Electronic Supplementary Material (ESI) for RSC Advances.

Supplementary materials to

Dual Mechanism of HIV-1 Integrase and RNase H Inhibition by Diketo Derivatives – A Computational Study

Poongavanam Vasanthanathan[¶*], N. S. Hari Narayana Moorthy[§], Jacob Kongsted[¶]

[¶] Department of Physics, Chemistry and Pharmacy, University of Southern Denmark, Odense M, Denmark, [§] Departamento de Química e Bioquímica, Faculdade de Ciências, Universidade do Porto, 687, Rua do Campo Alegre, 4169-007 Porto, Portugal

* Corresponding author: Vasanthanathan Poongavanam (nathan@sdu.dk), Tel: +45 65502304, Fax: +45 66158760.

Docking Procedure

Preparation of RNase H model

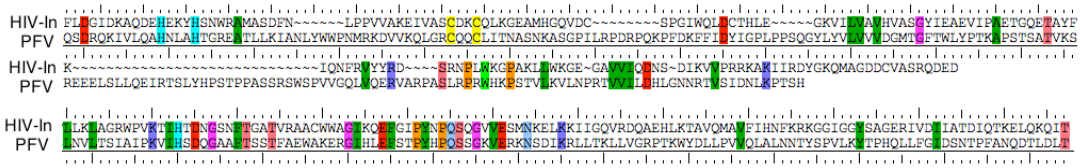
A computational model of the HIV-1 RT associated RNase H domain was constructed from an X-ray crystal structure with resolution of 1.4 Å from the Protein Data Bank (PDB ID: 3QIO). [1] In the deposited crystal structure of RNase H domain with N-hydroxy quinazolinediones (bound active site inhibitor), 12 residues were missing and the structure was determined with manganese (Mn^{++}) ions instead of the two catalytically active magnesium ions. Atomic coordinates for the missing residues were generated using the Swiss-Model[2]. Subsequently, the protein model was imported into the Maestro module available in the Schrödinger package and the protein was further optimized using the Protein Preparation Wizard[3]. This optimization includes adding hydrogen atoms, assigning correct bond orders and building di-sulfide bonds and replacing the Mn^{+2} ions with Mg^{+2} ions. The protonation states of all of the ionizable residues were predicted by PROPKA [4] provided in the Protein Preparation Wizard in the presence of the Mg^{+2} ions at the active site. An optimized structure model was derived by energy minimization (only hydrogen atoms) using the OPLS2005 force field. The receptor grid generation module of Glide [5] was used to define the active site for the docking experiments. As this protein model has a bound ligand (3-hydroxy 6-(phenylsulfonyl) quinazoline-2,4(1H,3H)-dione), the ligand was set as the centroid of the grid box (size of the active site is 20 Å from ligand position). Water molecules in the active site beyond 3 Å from the bound ligand were deleted.

Preparation of HIV-1 Integrase model

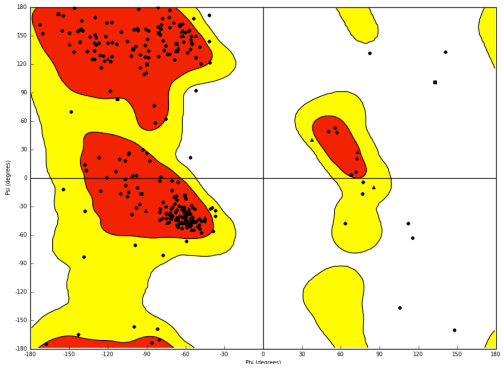
The model of the HIV-1 RT associated RNH domain was constructed from an X-ray crystal structure (resolution of 1.4 Å) obtained from the Protein Data Bank (PDB ID: 3QIO)[1] as shown in our previous study. [6] Since the appearance of a full-length IN crystal structure based on the PFV (prototype foamy virus) in complex with DNA with resolution of 2.65 Å (PDB ID: 3OYA),[7, 8] there has been many homology models reported and used for drug design.[9, 10] In the present study, we modeled the full-length IN structure based on the PFV structure reported by Hare et al.[7] Briefly, the HIV-1 IN sequence (288 amino acids) was obtained from UniProt.[11] This sequence was imported into Prime (version 3.4, Schrödinger, LLC, New York, NY, 2013), a homology modeling tool from Schrödinger, and the structure 3OYA was used as a template to build a homology model. The sequence of IN was aligned with the PFV structure as previously reported.[7, 12] In the secondary structure prediction, the bound ligand (RZL), magnesium ions and three water molecules, which lies close to the magnesium ions, were also included. Models were constructed using a Knowledge based method (construct insertion and

close the gaps based on the known structure). Subsequently the final model was used for the optimization process in the Protein Preparation Wizard as implemented in Schrödinger.[3] This protein structure optimization includes adding hydrogen atoms, assigning correct bond orders and building of di-sulfide bonds. The protonation states of all the ionizable residues were predicted by PROPKA[4] provided in the Protein Preparation Wizard in the presence of the magnesium ions at the active site. Finally, the optimized model was energy minimized (only hydrogen atoms) using the OPLS_2005 force field.

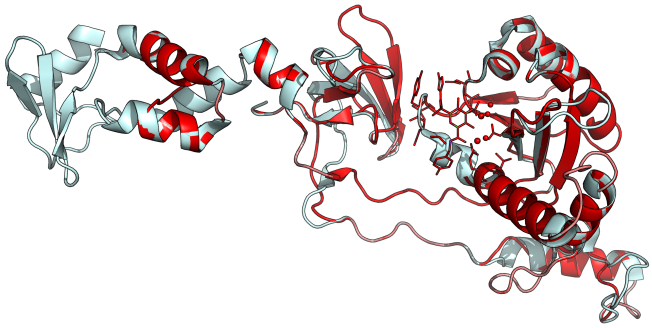
A: Sequence alignment of HIV-1 integrase (1148-1435) with PFV.



B. Ramachandran plot of HIV-1 Integrase Structure obtained from Procheck program



C: Comparison of homology model (red) with template (light blue)



Glide Docking: Protein-ligand complexes were obtained from docking using the Glide module of Schrodinger.[5] Glide (Version 5.8), a grid-based exhaustive search algorithm was used for all docking experiments[5]. Glide uses a series of hierarchical filters to find possible ligand pose in the active site, and the program has the option to treat the ligand fully flexible or rigid during the docking run. Glide uses an in-build docking scoring function resulting in a Glidescore (SP and XP). In the current setting, the SP docking with flexible ligand sampling mode was used. The van der Waals radii scaling factor was set to 0.8 with partial charge cutoff 0.15

Protein Ligand Interaction Fingerprint

In order to get the protein-ligand interactions of all compounds, the best docking pose for each compound/complex were analyzed using the Protein-Ligand Interaction Fingerprint (PLIF)

tool as implemented in the MOE software.[13] PLIF possesses a composition of seven visible fingerprint bits (side chain hydrogen bond donor (D), side chain hydrogen bond acceptor (A), backbone hydrogen bond donor (d), backbone hydrogen bond acceptor (a), solvent hydrogen bond (O), ionic attraction (I) and surface contact (C)). The hydrogen bond fingerprints are calculated using a method based on protein contact statistics, whereby a pair of atoms is scored by distance and orientation. Ionic interactions are scored by calculating the inverse square of the distance between atoms with opposite formal charge (e.g. a carboxylate oxygen atom and a protonated amine). Surface contact interactions are determined by calculating the solvent exposed surface area of the residue, first in absence of the ligand, and then in presence of the ligand.

Pharmacophore modeling

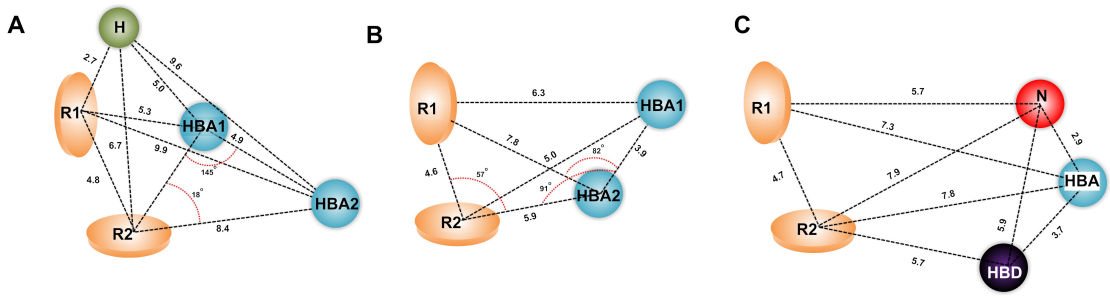
Pharmacophore modeling was carried out using the Phase (v3.1) module of the Schrödinger suit. [14] 3D conformers used for the pharmacophore building were generated using the ''Conf-Gen'' tool in the macro-model search algorithm with thorough sampling. The OPLS_2005 force field was used with a distance-dependent dielectric solvent model. The quality of the pharmacophore alignment was assessed using the RMSD (distance tolerance set to 1.2 Å) and the quality of the hypothesis was assessed using survival score, sites, vector, and selectivity scores. In the current setting, at least 4 common pharmacophore points must match with the selected active compounds and other settings were chosen as defaults.

Pharmacophore model results

For RNH modeling, among the 44 compounds, 25 compounds were set as a "must fit" option to derive the pharmacophore variance analysis, which results in 23 variances and two successful hypotheses (**Figure A/B**) survived in the post processing score. The *hypothesis 1* contains AARRH (A: Hydrogen bond acceptor, R: Ring, H: halogen) (survival score: 13.0, matches: 26, activity: 5.09) and *hypothesis 2* contains AARR (survival score: 43.5, matches: 40, activity: 4.8) features (A: Hydrogen bond acceptor (HBA), R: Ring). In both hypotheses, the inter-distance between the two rings is separated by 4.6 Å, and the distance between Ring 2 to hydrogen bond acceptor 1 range from 5 to 5.4 Å. However, the major difference is that the distance between two HBAs is quite large for *hypothesis 1* (4.9 Å) compared to *hypothesis 2* (3.9 Å). Although the distance is quite large for *hypothesis 1*, still the HBA position is favorable for metal chelation as shown in the X-ray crystal structure of RNH (PDB ID: 3QIO). In comparison to a previous pharmacophore model,[6] both *hypothesis 1* and *hypothesis 2* are in good agreement with models derived from a diverse set of RNH inhibitors. According to the previous model, the two HBAs are separated with a distance of 4.5 Å that favorably coordinates (in narrow angle) with the magnesium ions, which are separated with a distance of

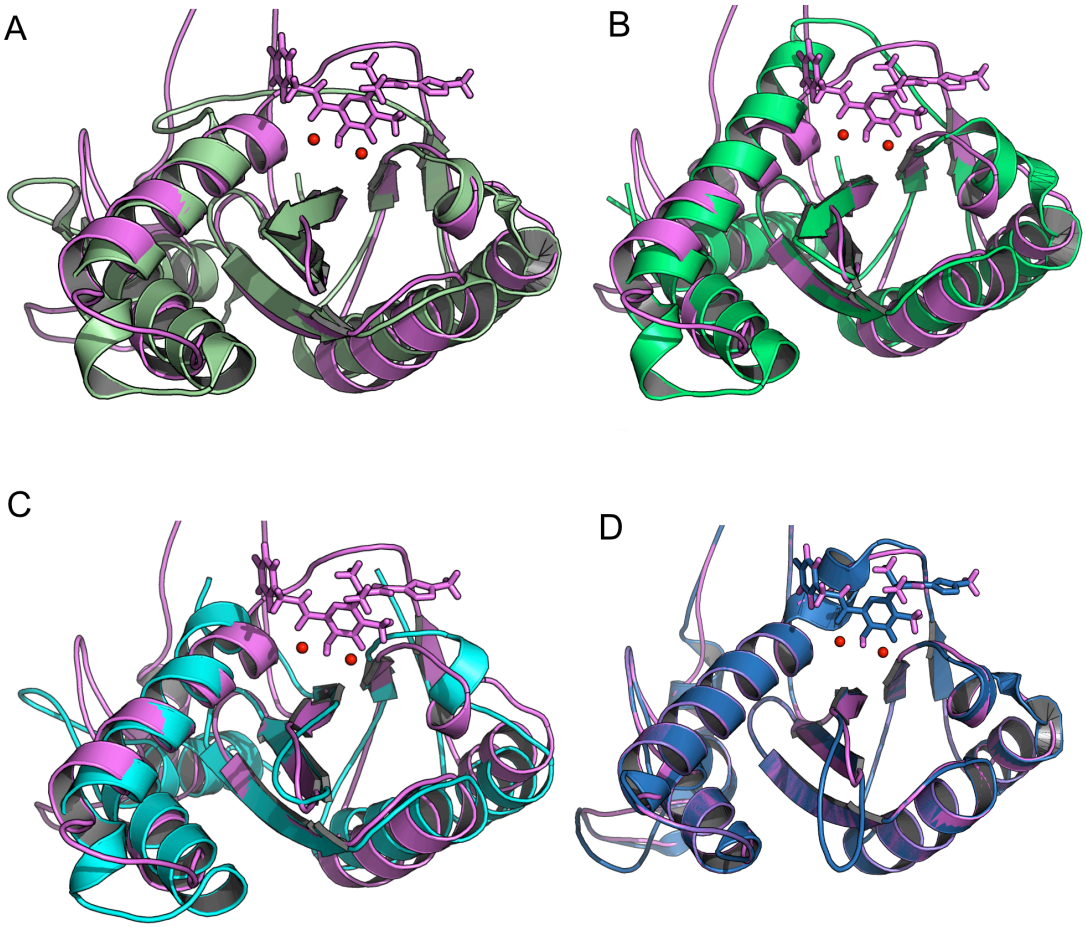
~4 Å in the active site. Moreover, the distance between the two hydrophobic rings is also found to be optimal, and ranges from 4 to 5 Å. Compared to *hypothesis* 2, *hypothesis* 1 is very restricted to this series of compounds, as only 26 compounds (mainly series 7) were aligned for model building. This observation is also in good agreement with experiment where the diketo-esters are more active against RNH compared to IN and a possible reason might be that the compounds in series 7 are aligned such that the two carbonyl oxygen atoms are down positioned, which can be noted with the angle between R2-HBA2-HBA1 and HB2-R2-HBA1 which are found to be 145° and 17°, respectively. This angle type is favoured when the terminal acidic group is esterified, because this was not observed for diketo-acid derivatives where the features are very plan. In addition, a halogen attached at the benzyl group (preferably meta position) might increase the RNH inhibition.

For IN modeling, with 46 active compounds, 3 hypotheses survived from 23 pharmacophore variances. All hypotheses share very similar features except inter feature distances. The *hypotheses 3-5* contain ADNRR (survival score: 9.4-9.7, matches: 22, activity: 6.5, A: HBA, R: ring, N: negative charge, D: HB donor). Similar to RNH models, *hypothesis 3* (**Figure C**) possesses similar ring-ring distance. In contrast to *hypotheses 1-2, hypothesis 3* was primilarly derived from diketo-acid derivatives. The distance pattern observed between N-HBA (2.9 Å) and HBA-D (3.7 Å) is quite different from *hypothesis 1,* as here D-N-HBA are arranged in plane which is highly preferable to chelation formation with diketo-acid derivatives.



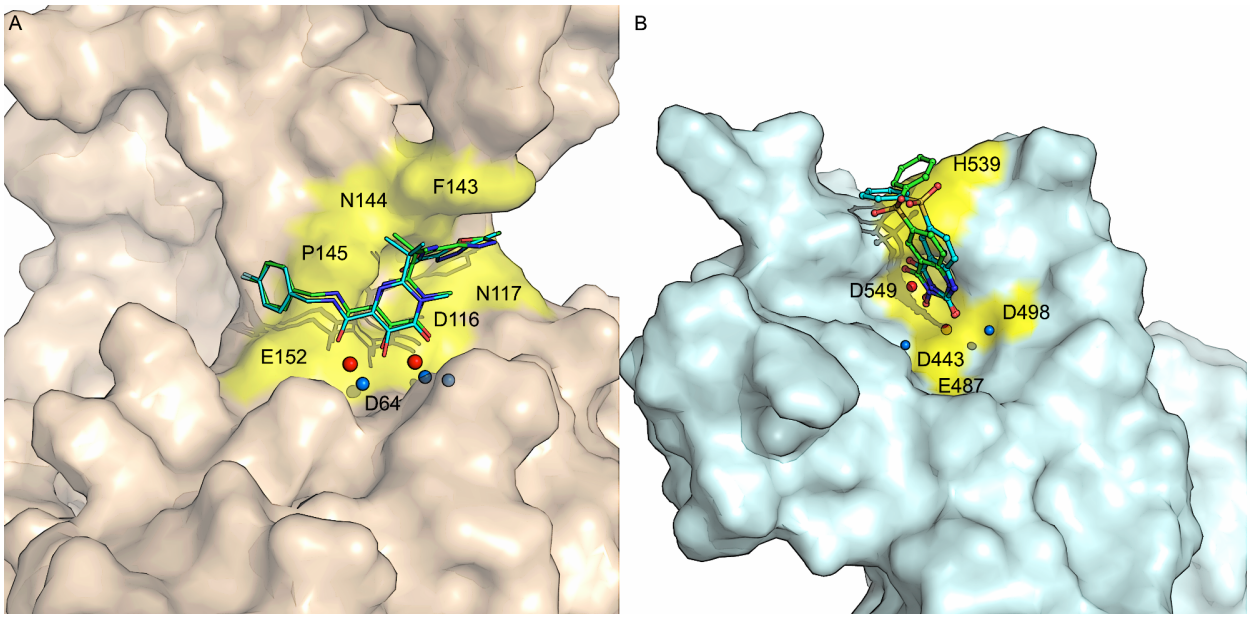
Diagrammatic representation of pharmacophore models for Rnase H and Integrase, A; Hypothesi 1, B: Hypothesis 2, C: Hypothesis 3.

S. Figure 1



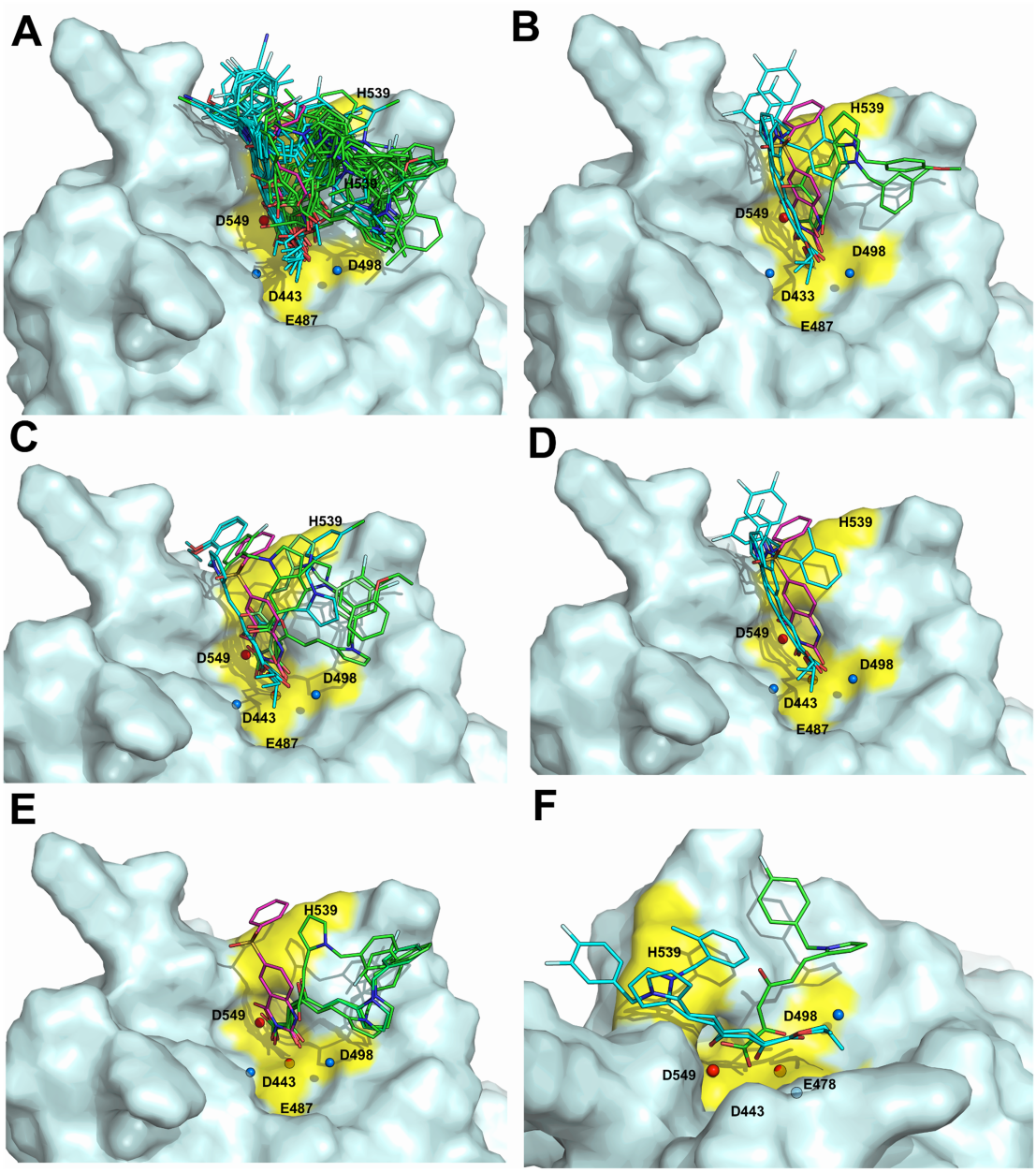
A comparison of the catalytic domains of the homology model (pink) and the crystal structures of HIV-1 integrase (3L3U (A), 3NF8 (B), 4DMN (C) and 1BIZ (D)).

S. Figure 2



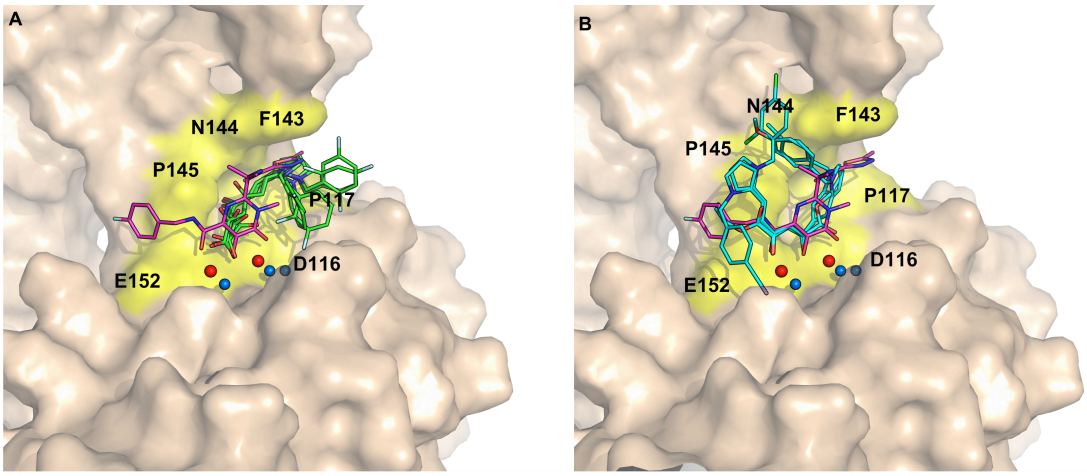
A: Comparison of the binding mode of crystal bound Raltegravir (green stick) with the result from the docking experiment (cyan stick). B: Comparison of binding mode of N-hydroxy quinazolinedinone (NHQD) (green stick) with docking experiment (cyan stick). Both active sites are highlighted with yellow surface and magnesium ions (red sphere) and waters (blue sphere) are also shown.

S. Figure 3.



Comparison of binding poses of compounds in series 7 and 8 at the RNase H binding site. A: Binding mode of highly active compounds against RNase H, C: Binding mode of poor inhibitory compounds against RNase H, D: Binding mode of highly active series 7 compounds, E: Binding mode of highly active series 8 compounds against RNase H, F: Comparing the chelation mode of series 7 and 8 compounds in the active site.

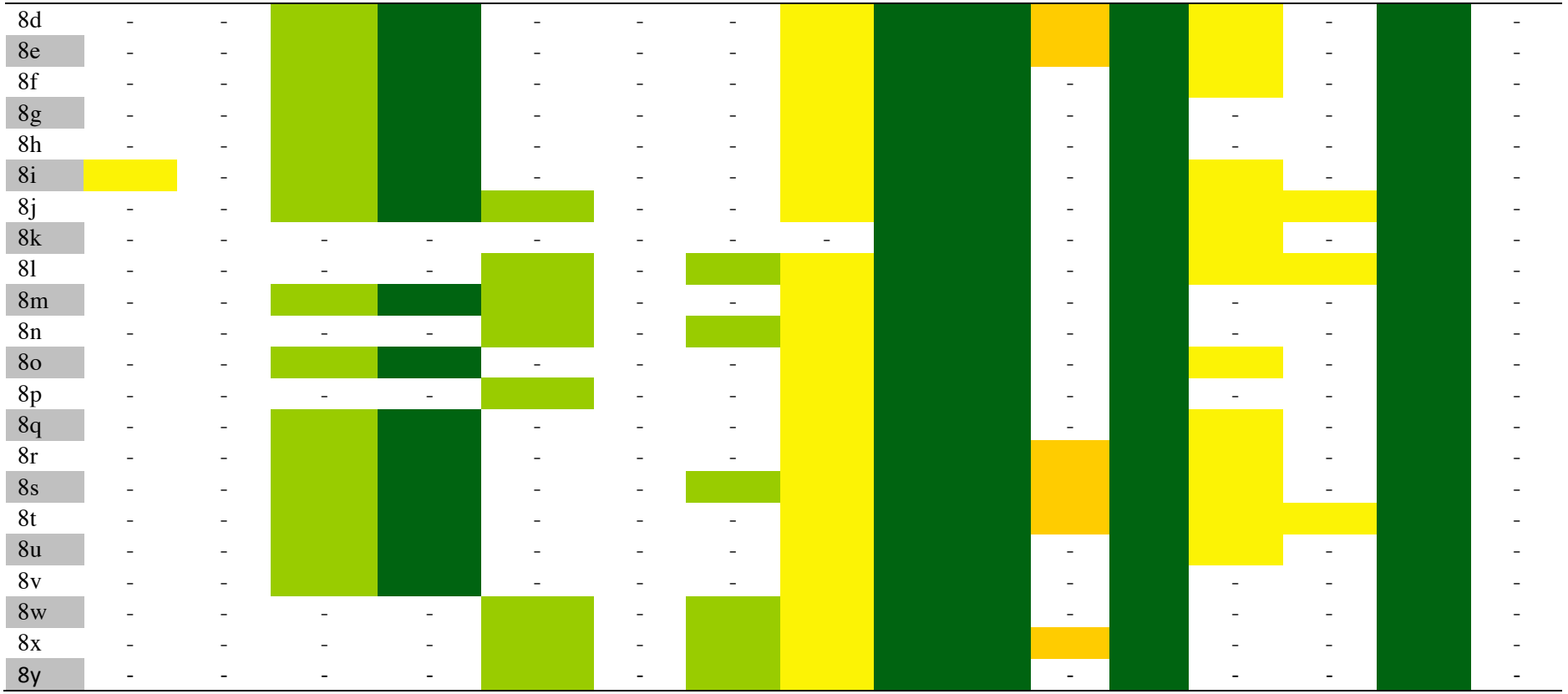
S. Figure 4



A. Binding mode of highly active compounds against Integrase, B: Binding mode of low inhibitory compounds against Integrase.

Table 1 A: Protein-Ligand Interaction Fingerprint for RNase H

ID	D443 Don	N474 Surf	E478 Don	D498 Don	H539 Acc/Don	D549 Don	R557 Acc	R557 Ionic	Mg1	Mg2	H_15	H_17	H_24	H_118	H_562	H_563
%	49.02	37.25	76.47	84.31	60.78	25.49	60.78	49.02	100	100	58.8	100	41.1	49.02	100	50.9
7a								-					-			
7b			-			-		-					-	-		
7c						-		-					-			
7d		-						-					-			
7e	-	-						-					-	-		
7f						-		-					-			
7g						-		-					-			
7h			-			-	-	-			-		-	-		
7i					-			-								
7j	-							-								
7k		-						-					-			
7l						-		-					-			
7m						-		-					-			
7n		-				-		-					-			
7o		-				-		-					-			
7p								-								
7q			-					-					-			
7r						-		-					-			
7s								-					-			
7t						-		-					-			
7u						-		-					-			
7v								-								



Note: Dark green: 80-10-0% abundance level, Light green: 60-80% abundance level, Red: <50% abundance level, White: No interaction.
Don: Hydrogen bond donor, ACC: Hydrogen bond acceptor, Surf: Surface contact, Ion: Ionic interaction.

Table 2 B: Protein-Ligand Interaction Fingerprint for Integrase

Comp No	D64 Don	D116 Don	G118 Surf	T143 Acc/Surf	P145 Surf	E152 Don	Mg1	Mg2	H_400	H_401	H_534
%	64.7	72.5	21.569	62.7	84.3	76.4	70.5	50.9	98.0	64.7	100
7a			-							-	
7b			-	-			-				
7c			-							-	
7d			-							-	
7e			-	-						-	
7f			-				-				
7g			-	-							
7h			-				-			-	
7i			-	-							
7j	-		-	-					-		
7k			-								
7l			-	-			-				
7m			-	-			-				
7n			-							-	
7o			-							-	
7p			-				-			-	
7q				-						-	
7r			-	-							
7s			-	-							
7t			-	-			-				
7u			-							-	
7v			-	-							
7w			-	-							
7x			-	-			-				
7y				-							
8a	-	-	-		-	-					
8b			-				-				
8c	-	-				-				-	

Note: Dark green: 80-10-0% abundance level, light green: 60-80% abundance level, red: <50% abundance level, White: No interaction.
Don: Hydrogen bond donor, ACC: Hydrogen bond acceptor, Surf: Surface contact, Ion: Ionic interaction.

Table 1: FLAP results for the studied compounds

LV	Exp VarX	Acc.Exp VarX	Exp SSX	Acc. Exp SSX	SDEC	SDEP	R^2	Q^2	Exp SSY	Acc Exp SSY
1	8.094	8.094	10.295	10.295	0.546	0.756	0.603	0.239	60.347	60.347
2	8.477	16.571	10.271	20.565	0.384	0.669	0.804	0.405	20.006	80.353
3	3.271	19.842	5.032	25.598	0.299	0.660	0.881	0.421	7.731	88.083
4	4.060	23.903	5.589	31.787	0.228	0.715	0.931	0.321	4.985	93.068
5	7.904	31.806	8.778	39.965	0.188	0.698	0.953	0.352	2.239	95.307

References:

1. E. B. Lansdon, Q. Liu, S. A. Leavitt, M. Balakrishnan, J. K. Perry, C. Lancaster-Moyer, N. Kutty, X. Liu, N. H. Squires, W. J. Watkins and T. A. Kirschberg, *Antimicrobial agents and chemotherapy*, 2011, 55, 2905-2915.
2. K. Arnold, L. Bordoli, J. Kopp and T. Schwede, *Bioinformatics*, 2006, 22, 195-201.
3. G. Madhavi Sastry, M. Adzhigirey, T. Day, R. Annabhimoju and W. Sherman, *Journal of computer-aided molecular design*, 2013, 27, 221-234.
4. H. Li, A. D. Robertson and J. H. Jensen, *Proteins*, 2005, 61, 704-721.
5. R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis and P. S. Shenkin, *Journal of medicinal chemistry*, 2004, 47, 1739-1749.
6. P. Vasanthanathan and J. Kongsted, *PloS one*, 2013, 8, 1-15.
7. S. Hare, S. S. Gupta, E. Valkov, A. Engelman and P. Cherepanov, *Nature*, 2010, 464, 232-236.
8. L. Krishnan, X. Li, H. L. Naraharisetty, S. Hare, P. Cherepanov and A. Engelman, *Proceedings of the National Academy of Sciences of the United States of America*, 2010, 107, 15910-15915.
9. S. Balasubramanian, M. Rajagopalan and A. Ramaswamy, *Journal of Biomolecular Structure and Dynamics*, 2012, 29, 1163-1174.
10. B. C. Johnson, M. Metifiot, A. Ferris, Y. Pommier and S. H. Hughes, *J Mol Biol*, 2013, 425, 2133-2146.
11. T. U. Consortium, *Nucleic Acids Research*, 2013, 41, D43-D47.
12. L. Krishnan and A. Engelman, *J Biol Chem*, 2012, 287, 40858-40866.
13. MOE, 2011.
14. S. L. Dixon, A. M. Smondyrev, E. H. Knoll, S. N. Rao, D. E. Shaw and R. A. Friesner, *Journal of computer-aided molecular design*, 2006, 20, 647-671.