Supplementary Information:

Ab Initio HIV-1 Tat-TAR Interactions Study using Hybrid Scoring-Enhanced Molecular Modeling Across Subtypes

Chengwei Zeng[1‡], Huiwen Wang[2‡], Jiaming Gao[1], Haoquan Liu[1], Chen Zhuo[1], Yunjie Zhao*[1]

*Corresponding authors

‡These authors contributed equally

[1]Institute of Biophysics and Department of Physics, Central China Normal University, Wuhan 430079, China

[2]School of Physics and Engineering, Henan University of Science and Technology, Luoyang 471023, China.

E-mail: yjzhaowh@ccnu.edu.cn

Supplementary Table 1. The templates used for structure prediction of the nine HIV-1 Tat subtypes with Swiss-model.

	Tat	Isolate	Entry	Origin length (AA)	Final Length (AA)	Structure	Template	Identity(%)	Coverage(%)	GMQE
HIV-1 group M	Subtype A	MAL	P04613	87	87	1K5K	/	/	/	/
	Subtype B	HXB2	P04608	86	86	1JFW	/	/	/	/
	Subtype C	92BR025	O12161	101	86	/	1TBC	72.09	100	0.61
	Subtype D	Z2/CDC Z34	P12506	86	86	1TBC	/	/	/	/
	Subtype F	VI850	Q9QSR0	101	86	/	1TBC	70.93	100	0.61
	Subtype G	SE6165	O89943	101	86	/	1TBC	72.09	100	0.61
	Subtype H	VI991	Q9Q717	86	86	/	1TBC	72.09	100	0.61
	Subtype J	SE9173	Q9WC66	101	86	/	1TBC	72.09	100	0.61
	Subtype K	96CM-MP535	P0C1K4	101	86	/	1TBC	72.09	100	0.61

Supplementary Table 2. The compilation of all methodological pipelines and their associated resource links.

Methods	Links
MAFFT	https://mafft.cbrc.jp/alignment/software
PSIPRED	http://bioinf.cs.ucl.ac.uk/psipred
DISOPRED3	http://bioinf.cs.ucl.ac.uk/psipred
Swiss-model	https://swissmodel.expasy.org
SAVES	https://saves.mbi.ucla.edu
ProSA-web	https://prosa.services.came.sbg.ac.at/prosa.php
ModRefiner	https://zhanggroup.org/ModRefiner
HADDOCK	https://rascar.science.uu.nl/haddock2.4
ITScore-PR	https://zoulab.dalton.missouri.edu/resources.html
3dRPC-Score	http://biophy.hust.edu.cn/new/3dRPC
DRPScore	https://github.com/Zhaolab-GitHub/DRPScore_v1.0
PLIP	https://plip-tool.biotec.tu-dresden.de/plip-web/plip/index

Supplementary Table 3. The detailed rankings provided by three scoring functions (ITScore-PR, 3dRPC-Score, and DRPScore) for the nine HIV-1 Tat subtypes.

Subtype	Scoring function	Top 1	Top 2	Top 3	Top 4	Top 5	Top 6	Top 7	Top 8	Top 9	Top 10	Final Pose
Subtype A	ITScore-PR	2-2	2-1	1-2	1-1	1-3	4-3	4-2	2-3	2-4	4-1	2-2
	3dRPC-Score	2-2	2-1	1-1	4-1	1-4	4-4	2-3	4-2	1-2	6-3	
	DRPScore	10-4	7-3	7-2	5-3	2-2	5-2	7-1	6-4	6-3	8-1	
Subtype B	ITScore-PR	10-1	2-1	2-2	6-2	7-1	10-3	2-3	2-4	7-2	10-2	2-2
	3dRPC-Score	2-1	2-3	2-4	2-2	10-1	7-2	4-2	1-3	1-2	4-3	
	DRPScore	5-4	5-3	2-2	3-4	6-3	5-2	6-4	8-4	4-1	10-1	
Subtype C	ITScore-PR	2-2	2-1	2-3	2-4	1-2	1-4	1-3	4-3	4-2	1-1	2-1
	3dRPC-Score	2-1	1-4	2-4	2-3	2-2	7-4	7-1	6-1	8-2	1-2	
	DRPScore	5-4	8-2	8-3	5-2	8-1	8-4	4-3	3-2	5-3	6-2	
Subtype D	ITScore-PR	1-4	1-2	1-1	2-4	1-3	2-2	3-1	2-1	3-4	3-2	1-2
	3dRPC-Score	1-2	1-3	2-4	1-1	1-4	2-2	2-1	3-3	3-1	3-4	
	DRPScore	1-2	2-3	4-4	4-3	4-2	1-3	3-3	1-1	5-3	3-1	
Subtype F	ITScore-PR	8-1	8-2	4-1	4-4	4-2	2-1	2-4	2-3	4-3	3-3	8-1
	3dRPC-Score	5-1	4-4	8-1	4-2	4-1	8- 2	4-3	5-3	8-3	3-3	
	DRPScore	7-2	6-2	7-1	7-3	6-1	1-4	7-4	1-1	1-2	1-3	
Subtype G	ITScore-PR	14-2	6-4	14-3	10-1	6-3	5-2	14-1	14-4	7-3	7-2	10-2
	3dRPC-Score	10-2	10-4	10-1	14-2	5-4	5-1	14-1	14-3	10-3	3-2	
	DRPScore	10-2	2-3	11-2	5-2	11-4	8-2	4-4	2-1	8-4	11-1	
Subtype H	ITScore-PR	1-1	8-3	3-1	2-4	1-3	2-1	8-1	2-2	7-1	6-2	1-1
	3dRPC-Score	1-1	1-2	1-4	1-3	8-3	8-1	7-1	8-2	7-3	2-4	
	DRPScore	1-2	5-1	5-4	3-4	3-2	5-2	5-3	4-1	4-4	4-3	
Subtype J	ITScore-PR	1-1	1-4	1-2	5-2	1-3	6-1	5-1	5-3	7-1	6-2	1-1
	3dRPC-Score	7-1	1-4	1-1	7-4	10-2	7-3	10-1	9-2	6-1	2-2	
	DRPScore	4-4	11-2	4-2	4-1	4-3	1-1	6-1	6-4	10-3	5-1	
Subtype K	ITScore-PR	10-1	2-4	10-2	2-1	2-3	2-2	3-1	9-3	3-4	7-1	2-3
	3dRPC-Score	2-3	3-4	3-2	2-4	2-2	3-1	3-3	5-3	7-3	7-1	
	DRPScore	6-1	6-3	6-2	4-3	1-1	6-4	8-4	1-4	9-4	1-3	

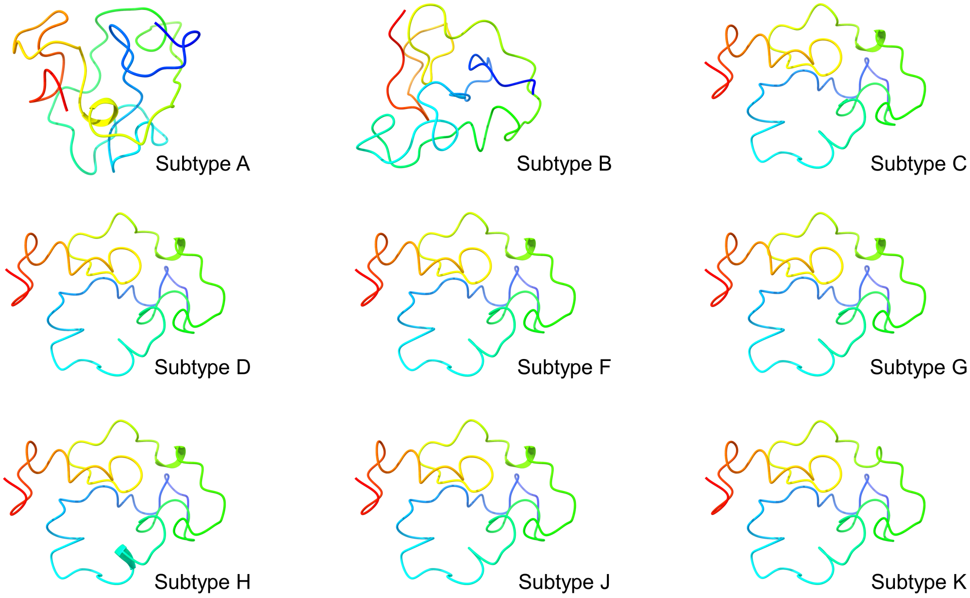
Supplementary Table 4. The interaction between Tat and TAR analyzed using the Protein-Ligand Interaction Profiler (PLIP) for the nine HIV-1 Tat subtypes.

Tat	H-Bonds (nucleotides)	π-Cation interactions (nucleotides)	Salt Bridges (nucleotides)	Total number
Subtype A	ARG53(C39), GLN54(A22), ARG55(G26), ARG56(C41, U40) Total: 7	ARG53(G26) Total: 1	ASP2(G26), LYS50(U25, C24), ARG52(C24, U23), ARG56(U42, C41) Total: 4	12
Subtype B	GLY48(C41), ARG49(A22, C41), LYS51(U40, C41), ARG52(G26) Total: 7	ARG52(G26) Total: 2	LYS50(C41, U40), LYS51(C41, U40), LYS51(U40, C39), ARG56(G26, U25) Total: 4	13
Subtype C	ARG49(A22), LYS51(C41), ARG52(G26, U40), ARG55(U25), SER57(G26) Total: 8	ARG52(U40) Total: 1	LYS50(C24, U23) Total: 1	10
Subtype D	ARG49(A22), LYS51(C41), ARG52(U40, G26), ARG55(C24, G26) Total: 6	N/A	LYS50(U23, A22), ARG55(U25, C24), ARG57(U40, C39), ARG57(C41, U40) Total: 4	10
Subtype F	LYS51(C41), ARG52(U40), ARG55(G26), ARG57(G26, U40), THR58(A27) Total: 8	ARG52(G26) Total: 1	LYS50(U23, A22), ARG55(U25, C24), ARG55(G26, U25) Total: 3	12
Subtype G	SER46(U23), ARG49(U23), ARG56(G26), PRO68(G26), LYS71(G28), GLN72(C29) Total: 6	LYS51(G26) Total: 2	ARG52(G26, U25), ASP67(G26) Total: 2	10
Subtype H	HIS33(U40), GLY48(A22), LYS51(G26), GLY57(G26), THR58(A27), LYS71(G28) Total: 6	N/A	HIS33(C41, U40), ARG52(U25, C24), ARG52(G26, U25), LYS60(G28, A27), LYS71(C29, G28) Total: 5	11
Subtype J	GLY48(A22), LYS51(C41), ARG52(C39, U40), SER57(G26), ASP67(U40) Total: 6	ARG52(G26) Total: 1	ARG55(U25, C24), ARG55(G26, U25) Total: 2	9
Subtype K	ARG49(A22), KYS50(U23), ARG52(G26, U40), ARG53(C24), THR57(G26), THR58(A27, G26) Total: 9	ARG52(G26) Total: 1	LYS50(C24, U23), ARG55(U25, C24), ARG55(G26, U25) Total: 3	13

Supplementary Table 5. The Average Backbone Root Mean Square Deviation (RMSD) and Standard deviation for HIV-1 Tat subtypes A, B, and C over 500 ns of molecular dynamics (MD) simulations.

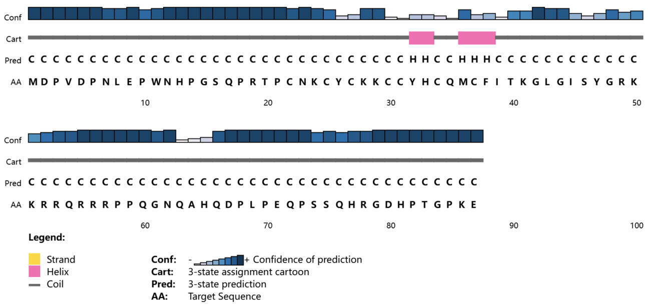
	Subtype A		Subtype B		Subtype C	
Time	Average RMSD (nm)	Standard deviation (nm)	Average RMSD (nm)	Standard deviation (nm)	Average RMSD (nm)	Standard deviation (nm)
0-50	0.46	0.10	0.44	0.09	0.50	0.10
50-100	0.58	0.05	0.58	0.06	0.70	0.04
100-150	0.55	0.02	0.66	0.05	0.74	0.03
150-200	0.54	0.02	0.63	0.03	0.76	0.03
200-250	0.57	0.03	0.60	0.03	0.79	0.01
250-300	0.61	0.02	0.57	0.02	0.88	0.03
300-350	0.59	0.02	0.54	0.02	0.89	0.03
350-400	0.64	0.02	0.56	0.03	0.84	0.03
400-450	0.65	0.04	0.56	0.02	0.84	0.02
450-500	0.60	0.01	0.57	0.02	0.84	0.02

Supplementary Figure 1. The structural models of HIV-1 Tat subtypes (A, B, C, D, F, G, H, J, and K) predicted by Swiss-model. These models were used for subsequent analyses of Tat-TAR interactions and binding dynamics.

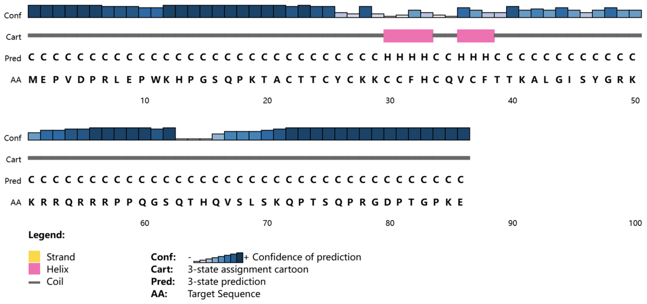


Supplementary Figure 2. The secondary structure prediction for the nine HIV-1 Tat subtypes from PSIPRED.

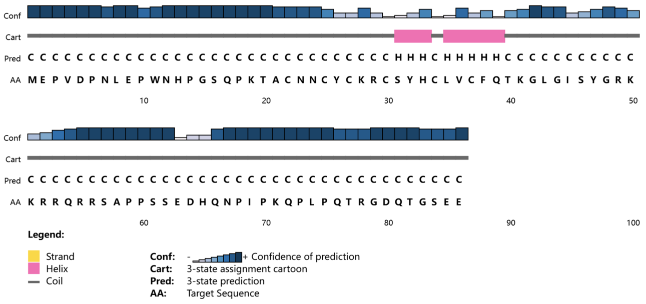
Subtype A



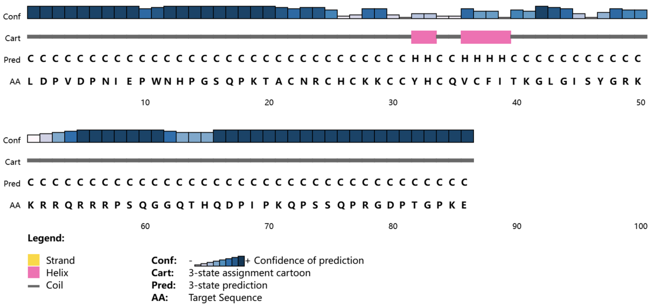
Subtype B



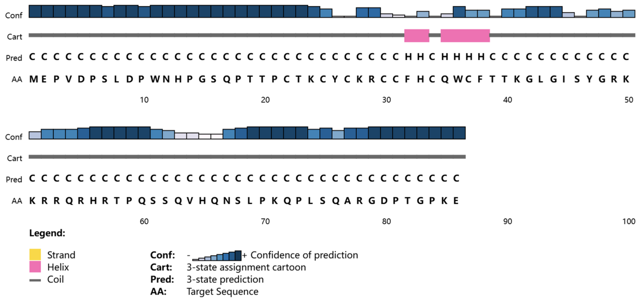
Subtype C



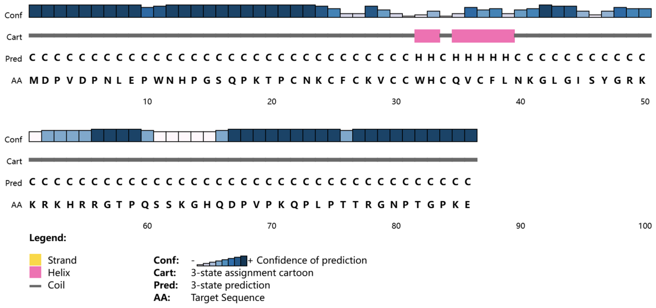
Subtype D



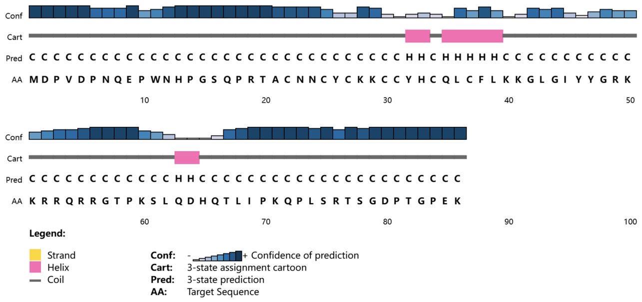
Subtype F



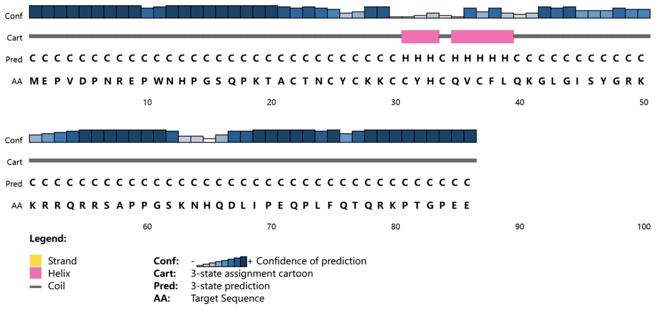
Subtype G



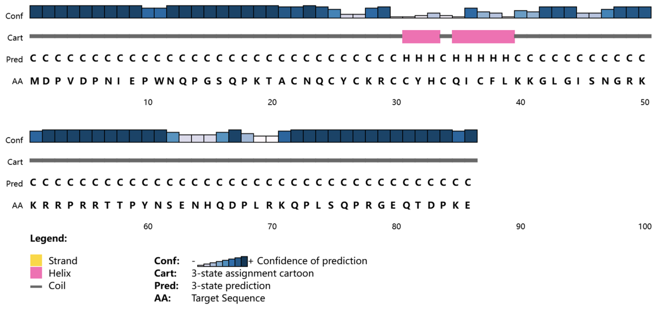
Subtype H



Subtype J

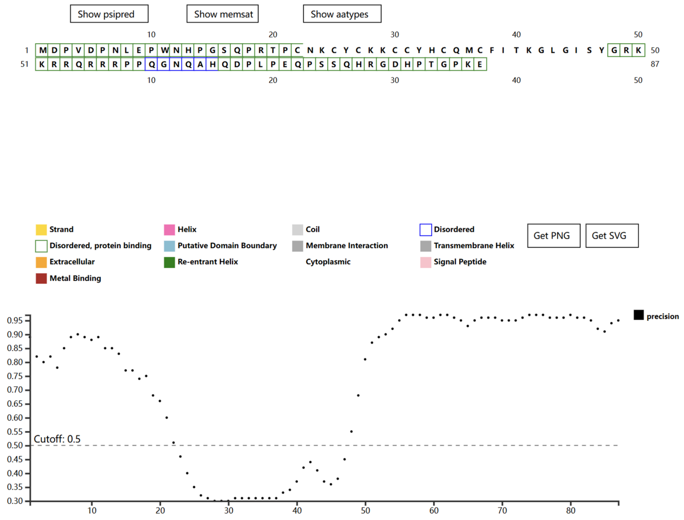


Subtype K



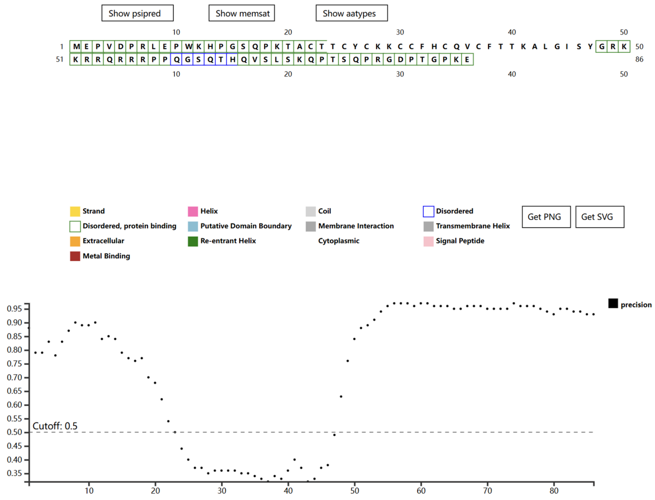
Supplementary Figure 3. The disordered state prediction for the nine HIV-1 Tat subtypes from DISOPRED3.

Subtype A



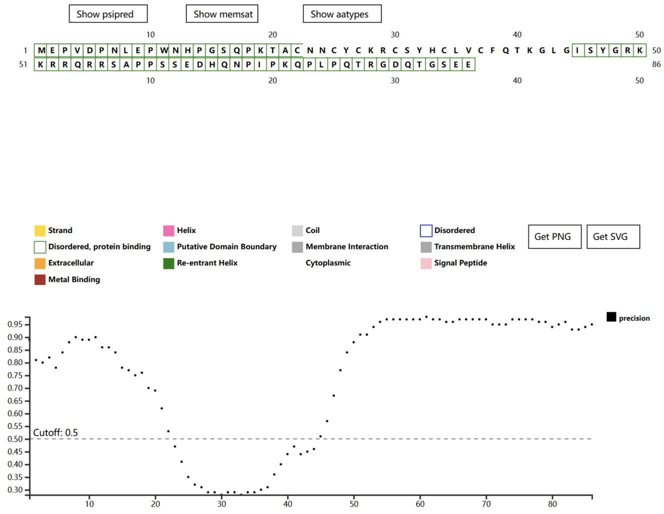
```
1  MDPVDPNLEPWNHPGSQPRTPCNKCYCKKCCYHCQMCFITKGLGISYGRK 50
51 KRRQRRRPPQGNQAHQDPLPEQPSSQHRGDHPTGPKE 87
```

Subtype B



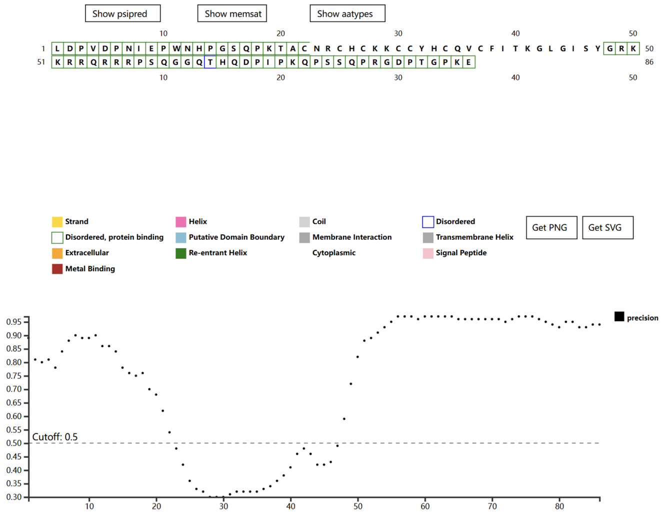
```
1  MEPVDPRLEPWKHPGSQPKTACTTCYCKKCCFHCQVCFTTKALGISYGRK 50
51 KRRQRRRPPQGSQTHQVSLSKQPTSQPRGDPTGPKE 86
```

Subtype C



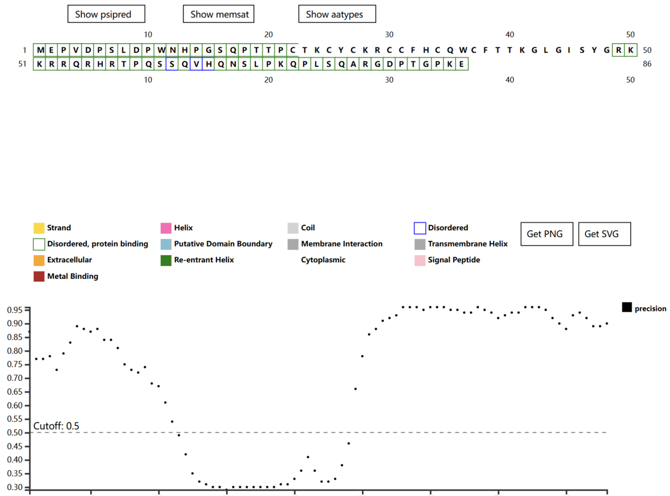
```
1  MEPVDPNLEPWNHPGSQPKTACNNCYCKRCSYHCLVCFQTKGLGISYGRK 50
51 KRRQRRSAPPSSEDHQNPIPKQPLPQTRGDQTGSEE 86
```

Subtype D



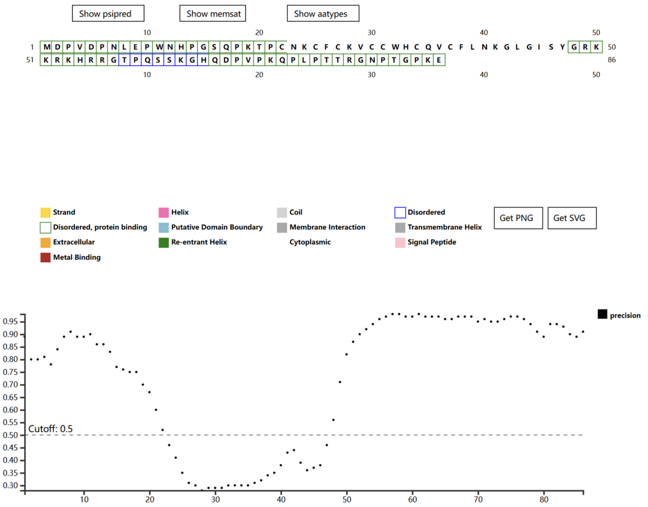
```
1  LDPVDPNIEPWNHPGSQPKTACNRCHCKKCCYHCQVCFITKGLGISYGRK 50
51 KRRQRRRPSQGGQTHQDPIPKQPSSQPRGDPTGPKE 86
```

Subtype F



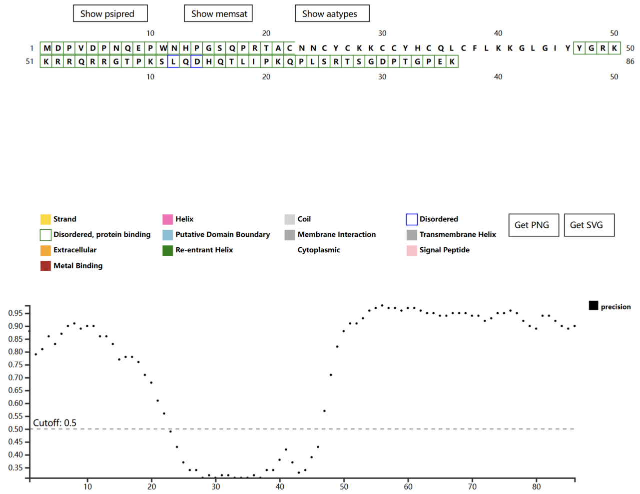
```
1  MEPVDPSLDPWNHPGSQPTTPCTKCYCKRCCFHCQWCFTTKGLGISYGRK 50
51 KRRQRHRTPQSSQVHQNSLPKQPLSQARGDPTGPKE 86
```

Subtype G



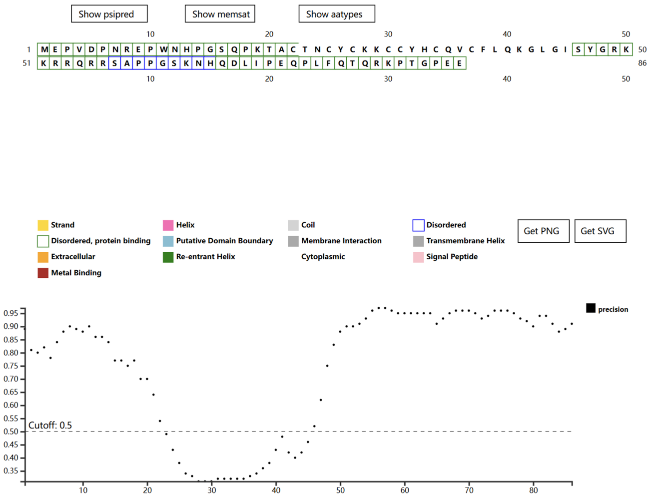
```
1  MDPVDPNLEPWNHPGSQPKTPCNKCFCKVCCWHCQVCFLNKGLGISYGRK 50
51 KRKHRRGTPQSSKGHQDPVPKQPLPTTRGNPTGPKE 86
```

Subtype H



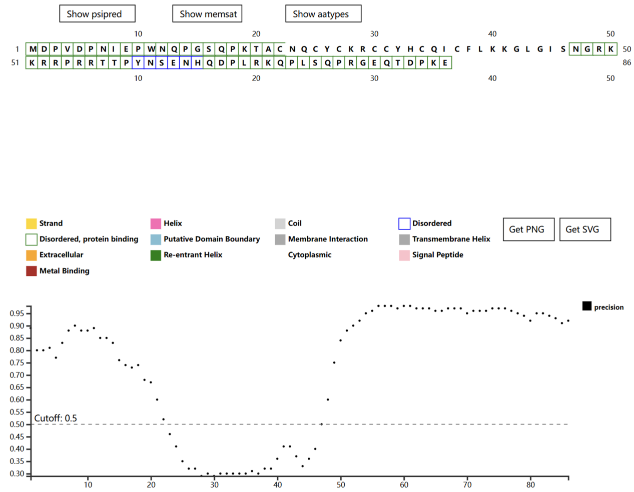
1 MDPVDPNQEPWNHPGSQPRTACNNCYCKKCCYHCQLCFLKKGLGIYYGRK 50
51 KRRQRRGTPKSLQDHQTLIPKQPLSRTSGDPTGPEK 86

Subtype J



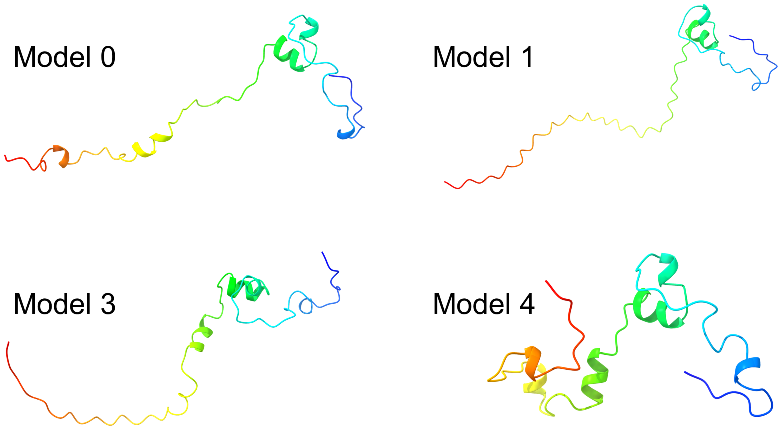
1 MEPVDPNREPWNHPGSQPKTACTNCYCKKCCYHCQVCFLQKGLGISYGRK 50
51 KRRQRRSAPPGSKNHQDLIPEQPLFQTQRKPTGPEE 86

Subtype K

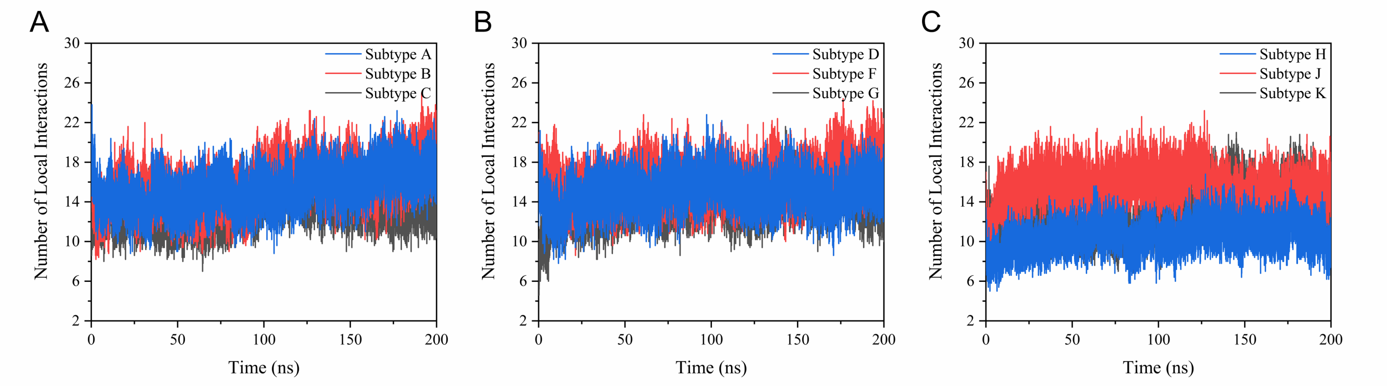


1 MDPVDPNIEPWNQPGSQPKTACNQCYCKRCCYHCQICFLKKGLGISNGRK 50
51 KRRPRRTTPYNSENHQDPLRKQPLSQPRGEQTDPKE 86

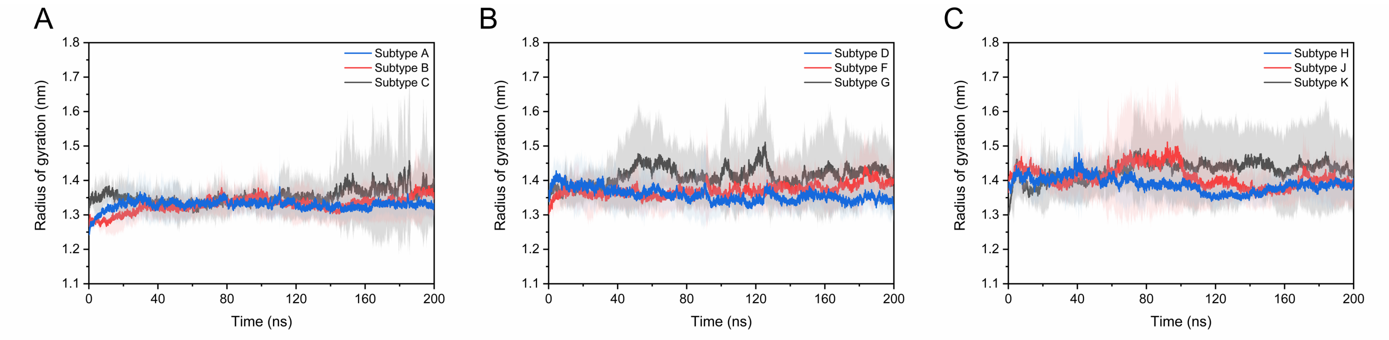
Supplementary Figure 4. The 5 structural models of HIV-1 Tat subtype C predicted by Alphafold3.



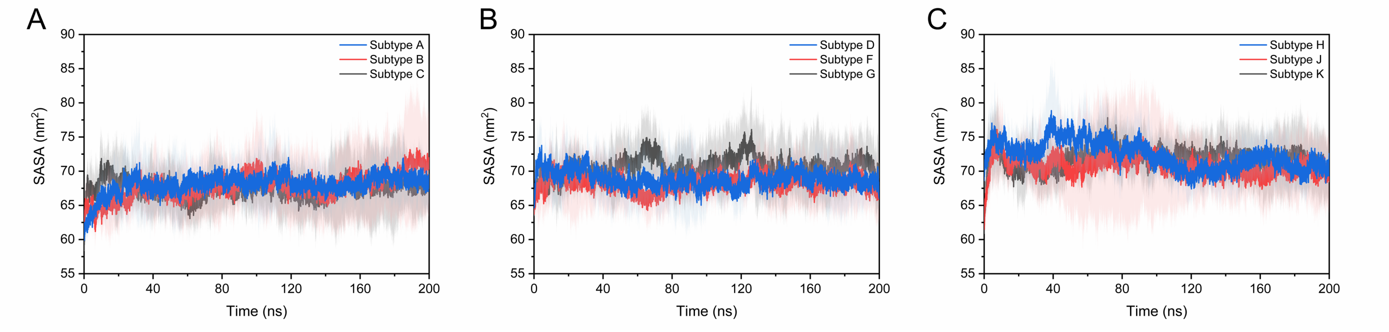
Supplementary Figure 5. The average number of pairs within 0.35 nm calculated from five independent simulations for the nine HIV-1 Tat subtypes over 200 ns molecular dynamics (MD) simulations. (A) The number of pairs within 0.35 nm for HIV-1 Tat subtypes A (blue curve), B (red curve), and C (black curve). (B) The number of pairs within 0.35 nm for HIV-1 Tat subtypes D (blue curve), F (red curve), and G (black curve. (C) The number of pairs within 0.35 nm for HIV-1 Tat subtypes H (blue curve), J (red curve), and K (black curve).



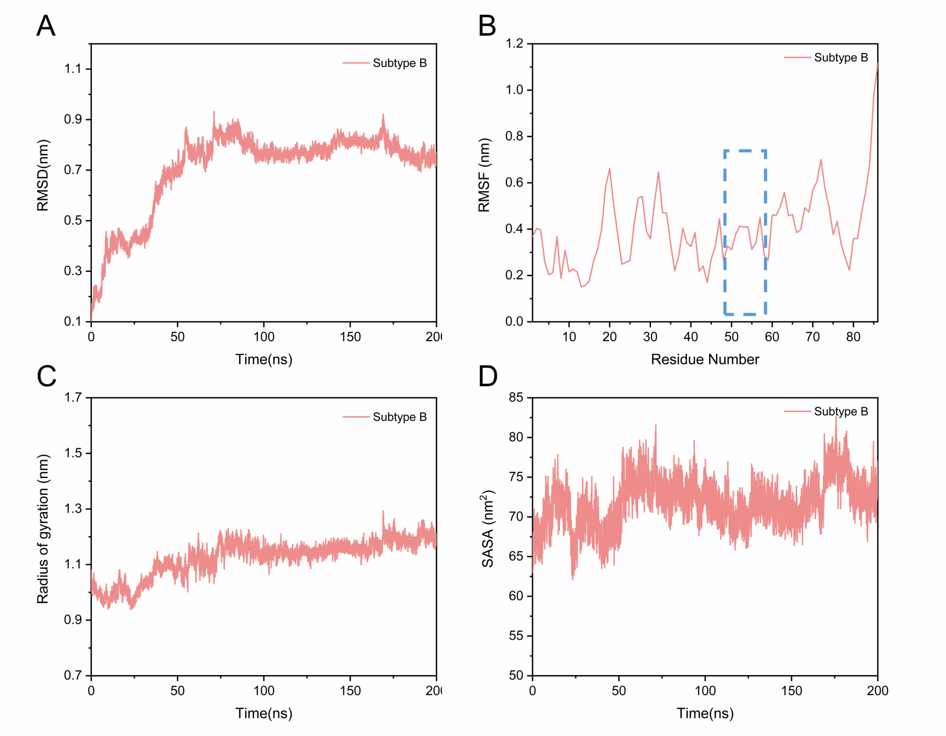
Supplementary Figure 6. The average Radius of gyration (Rg) and standard deviations calculated from five independent simulations for the nine HIV-1 Tat subtypes over 200 ns molecular dynamics (MD) simulations. (A) The Rg for HIV-1 Tat subtypes A (red curve), B (blue curve), and C (green curve). (B) The Rg for HIV-1 Tat subtypes D (red curve), F (blue curve), and G (green curve. (C) The Rg for HIV-1 Tat subtypes H (red curve), J (blue curve), and K (green curve).



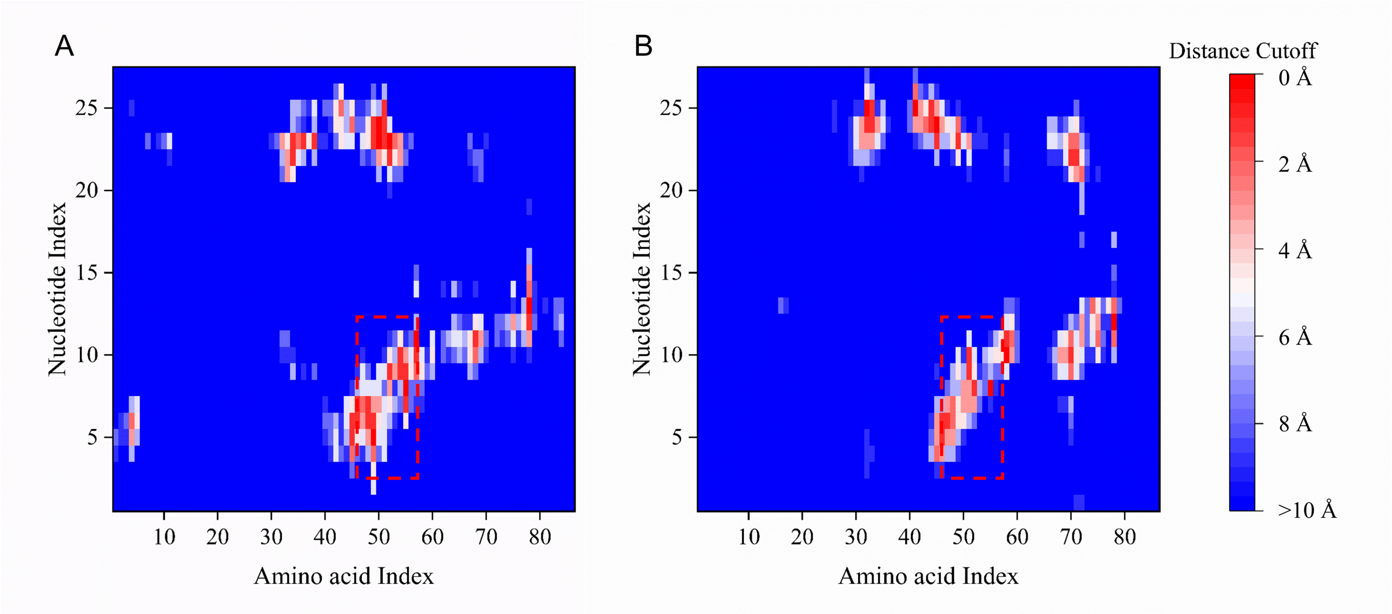
Supplementary Figure 7. The average Solvent Accessible Surface Area (SASA) and standard deviations calculated from five independent simulations for the nine HIV-1 Tat subtypes over 200 ns molecular dynamics (MD) simulations. (A) The SASA for HIV-1 Tat subtypes A (red curve), B (blue curve), and C (green curve). (B) The SASA for HIV-1 Tat subtypes D (red curve), F (blue curve), and G (green curve. (C) The SASA for HIV-1 Tat subtypes H (red curve), J (blue curve), and K (green curve).



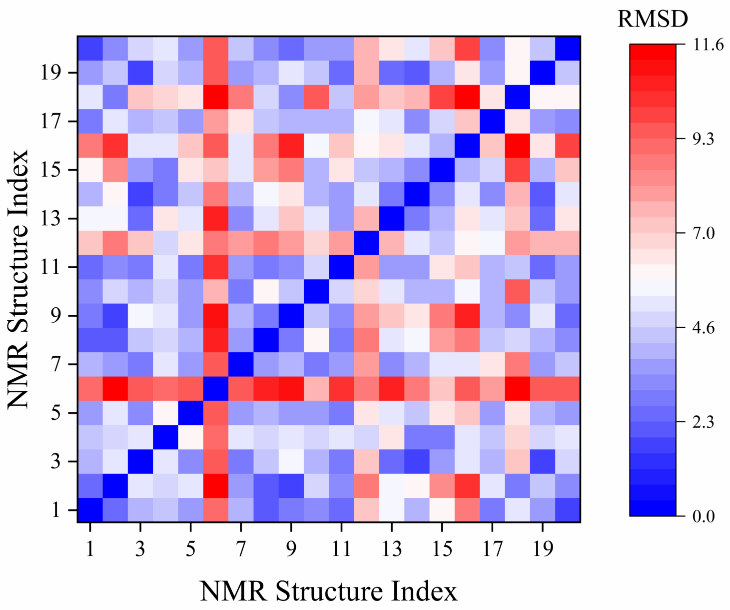
Supplementary Figure 8. The (A) Backbone Root Mean Square Deviation (RMSD), (B) Root Mean Square Fluctuation (RMSF), (C) Radius of gyration (Rg), and (D) Solvent Accessible Surface Area (SASA) for the HIV-1 Tat subtype B over 200 ns molecular dynamics without three hybrid scoring functions.



Supplementary Figure 9. Comparison of the binding interfaces of HIV-1 Tat subtype D (A) and subtype G (B) with TAR RNA. Red circles highlight the conserved core binding region and the color scale from red to blue denotes inter-residue distances ranging from 0 Å to over 10 Å.



Supplementary Figure 10. Pairwise RMSD matrix among the 20 NMR conformers of TAR RNA (PDB ID: 1ANR).



Supplementary Figure 11. Comparison between MD sampled RNA conformations and NMR ensembles. (A-C) Distribution of minimum RMSD between each MD sampled RNA conformation and the 20 NMR conformers.