In Silico Design of New Inhibitors Against Hemagglutinin of Influenza

Aurélie PERRIER,[*,†,‡,¶] Matthias ELUARD,[†,§,¶] Michel PETITJEAN,[‖,¶] and Anne VANET[*,§,¶]

†*Chimie ParisTech, PSL Research University, CNRS, Institut de Recherche de Chimie Paris (IRCP), F-75005 Paris, France*

‡*Université Paris Diderot, Sorbonne Paris Cité, 5 rue Thomas Mann, F-75205 Paris Cedex 13, France*

¶*Epôle de Génoinformatique, Institut Jacques Monod, UMR7592, CNRS, F-75013 Paris, France*

§*Pathologies de la réplication de l'ADN, Institut Jacques Monod, UMR7592, CNRS, F-75013 Paris, France*

‖*MTi, UMR-S 973, INSERM, University Denis Diderot, Paris 7, France*

E-mail: aurelie.perrier-pineau@univ-paris-diderot.fr; anne.vanet@univ-paris-diderot.fr

H1 structures

Table S1: List of the 32 PBD structures of H1 used for the clustering.

PDB Code	Strain
1RD8	A/Brevig Mission/1/1918 H1N1
1RUZ	A/Brevig Mission/1/1918 H1N1
1RU7. 1RVX, 1RVZ	A/Puerto Rico/8/1934 H1N1
2WRG	A/Brevig Mission/1/1918 H1N1
3LZG	swl A/California/04/2009 H1N1
3M6S	A/Darwin/2001/2009 H1N1
3UBE, 3UBJ, 3UBN, 3UBQ	swl A/California/04/2009 H1N1
4EDB	swl A/California/04/2009 H1N1
4EEF	A/Brevig Mission/1/1918 H1N1
4GXU	A/Brevig Mission/1/1918 H1N1
4JTV, 4JTX, 4JU0	swl A/California/04/2009 H1N1
4JUG, 4JUH, 4JUJ	A/Brevig Mission/1/1918 H1N1
4LXV	A/Washington/5/2011 H1N1
4M4Y	swl A/California/04/2009 H1N1
5UGY	A/Solomon Islands/3/2006 H1N1
5UJZ, 5UK0. 5UK1. 5UK2	A/Solomon Islands/3/2006 H1N1
5VMC, 5VMF, 5VMG, 5VMJ	A/Brevig Mission/1/1918 H1N1

The multiple alignment (Fig. S1) shows a modification of 4 amino acids (AAs) in the area of interest. The AAs K395. K398, M400 and N405 of 1RU7 can be mutated to H395 or N395. R398, I400 and K405. respectively. The characteristics of the AAs in positions 395 and 398 in synthetic lethality are thus preserved. Position 405 is invariant, a position is considered invariant if the mutation rate is less than 0.3 %. The K405N mutation at this position is under-represented in the multiple alignment (3 out of 32 sequences) and should correspond to a mutation rate of less than 0.3 %. Position 400 is neither a synthetic lethal nor an invariant. The mutation at this position is therefore not lethal.

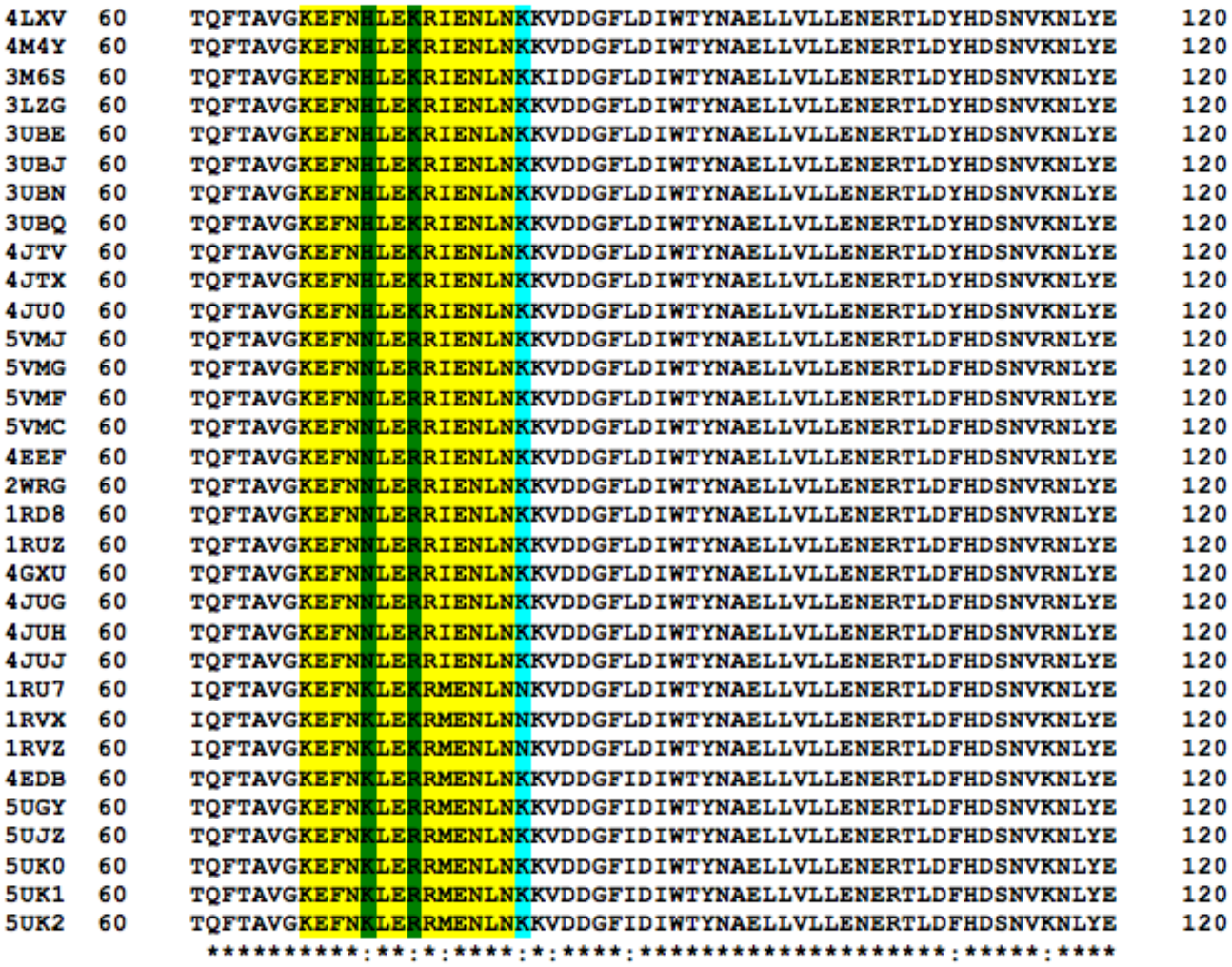

Figure S1: Multiple alignment (with ClustalΩ) of a portion of the HA2 subunit of the 32 PDB structure primary sequences. The area of interest is in color (position 67 to 81 corresponding to residues 391 to 405). Three mutations are observed: (1) on residues in synthetic lethality 395. 398 (in green) (the conservation of the physicochemical properties should not lead to lethality), (2) on residue 405 (in blue), (3) on the residue 400.

Analysis of the MD simulation of 1RU7 without ligand

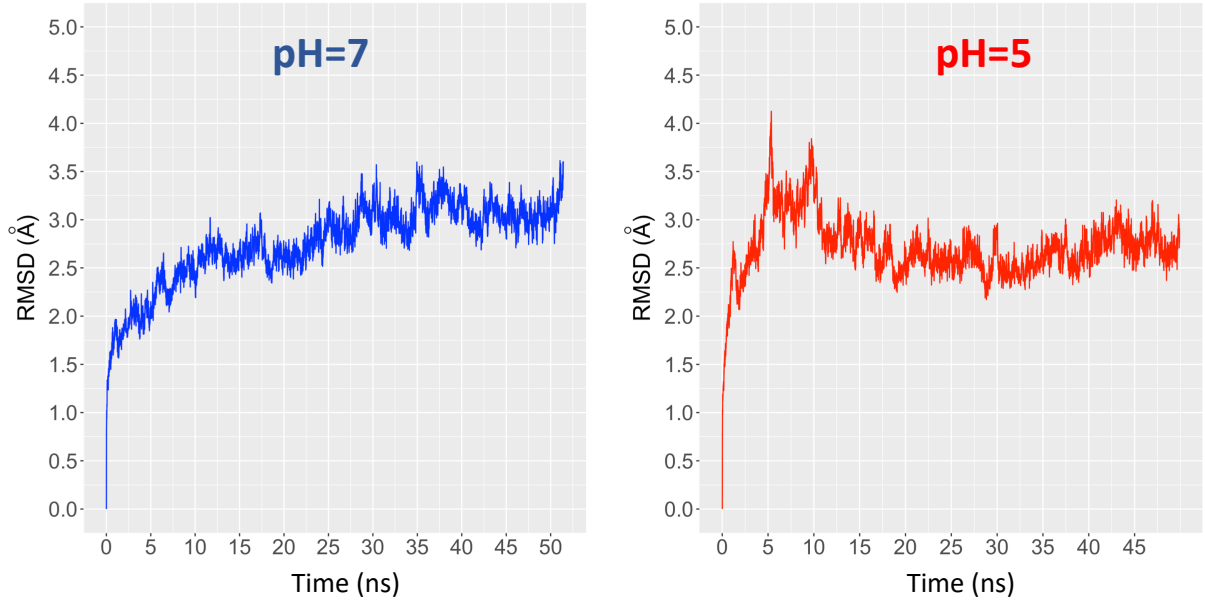

Figure S2: RMSD (in Å) of the MD simulations at pH=7 (left) and pH=5 (right) for the 1RU7 protein.

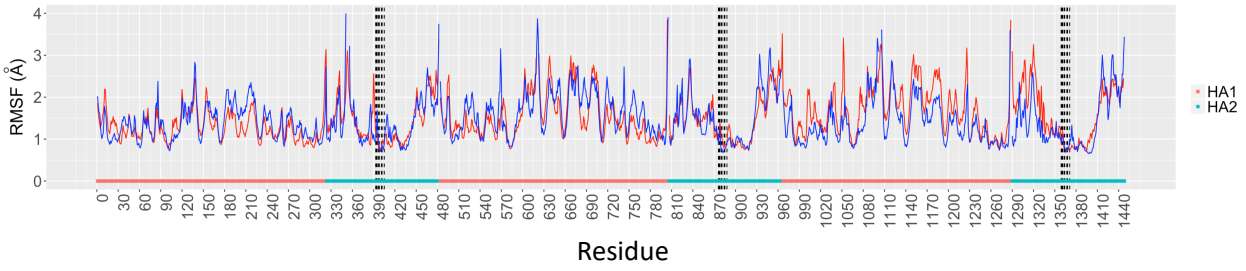

Figure S3: RMSF (in Å) calculated for the Cα of the AAs of the protein 1RU7 at pH=7 (blue) and pH=5 (red). The three monomers of Figure 1 are taken into account. The HA1 and HA2 subsunits are respectively identified with orange and turquoise lines under the graph. The position of the residues of interest are represented with dashed lines.

Interaction of the EQRRS linear peptide with 1RU7

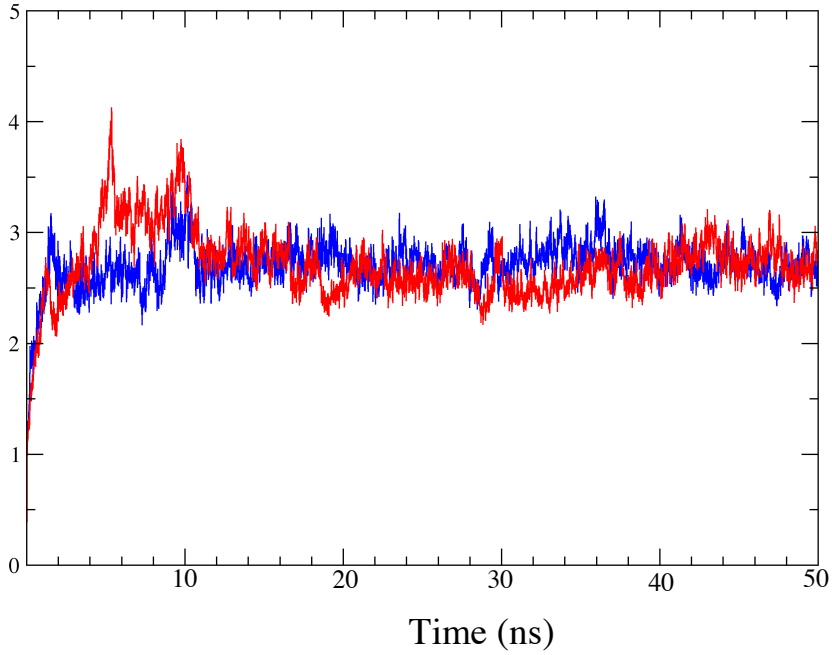

Figure S4: RMSD (in Å) of the MD simulations at pH=7 (blue) and pH=5 (red) for the EQRRS - 1RU7 system.

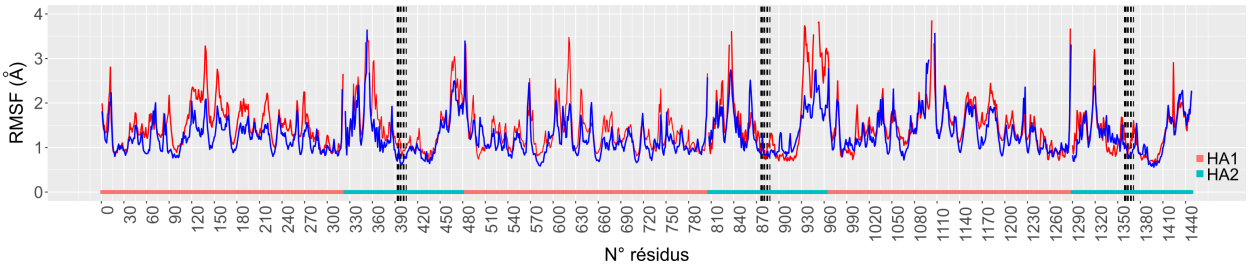

Figure S5: RMSF (in Å) at pH=7 (blue) and pH=5 (red) for the EQRRS (one molecule) - 1RU7 system. See Fig. S3 for more details.

Table S2: Free energy of binding calculated with the MM/PBSA and MM/GBSA apporaches for the EQRRS-1RU7 and DQRRD-1RU7 complexes at pH=7.

	Energy components	**EQRRS**	**DQRRD**
PBSA			
	ΔE_{ele}	-462.9	-315.5
	ΔE_{vdw}	-43.9	-25.5
	$\Delta G_{ele,sol}$	466.6	300.0
	$\Delta G_{nonpolar,sol}$	-6.1	-3.9
	ΔE_{MM}	-506.8	-341.1
	ΔG_{solv}	460.5	296.0
	ΔG_{tot}	-46.4 (σ: 9.3)	-45.0 (σ: 11.0)
GBSA			
	ΔE_{ele}	-462.9	-315.5
	ΔE_{vdw}	-43.9	-25.5
	$\Delta G_{ele,sol}$	487.8	309.2
	$\Delta G_{nonpolar,sol}$	-8.8	-5.7
	ΔE_{MM}	-506.8	341.1
	ΔG_{solv}	478.9	303.4
	ΔG_{tot}	-27.9 (σ: 7.4)	-37.6 (σ: 10.9)

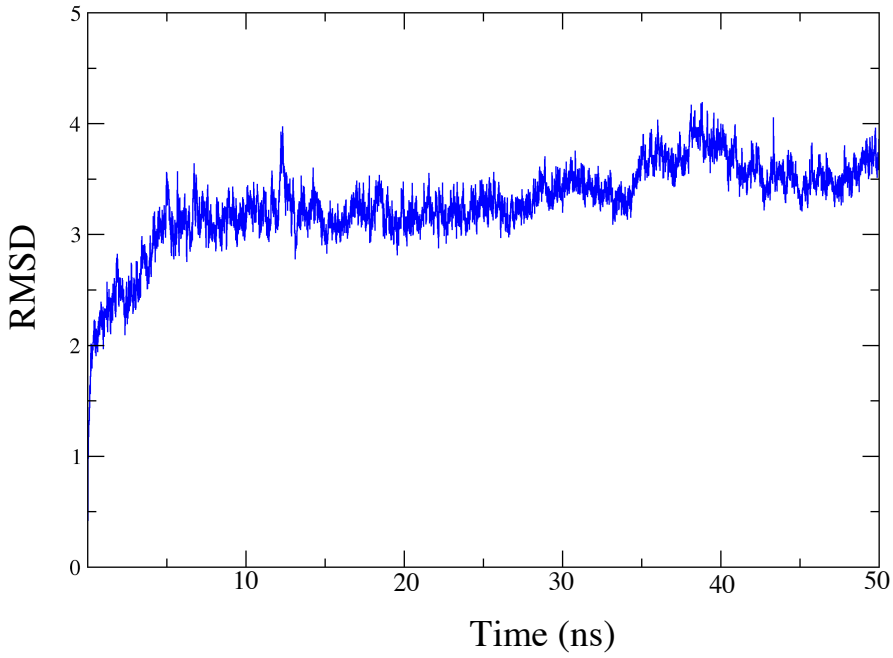

Figure S6: RMSD (in Å) of the MD simulations at pH=7 for the system with **three EQRRS ligands** interacting with the 1RU7 protein.

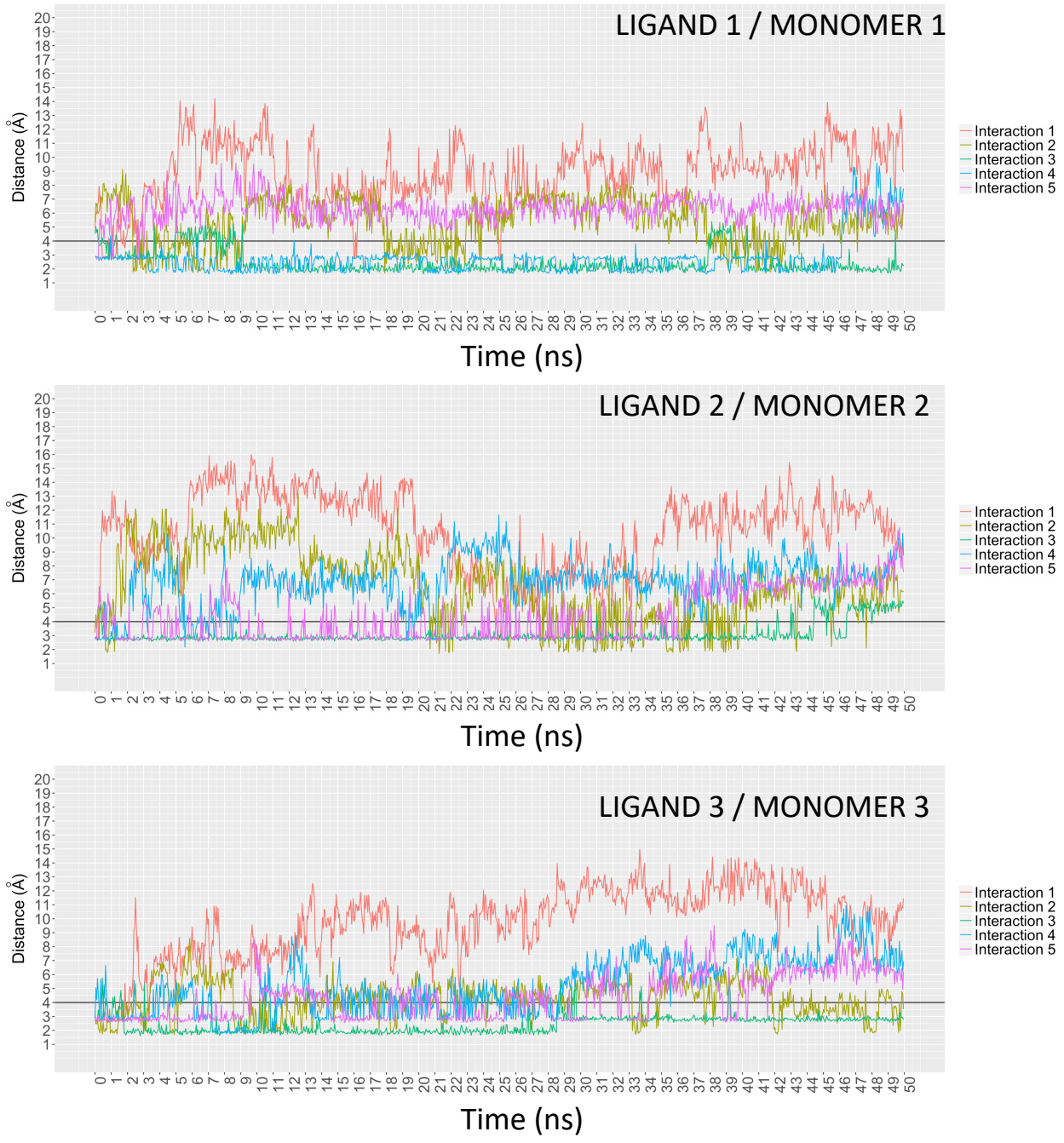

Figure S7: **Three EQRRS ligands interacting with the 1RU7 protein**: distances (in Å) caracterizing the 5 target conctacts between one AA of the EQRRS ligand and one of the 5 residues of 1RU7 (pH=7). Interaction 1: E-K395, Interaction 2: Q-N394, Interaction 3: R-E392, Interaction 4: R-E401, Interaction 5: S-K398. These distances are given for the three ligand /monomer couples.

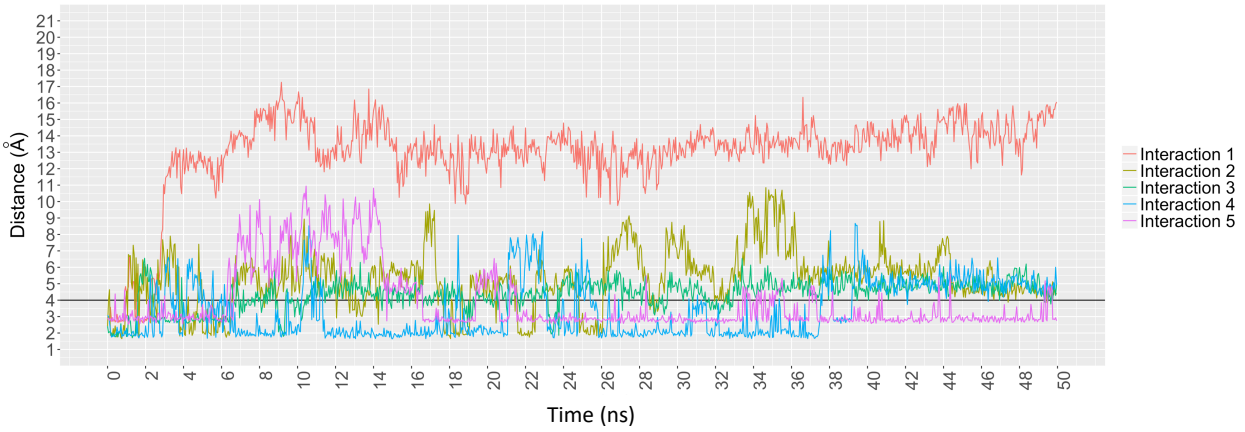

Figure S8: EQRRS ligand interacting with the 1RU7 protein at **pH=5**: distances (in Å) caracterizing the 5 target conctacts between one AA of the EQRRS ligand and one of the 5 residues of 1RU7 (pH=7). Interaction 1: E-K395, Interaction 2: Q-N394, Interaction 3: R-E392, Interaction 4: R-E401, Interaction 5: S-K398.

Effect of the ligand on the pKa values

We evaluated the effect of the ligand on the pKa values of the Glu residues E392, E397 and E401 inside the area of interest. We extracted 40 conformations generated during the 20 last nanoseconds of the MD simulations at pH=7 for protein 1RU7 with and without the ligand EQRRS. The pKa values were computed at pH=5 with PROPKA 3.0.[4] The 40 files were prepared according to PROPKA requirements: water molecules and ions were suppressed, the residues were renamed according to the protonation state at pH=7 and disulfure bridges were suppressed (residues CYX were renamed CYS). Then the 40 conformations were analyzed with PROPKA 3.0 through the on-line server PDB2PQR[1] in order to estimate the pKa of the residues at pH=5. A pKa value smaller than 5 indicates that the Glu residue is protonated. Hemagglutinin is an homotrimer, thus we analyzed 9 residues (3 per monomer): E392, E397. E401. E875. E880. E884, E1358, E1363 and E1367. Each residue is analyzed either in presence of the ligand or not, so that we analyzed a total of 18 populations of 40 pKa values. For clarity, to each of these populations is associated an identifier (ID), which is an integer value in the range 1–18.

The means and the standard deviations of these 40 populations are reported in Table S3.

Several statistical tests were realized with the R software.[5] We assumed that, for each of the 18 populations of 40 observed pKa values, these 40 values constitute a sample of 40 independant observations. First, a Shapiro-Wilk test of normality was performed on the means.[6] The normality hypothesis was rejected with a p-value of 4.3e-8. Thus parametric tests were needed. Then, assuming the independance of the 18 samples, a Kruskal-Wallis rank sum test was performed.[3] The hypothesis that the 18 samples were issued from a common population was rejected with a p-value of 2.2e-16. We conclude that the pKa values are significantly depending on the ID.

The next step was to do pairwise comparisons of the 18 populations in order to determine, for each pair of samples assumed to be independant, whether or not they come from populations having the same distribution. We used the parametric Mann-Whitney U test[7] corrected by

Table S3: Means and standard deviations of the 18 populations of the 40 pKa values. W-pval: p-value associated to the one-sample Wilcoxon test.

ID	Residue	Ligand	mean	standard deviation	W-pval
1	E392	yes	3.55900	0.3791948	1.455e-11
2	E392	no	4.76000	0.3520198	0.0006557
3	E397	yes	3.33725	0.4427362	1.164e-10
4	E397	no	2.91850	0.3637874	2.328e-10
5	E401	yes	3.68925	0.3404325	5.821e-11
6	E401	no	4.03050	0.3813804	4.657e-10
7	E875	yes	4.92875	0.5027167	0.5562
8	E875	no	4.73000	0.5448971	0.002751
9	E880	yes	2.72725	0.2365347	2.328e-10
10	E880	no	3.58350	0.2424612	2.91e-11
11	E884	yes	4.27525	0.1913111	2.328e-10
12	E884	no	4.14100	0.3477459	2.328e-09
13	E1358	yes	4.82750	0.1833695	0.001623
14	E1358	no	5.06300	0.1525543	0.9735
15	E1363	yes	3.49625	0.5007158	2.91e-11
16	E1367	no	3.66625	0.2539199	2.328e-10
17	E1367	yes	4.05600	0.1947095	1.863e-09
18	E1367	no	4.18125	0.1979535	9.313e-10

the Bonferroni adjustment.[2]

The results are in Table S4. The mean pKa values for a given residue are significantly depending on which monomer is considered. For E401, the pKa is lowered for the monomer interacting with the ligand, but not for the two other monomers. We conclude that the ligand is responsible of this effect. For E397 and E392, there is also a decrease of the pKa, but this decrease occurs both on the monomer interacting with the ligand and on one of the two other monomers: there is a possible effect of the ligand, but this effect remains hypothetical.

The one-sample Wilcoxon test[8] was used to compare each of the 18 mean pKa values to the theoretical pKa value at pH=5. The 18 p-values are reported in Table S3. These tests permit to know which mean pKa values are significantly less than 5. and thus to know which residues are protonated at acidic pH. When the p-value is less than 0.01. we conclude with an error risk of 1% that the pKa of the residue is smaller than 5. thus indicating that the

residue is charged at pH=5.

We got p-values less than 0.01 in all cases except in the following ones: for E392 with ligand (ID=7) and for its corresponding residue on an other monomer without ligand (ID=14, E1358). Thus we may deduce that the ligand induces a deprotonation at pH=5 for E1358, so that the carboxy group of Glu is charged in presence in the ligand. For E392, the rejection of the hypothesis pKa>5 is much stronger in presence of the ligand (ID=1) than in the absence of the ligand (ID=2): we may consider that, in some extent, this is not a disagreement with our deduction that the carboxy group of Glu is charged in presence in the ligand. It is recalled that, in test theory, the fact that the hypothesis passes a test that does not prove this hypothesis (while there is few risk to reject the hypothesis when the test is not passed). Thus, the large p-value got for E875 with ligand (ID=7) does not prove any protonation in presence of the ligand. Furthermore, the p-values must be interpreted with some caution because they are computed on the independancy assumptions that we mentioned earlier. To summarize, we may reasonably think that that the ligand induces a deprotonation at pH=5 of Glu392 and its corresponding residues on the two other monomers, so that the carboxy group of this Glu residue is charged.

Table S4: The symmetric matrix of the p-values of the pairwise Mann-Whitney U tests corrected by the Bonferroni adjustment. Only the lower triangular part of the matrix is displayed, without the diagonal.
ID: identifier of the residue (see Table S3).

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
2	1.3e-11	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3	0.40298	4.4e-11	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4	3.9e-08	2.5e-12	6.0e-05	-	-	-	-	-	-	-	-	-	-	-	-	-	-
5	1.00000	4.6e-11	0.00097	2.8e-09	-	-	-	-	-	-	-	-	-	-	-	-	-
6	3.1e-05	2.2e-07	3.8e-08	2.3e-10	0.00063	-	-	-	-	-	-	-	-	-	-	-	-
7	7.6e-11	1.00000	4.4e-11	4.3e-12	3.5e-10	1.6e-07	-	-	-	-	-	-	-	-	-	-	-
8	4.5e-10	1.00000	1.1e-10	6.7e-12	2.6e-09	4.0e-05	1.00000	-	-	-	-	-	-	-	-	-	-
9	4.8e-11	2.2e-12	2.3e-09	0.47115	1.1e-11	2.2e-12	2.4e-12	2.2e-12	-	-	-	-	-	-	-	-	-
10	1.00000	3.3e-12	0.01545	7.8e-09	1.00000	4.8e-07	4.3e-11	8.7e-11	1.5e-11	-	-	-	-	-	-	-	-
11	2.7e-09	1.9e-07	4.8e-10	8.6e-12	7.8e-09	0.00049	1.3e-06	0.00881	2.2e-12	1.2e-11	-	-	-	-	-	-	-
12	5.6e-07	3.2e-07	2.2e-09	5.5e-11	1.3e-06	0.53340	1.3e-06	0.00048	2.2e-12	1.4e-09	0.02826	-	-	-	-	-	-
13	2.2e-12	1.00000	2.3e-11	2.2e-12	8.9e-12	6.0e-08	1.00000	1.00000	2.2e-12	2.2e-12	6.4e-12	5.2e-08	-	-	-	-	-
14	2.2e-12	0.00761	5.7e-12	2.2e-12	2.2e-12	1.6e-11	1.00000	1.00000	2.2e-12	2.2e-12	2.2e-12	1.5e-10	2.0e-05	-	-	-	-
15	1.00000	1.5e-10	1.00000	5.4e-06	1.00000	9.2e-06	2.2e-10	1.1e-09	1.6e-09	1.00000	3.8e-09	1.4e-07	1.5e-10	1.8e-11	-	-	-
16	1.00000	3.7e-12	0.00095	3.5e-09	1.00000	0.00012	1.6e-10	5.3e-10	7.7e-12	1.00000	7.8e-11	1.6e-07	2.2e-12	2.2e-12	1.00000	-	-
17	1.3e-06	1.8e-10	4.6e-09	5.5e-11	5.2e-06	1.00000	2.6e-08	2.8e-05	2.2e-12	2.2e-09	0.00182	1.00000	2.2e-12	2.2e-12	2.4e-07	1.1e-07	-
18	3.6e-08	8.3e-09	6.7e-10	1.7e-11	3.7e-08	0.00381	3.4e-07	0.00068	2.2e-12	4.4e-11	1.00000	1.00000	5.3e-12	2.2e-12	1.4e-08	2.1e-09	1.00000

Interaction of the DQRRD cyclic peptide with 1RU7

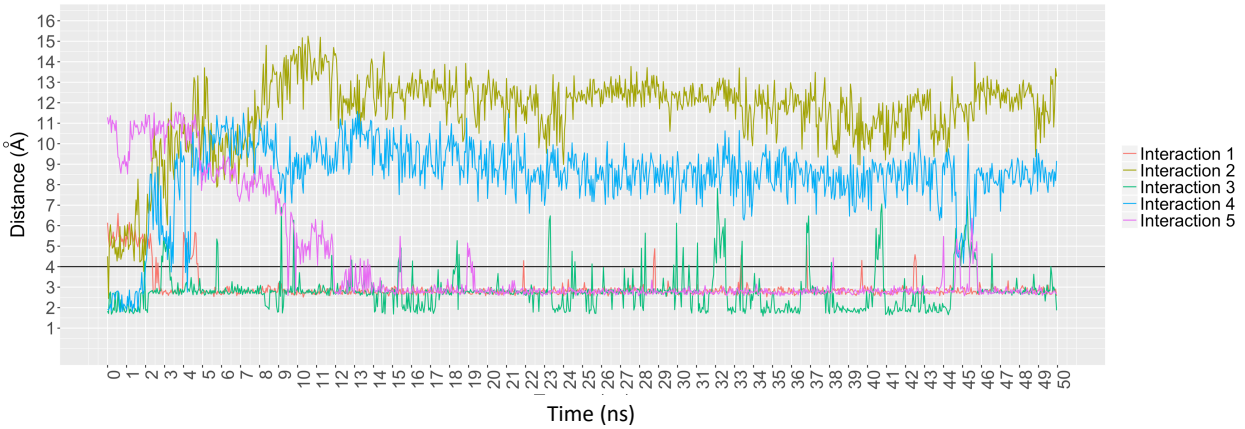

Figure S9: DQRRD ligand interacting with the 1RU7 protein at pH=7: distances (in Å) caracterizing the 5 target conctacts between one AA of the EQRRS ligand and one of the 5 residues of 1RU7 (pH=7). Interaction 1: E-K395, Interaction 2: Q-N394, Interaction 3: R-E392, Interaction 4: R-E401, Interaction 5: S-K398.

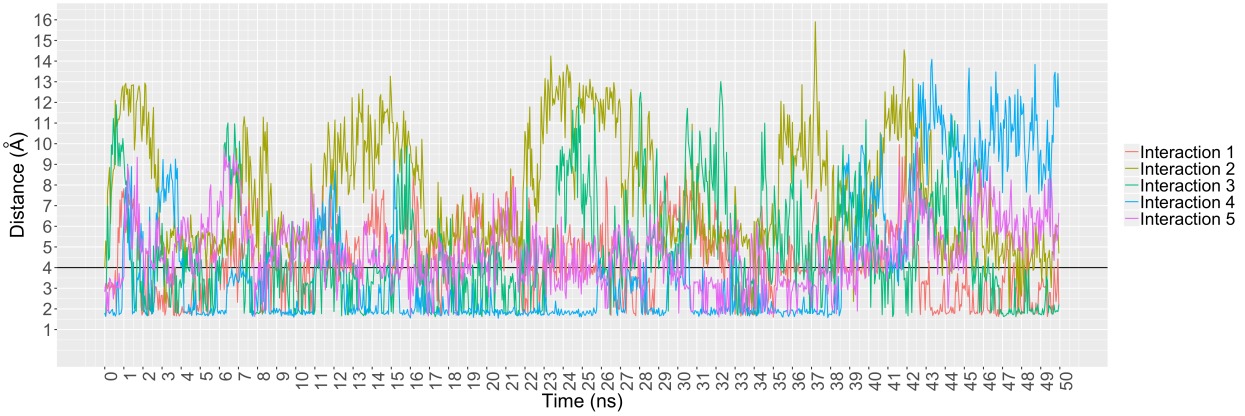

Figure S10: DQRRD ligand interacting with the 1RU7 protein at **pH=5**: distances (in Å) caracterizing the 5 target conctacts between one AA of the EQRRS ligand and one of the 5 residues of 1RU7 (pH=7). Interaction 1: E-K395, Interaction 2: Q-N394, Interaction 3: R-E392, Interaction 4: R-E401, Interaction 5: S-K398.

References

(1) Dolinsky, T.J.; Nielsen, J.E.; McCammon, J.A.; Baker, N.A.; PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* **2004**, *32*, W665–W667. doi 10.1093/nar/gkh381

(2) Dunn, O.J. Multiple comparisons among means. *J. Am. Stat. Assoc.* **1961**, *56*[293], 52–64. doi 10.1080/01621459.1961.10482090

(3) Kruskal, W.H.; Wallis, W.A. Use of ranks in one-criterion variance analysis. *J. Am. Stat. Assoc.* **1952**, *47*[260], 583–621. doi 10.1080/01621459.1952.10483441

(4) Olsson, M.H.M.; SøndergAArd, C.R.; Rostkowski, M.; Jensen, J.H. PROPKA3: Consistent treatment of internal and surface residues in Empirical pK_a Predictions. *J. Chem. Theor. Comput.* **2011**, *7*[2], 525–537. doi 10.1021/ct100578z

(5) R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. http://www.R-project.org (accessed June 27. 2018)

(6) Shapiro,S.S.; Wilk, M.B. An analysis of variance test for normality (complete samples). *Biometrika* **1965**, *52*[3–4], 591–611. doi 10.1093/biomet/52.3-4.591

(7) Mann H.B.; Whitney, D.R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.* **1947**, *18*[1], 50–60. doi 10.1214/aoms/1177730491

(8) Wilcoxon, F. Individual comparisons by ranking methods. *Biometrics Bull.* **1945**, *1*[6], 80–83. doi 10.2307/3001968