mesh contoured at 1.5σ. High density for R17′ and R18′ in the major groove of dsRNA was observed in the WT cpzA3H structure (left), whereas low density was observed for the cpzA3H ΔN15 (right). (**E**) Diagrams of the salt bridges and hydrogen bonds between cpzA3H and dsRNA based on the MDs. The dsRNA-binding residues on L1 and α6 are colored with gray and red circles, respectively. The black ellipses and white rectangles depict the phosphate backbone (O5′, ribose O5′; OP, phosphate group; and O3′, ribose O3′) and nucleotide bases.



Supplementary Figure S4. The helix structure of α6 is distorted. (A) The α6 main chain (cyan) is distorted between K168 and A172. Hydrogen bonds of the α6 helix main chain are broken at the red arrowhead. N15 stabilizes the α6 position via a hydrogen bond between the N15 Nδ (side chain) and the L166 O (main chain). The R175, R176, and R179 side chains of α6 also form salt bridges with the RNA phosphate backbones. (B) Instability of the distorted α6 in the apo form of cpzA3H ΔN15 during MD simulations. Distances between the A172 N (main-chain NH) and the K168 O (main-chain C=O) and between the R176 N (main-chain NH) and the A172 O (main-chain C=O) were analyzed during the equilibrated 2.0-20.0-ns periods of the respective MDs to examine the maintenance of the bent α6 conformations for cpzA3H WT (upper) and cpzA3H ΔN15 (lower). The maximum hydrogen bond distance is highlighted at 3.5 Å with a black solid horizontal line. The distance between A172 N and K168 O is stably long in the WT compared with the length between R176 N and A172 O. In contrast, these distances are all similar in cpzA3H ΔN15. These results suggest that the WT retains the helix structure with a distorted α6 between K168 and A172, whereas ΔN15 has no distortion between K168 and A172.



Supplementary Figure S5. Predicted interaction between cpzA3H and the HIV-1 Vif-CBF-ß complex. (**A**) Ribbon representation of the predicted cpzA3H-Vif-CBF-ß complex structure. The structures of cpzA3H, Vif, and CBF-ß are colored in gray, yellow and navy, respectively. The cpzA3H and Vif residues involved in their interaction are represented by magenta and yellow sticks, respectively. (**B**) The electrostatic surface potential of Vif. The accessible surface area is colored according to the calculated potential from −5 kT/e (red) to +5 kT/e (blue). The buried surface area of the interaction was 877.9 Å^2, based on analysis with PDBe PISA v1.52 (http://www.ebi.ac.uk/pdbe/pisa/).



Supplementary Figure S6. Structural comparison of cpzA3H with pgtA3H and hA3H. Ribbon representations of the superimposed structures of cpzA3H-dsRNA (gray) with the pgtA3H-dsRNA (green; PDB 5W3V) or with hA3H hap II-dsRNA (cyan; PDB 6B0B) are shown. All three A3H complexes form a similar dimer structure through the dsRNA. The last base-paired purine rings of dsRNA form a π-stacking interaction with W115. The W115 residue is underpinned by the unique T-shaped "π-π networks" of W82-Y112-Y113-W115 and H114-W115 in the cpzA3H and pgtA3H structures, but not in the hA3H hap II structure. As shown enclosed with the circular dotted line, a unique kinked conformation of the single-stranded end of the RNA alongside Y23 is observed in the cpzA3H and pgtA3H structures, but not in the hA3H hap II structure.

Supplementary Methods

Sequence Analysis of Primate A3Hs

The nucleotide sequences of primate A3Hs were obtained from the GenBank databases for the accession numbers FJ376614.1 (*Homo sapiens*, hap II), EU861357.1 (*Pan troglodytes*), XM_003821609.1 (*Pan paniscus*), EU861358.1 (*Gorilla gorilla*), EU861359.1 (*Pongo pygmaeus*), EU861360.1 (*Symphalangus syndactylus*), XM_007975821.1 (*Chlorocebus sabaeus*), XM_005567258.1 (*Macaca fascicularis*), XM_011712322.1 (*Macaca nemestrina*), NM_001042372.1 (*Macaca mulatta*), XM_011978835.1 (*Mandrillus leucophaeus*), and XM_012032558.1 (*Cercocebus atys*). We generated codon alignments of these sequences with the MEGA7 program (https://www.megasoftware.net/)(1) and then constructed a phylogenic tree using the neighbor-joining method with a compete deletion option. Amino acid residues with high posterior probabilities (greater than 0.95 or 0.99) of having evolved under positive selection were identified using the PAML 4.8 program (http://abacus.gene.ucl.ac.uk/software/paml.html) (2) under an M8 model (3) as positive selection sites.

Docking Simulation of the cpzA3H-Vif-CBF-ß Complex

A structural model of the Vif-CBF-ß complex bound to cpzA3H was predicted using a method similar to that reported by Ooms *et al*. (4) for the hA3H-Vif-CBF-ß complex. Briefly, assignments were made for the A3H and Vif residues critical for their interactions, as previously identified (4,5). Subsequently, the cpzA3H crystal structure was docked into the Vif-CBF-ß complex structure (PDB 4N9F) (6), using the ClusPro 2.0 program (https://cluspro.bu.edu/login.php)(7).

Molecular Dynamics (MD) Simulations

We performed 20.0-ns MD simulations for the cpzA3H WT dimer and the ΔN15 dimer in complex with (as a dimer) or without (as the apo form) dsRNA, using the Amber16 software package (http://ambermd.org/)(8). The initial structure of cpzA3H ΔN15 was constructed by homology modeling with Modeller 9v8 (9) based on our crystal structure of the WT. The residues for cpzA3H models of dimer and monomer were surrounded by ~70,000 and ~13,000 water molecules (the TIP3P model)(10), respectively. To neutralize electric charges of cpzA3H and/or dsRNA, potassium or chloride ions were further placed around residues for the cpzA3H models. The ff14SB(11) was primarily used for energy and force calculations, while the ff99bsc0χOL3(12) and the force field proposed by Pang *et al*.(13) were adopted for energy and force calculations of

dsRNA and zinc ions at the cpzA3H catalytic centers. The constituted molecular system including cpzA3H, water molecules, ions, and/or RNA was subjected to energy minimization to relax unfavorable conformations and/or interactions. Energy minimization was performed only for solvent molecules (water molecules and ions) and subsequently for the whole system. In each minimization process, 10,000 steps of the steepest descent method followed by 10,000 steps of the conjugated gradient method were adopted. The energy-minimized system was heated to 273K under the NTV ensemble condition for 0.02 ns and from 273K to 310K at 1 atm under the NTP ensemble condition for 0.08 ns. During the heating periods, positions of heavy atoms in cpzA3H and RNA were constrained with a harmonic potential of 1.0 kcal/mol/Å^2. Then, the 20.0 ns-timescale MD simulation of the system was performed at 310K at 1 atm under the NTP ensemble condition. For the heating and MD simulation processes, the time step was set at 0.2 fs. Periodic boundary conditions were applied to avoid problems with the boundary effects caused by finite size, while the SHAKE algorithm was used to constrain bond distances involving any hydrogen atoms. The structural characteristics of these molecules were examined from the equilibrated trajectories in 2.0-20.0 ns simulations created by the cpptraj module in the AmberTools17 program (http://ambermd.org/)(14). Hydrogen bond formation was defined here by <3.5 Å of acceptor-donor distance as well as by <135 degrees of the bond angle.

REFERENCES

1. Kumar, S., Stecher, G. and Tamura, K. (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.*, **33**, 1870-1874.
2. Yang, Z. (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.*, **24**, 1586-1591.
3. Yang, Z., Wong, W.S. and Nielsen, R. (2005) Bayes empirical bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.*, **22**, 1107-1118.
4. Ooms, M., Letko, M. and Simon, V. (2017) The Structural Interface between HIV-1 Vif and Human APOBEC3H. *J. Virol.*, **91**, e02289-02216.
5. Nakashima, M., Tsuzuki, S., Awazu, H., Hamano, A., Okada, A., Ode, H., Maejima, M., Hachiya, A., Yokomaku, Y., Watanabe, N. *et al.* (2017) Mapping Region of Human Restriction Factor APOBEC3H Critical for Interaction with HIV-1 Vif. *J. Mol. Biol.*, **429**,

1262-1276.

6. Guo, Y., Dong, L., Qiu, X., Wang, Y., Zhang, B., Liu, H., Yu, Y., Zang, Y., Yang, M. and Huang, Z. (2014) Structural basis for hijacking CBF-beta and CUL5 E3 ligase complex by HIV-1 Vif. *Nature*, **505**, 229-233.
7. Kozakov, D., Hall, D.R., Xia, B., Porter, K.A., Padhorny, D., Yueh, C., Beglov, D. and Vajda, S. (2017) The ClusPro web server for protein-protein docking. *Nat. Protoc.*, **12**, 255-278.
8. Case, D.A., Cheatham, T.E., 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K.M., Jr., Onufriev, A., Simmerling, C., Wang, B. and Woods, R.J. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem.*, **26**, 1668-1688.
9. Sali, A. and Blundell, T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.*, **234**, 779-815.
10. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. and Klein, M.L. (1983) Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, **79**, 926-935.
11. Maier, J.A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K.E. and Simmerling, C. (2015) ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.*, **11**, 3696-3713.
12. Zgarbova, M., Otyepka, M., Sponer, J., Mladek, A., Banas, P., Cheatham, T.E., 3rd and Jurecka, P. (2011) Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theory Comput.*, **7**, 2886-2902.
13. Tang, J., Park, J.G., Millard, C.B., Schmidt, J.J. and Pang, Y.P. (2007) Computer-aided lead optimization: improved small-molecule inhibitor of the zinc endopeptidase of botulinum neurotoxin serotype A. *PLoS One*, **2**, e761.
14. Roe, D.R. and Cheatham, T.E., 3rd. (2013) PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.*, **9**, 3084-3095.