



**HAL**  
open science

**Structural and functional analysis of physicochemical interactions of the hemagglutinin of influenza A viruses : evolutionary constraints and perspectives for the development of broad-spectrum therapeutic antivirals and universal vaccines**

Valentin Ozeel

► **To cite this version:**

Valentin Ozeel. Structural and functional analysis of physicochemical interactions of the hemagglutinin of influenza A viruses : evolutionary constraints and perspectives for the development of broad-spectrum therapeutic antivirals and universal vaccines. Human health and pathology. Université Paris Cité, 2022. English. NNT : 2022UNIP5179 . tel-04658349

**HAL Id: tel-04658349**

**<https://theses.hal.science/tel-04658349>**

Submitted on 22 Jul 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**Université Paris Cité**  
École Doctorale 562 – **Bio Sorbonne Paris Cité**  
**Institut Pasteur, Paris**  
Unité de Génétique Moléculaire des Virus à ARN

Thèse de doctorat d'infectiologie

# **Structural and functional analysis of physicochemical interactions of the hemagglutinin of influenza A viruses: evolutionary constraints and perspectives for the development of broad-spectrum therapeutic antivirals and universal vaccines**

**Valentin OZEEL**

Dirigée par Pr. Sylvie van der Werf / Pr. Anne Vanet

Présentée et soutenue publiquement à Paris, le **15 décembre 2022**

Devant un jury composé de :

Rapporteur :	M. Yves Gaudin	DR	- Université Paris-Saclay, Institut de Biologie Intégrative de la Cellule
Rapporteuse :	Mme. Mariette Ducatez	DR	- IRAE, École Nationale Vétérinaire de Toulouse
Examineur :	M. Olivier Sperandio	CR	- Université Paris Cité, Institut Pasteur
Examineur :	M. Manuel Rosa-Calatrava	DR	- Université Claude Bernard Lyon 1, Centre International de Recherche en Infectiologie
Examinatrice :	Mme. Caroline Goujon	CR	- Université de Montpellier, Institut de Recherche en Infectiologie
Examineur :	M. Pierre-Emmanuel Ceccaldi	PU	- Université Paris Cité, Institut Pasteur
Directrice :	Mme. Sylvie van der Werf	PU	- Université Paris Cité, Institut Pasteur
Co-directrice :	Mme. Anne Vanet	PU	- Université Paris Cité, Institut Jacques Monod



**Université Paris Cité**  
École Doctorale 562 – **Bio Sorbonne Paris Cité**  
**Institut Pasteur, Paris**  
Unit of Molecular Genetics of RNA Viruses

Doctoral dissertation of infectiology

# **Structural and functional analysis of physicochemical interactions of the hemagglutinin of influenza A viruses: evolutionary constraints and perspectives for the development of broad-spectrum therapeutic antivirals and universal vaccines**

**Valentin OZEEL**

Directors: Pr. Sylvie van der Werf / Pr. Anne Vanet

Publicly presented and defended in Paris, the **15th of December 2022**

In front of a jury composed of:

Reviewer:	M. Yves Gaudin	DR	- Université Paris-Saclay, Institut de Biologie Intégrative de la Cellule
Reviewer:	Mme. Mariette Ducatez	DR	- IRAE, École Nationale Vétérinaire de Toulouse
Examinator:	M. Olivier Sperandio	CR	- Université Paris Cité, Institut Pasteur
Examinator:	M. Manuel Rosa-Calatrava	DR	- Université Claude Bernard Lyon 1, Centre International de Recherche en Infectiologie
Examinator:	Mme. Caroline Goujon	CR	- Université de Montpellier, Institut de Recherche en Infectiologie
Examinator:	M. Pierre-Emmanuel Ceccaldi	PU	- Université Paris Cité, Institut Pasteur
Director:	Mme. Sylvie van der Werf	PU	- Université Paris Cité, Institut Pasteur
Co-director:	Mme. Anne Vanet	PU	- Université Paris Cité, Institut Jacques Monod

Do not whine.

Do not complain.

In some situations, working harder is the only thing you can do.



## Acknowledgements

I would like to sincerely acknowledge the members of my jury for agreeing to take some of their valuable time to evaluate this work. I thank Dr. Yves Gaudin and Dr. Mariette Ducatez for accepting to dedicate their time to read and judge my work in depth. I also thank Dr. Olivier Sperandio, Dr. Manuel Rosa-Calatrava, Dr Caroline Goujon and Pr. Pierre-Emmanuel Ceccaldi for examining my work. I look forward to discussing and debating my thesis with you all.

I thank Sylvie for her support and for welcoming me in her laboratory where I had the opportunity to evolve as a scientist and to develop my avidity to discover and learn always more. I thank Anne for opening my eyes on the crucial importance of multidisciplinary approaches, I will continue along this path. I also thank the Pasteur Institute for providing a nice environment to perform my research and the Université Paris Cité (Doctoral school 562, BioSPC) for financing my thesis grant.

I would like to thank Lius for being by my side and supporting me in so many circumstances. We met at the worst possible time: at the very beginning of my thesis. You still chose to be with me even though I could not give you the time and attention you deserve. I will do my best to make sure you thrive in the future. Have faith in yourself, you are exceptional and I couldn't be happier to be with you.

Merci à toute ma famille de m'avoir soutenu et de ne pas avoir douté de mes choix depuis toujours. Merci à mes parents Nathalie et Bruno et à ma sœur Alexiane de s'occuper de papi et mamie. Alex, bon courage pour l'école, ça va le faire tranquille ! Papa, maman je vous souhaite de pouvoir bronzer au bord d'une piscine toute l'année dans le sud au plus vite ! Irène et Gabiz je pense fort à vous ! Je vous souhaite le meilleur à tous et vous serez les bienvenus à la montagne.

Thanks to Marquinhos for coming with me to Panam. These years weren't the best time for me to get some fun but I hope we could change that in the future and that we could regularly hang out along the lake soon! I wish you all the best for your thesis and for your life in general brother. If you need my help for anything, you can always call me.

I would like to thank Catoch the tardigrade and Andrew for their continuous support. You are such amazing individuals with good values. You are an inspiration. I wish you and your family all the best. Raphael and Thomas, I wish you guys to continue along this path and to build a great future.

Floflo and Tim, thanks for your support. Fortunately, you were there to brighten my journey whether it was in the office or in the P2+. Flo thanks for the kind messages I could find late in the evening after finishing my experiments. Tim thanks for never stealing my pipets. I'm happy for you guys. I promise we'll do the escape game and we'll also hike! I wish you guys all the best at the lab and in your life.

Mariem, thanks for trusting me and for your kindness. I hope I taught you some great things for you to succeed in whatever you want to accomplish. I'm not worried at all for you, you're going to succeed at whatever you want to do. Thanks again for the figurine I really appreciated that. It's on my desk btw!

Merci à Sylvain de m'avoir fait prendre des pauses pour partager un café et rigoler un peu. J'espère vraiment que tu auras ton post d'ingé très prochainement et qu'on pourra boire un petit verre pour fêter ça ! T'es un bon. Mec j'y pense j't'ai toujours pas vu avec tes baskets bicolores de jeunes là du coup !

KamKam merci de prendre de mes nouvelles dans les moments compliqués. Je ne doute pas que tu feras de belles choses à Pasteur. Essaie d'être moins gentille un peu. On se capte à Sainte-Suzanne !

Merci Bernadette pour toutes les petites discussions, pour avoir partagé ton bureau avec moi et pour m'avoir encouragé ! J'ai été ravi de pouvoir visiter ton palace doré à Neuilly. Bientôt la retraite là c'est bon ! Profite en bien quand ça sera le grand jour. Tout le meilleur pour ta petite famille.

Thanks Sinem for trusting me. It was a pleasure to supervise you even for such a short time. Thanks for bringing a good mood every day!

Modzilla, thanks for your calm and kindness. I wish you all the best for your thesis and then to find a position in Canada. You'll be a great teacher.

J-b thanks for the few times we had the opportunity to have a drink. I hope you'll enjoy your next position, hopefully a permanent one!

Thanks MélaNight for the great times, your smile and all the stupid songs you made for me! Thanks also for the kind message I could find during the evening. I hope you're enjoying your new life and that you'll evolve in your new environment!

Thanks also to Pablito, Thomas, Chloe and Adil (and Camille again) for the few but great times we had together during these three last years. You're welcome at home whenever you want.

Anthony, a special thank you for your ears. You know what I think about all of this, nevertheless it was great to have your company from time to time. You a nice guy, I wish you all the best.

# Summary

Influenza A viruses (IAVs) cause yearly epidemics and occasional pandemics that represent a global health threat of major concern. Virus entry within cells is carried out by the hemagglutinin (HA), a trimeric surface glycoprotein of which each protomer is constituted of two sub-units (HA<sub>1</sub>, HA<sub>2</sub>). The HA binds its receptor at the cell surface and induces the endocytosis of the particle. The exposure of the HA to acidic pH upon maturation of the endosome triggers a drastic conformational change (CC), which enables the fusion between endosomal and viral membranes. The HA is intrinsically extremely variable, it has been classified in subtypes, clades and groups.

Based on 269 trimeric prefusion structures, we computed six interaction types, including electrostatic interactions, along with their distribution intra- and inter-protomer (ITA-p/ITR-p) and intra- and inter-sub-unit. We revealed distinct clade-dependent physicochemical profiles of interaction networks. These different profiles of networks enable the folding of highly similar trimeric metastable structures carrying identical function and thus suggested that HA evolution is constrained accordingly and that the function may be carried out by different CC molecular drivers.

We then computed the conserved interactions across HA classifications. We characterized a buried complex consisting of two connected ITR-p salt-bridges, located at a crucial area for the CC. The first is the ITR-p HA<sub>2</sub>-HA<sub>2</sub> (E74<sub>2</sub>-R76<sub>2</sub>) salt-bridge, strictly conserved among all HAs except clade 7 HAs and H17. The second is the ITR-p HA<sub>1</sub>-HA<sub>2</sub> (E104<sub>1</sub>-R76<sub>2</sub>) salt-bridge, conserved among almost every subtype of group 1 HAs, whereas the ITR-p HA<sub>2</sub>-HA<sub>2</sub> (R76<sub>2</sub>-E81<sub>2</sub>) salt-bridge is strictly conserved among clade 3 HAs. The functional importance of the complex was analyzed through mutagenesis, fusion assays and reverse genetics in two viruses from clade 1 (H1) and 3 (H3), which displayed distinct physicochemical profiles. Mutations were tolerated in the H3 context, *in vitro*, whereas in the H1 context any mutation disrupting the complex resulted in a loss of function and unviable mutant viruses. The results therefore indicated that HA evolution is constrained according to the underlying physicochemical profile of interaction networks. The strict conservation of the complex in structures sharing the H1 physicochemical profile and several other factors, that include the identical geometric arrangement of the complex in space, suggested that the essential nature of the complex in H1 is likely to be extrapolated to other HAs sharing the same physicochemical profile.

Finally, we tracked the modulation of interactions during the CC based of structures of fusion intermediates (H3). We revealed various molecular drivers, especially the disruption of highly conserved ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges located at the membrane-proximal domain, which might be key in the context of the clade 3 physicochemical profile.

Overall, assessments of the functional importance of the conserved ITR-p salt-bridges complex in different environments highlighted that underlying clade-dependent physicochemical profile of interaction networks influences CC molecular drivers and how function is performed. These profiles influence the mutational tolerance of specific residues and thereby limit HA evolution. Our results demonstrated that comprehensive analysis of interaction networks in the context of HA variability enables to better understand its evolution as well as to take advantage of the variability to identify crucial conserved interactions. Most importantly, we highlighted the value of combining interaction conservation assessments and functional validation to provide insightful data for the research of broad and non-escapable therapeutic targets and vaccine epitopes.

**Keywords:** influenza, hemagglutinin, conservation, interactions, salt-bridges, physicochemical profile

# Résumé

Les virus de la grippe provoquent chaque année des épidémies et parfois des pandémies qui représentent une menace mondiale. L'entrée du virus dans les cellules est assurée par l'hémagglutinine (HA), une glycoprotéine trimérique de surface dont chaque protomère est constitué de deux sous-unités (HA<sub>1</sub>, HA<sub>2</sub>). La HA se lie à son récepteur à la surface de la cellule, ce qui induit l'endocytose de la particule. L'exposition de la HA au pH acide lors de la maturation de l'endosome déclenche un changement conformationnel (CC) drastique, qui permet la fusion entre les membranes endosomale et virale. La HA est extrêmement variable, et a été classée en sous-types, clades et groupes.

Sur la base de 269 structures trimériques pré-fusion, nous avons calculé six types d'interactions ainsi que leur distribution intra- et inter-protomères (ITA-p/ITR-p) et intra- et inter-sous-unités. Nous avons révélé des profils physico-chimiques de réseaux d'interactions distincts en fonction des clades. Ces réseaux permettent le repliement de structures métastables trimériques très similaires, ayant une fonction identique, ce qui suggère que l'évolution des HA est limitée en conséquence.

Nous avons ensuite calculé les interactions conservées en fonction des classifications et caractérisé un complexe composé de deux ponts salins ITR-p connectés, situé dans une zone cruciale pour le CC. Le premier est le pont salin ITR-p HA<sub>2</sub>-HA<sub>2</sub> (E74<sub>2</sub>-R76<sub>2</sub>), strictement conservé parmi toutes les HA sauf celles du clade 7 et H17. Le second est le pont salin ITR-p HA<sub>1</sub>-HA<sub>2</sub> (E104<sub>1</sub>-R76<sub>2</sub>), conservé parmi presque tous les sous-types de HA du groupe 1, tandis que le pont salin ITR-p HA<sub>2</sub>-HA<sub>2</sub> (R76<sub>2</sub>-E81<sub>2</sub>) est strictement conservé au sein du clade 3. L'importance fonctionnelle du complexe a été caractérisée par mutagenèse, test de fusion et génétique inverse dans deux virus du clade 1 (H1) et 3 (H3), qui présentent des profils physico-chimiques distincts. Les mutations ont été tolérées dans le contexte H3, *in vitro*, tandis que dans le contexte H1, toute mutation perturbant le complexe a donné lieu à une perte de fonction et des virus mutants non viables. Les résultats indiquent donc que l'évolution de la HA est limitée par le profil physico-chimique des réseaux d'interactions et suggèrent différents moteurs moléculaires du CC. La conservation du complexe au sein des structures partageant le profil physico-chimique H1, et plusieurs autres facteurs dont l'arrangement géométrique identique du complexe dans l'espace, suggèrent que la nature essentielle du complexe est susceptible d'être extrapolée aux HA partageant le même profil physico-chimique.

Enfin, nous avons suivi la modulation des interactions au cours du CC à partir des structures des intermédiaires de fusion (H3). Nous avons révélé divers moteurs moléculaires, notamment la rupture de nombreux ponts salins ITR-p HA<sub>2</sub>-HA<sub>2</sub> très conservés, à la base du trimère.

Dans l'ensemble, l'évaluation de l'importance fonctionnelle, dans différents contextes, du complexe de ponts salins ITR-p largement conservé a mis en évidence que le profil physico-chimique des réseaux d'interactions influence les moteurs moléculaires du CC et la façon dont la fonction est réalisée. Les profils physico-chimiques des réseaux d'interactions influencent la tolérance mutationnelle de résidus spécifiques, contraignant ainsi l'évolution de la HA. Nos résultats ont démontré que l'analyse des réseaux d'interactions dans le contexte hypervariable de la HA permet de mieux comprendre son évolution et de tirer parti de la variabilité pour identifier des interactions conservées essentielles. Enfin, nous avons souligné l'intérêt de combiner l'évaluation de la conservation des interactions et leur validation fonctionnelle afin de fournir des données bénéfiques pour la recherche de cibles thérapeutiques et d'épitopes vaccinaux larges ne pouvant muter.

**Mots clés :** grippe, hémagglutinine, conservation, interactions, pont-salin, profile physico-chimique

Titre en français : **Analyse structurale et fonctionnelle des interactions physico-chimiques de l'hémagglutinine des virus de la grippe de type A : contraintes évolutives et perspectives pour le développement d'antiviraux thérapeutiques à large spectre et de vaccins universels**



## Résumé étendu

Les virus de la grippe provoquent chaque année des épidémies et parfois des pandémies qui représentent une menace sanitaire mondiale très préoccupante. L'entrée du virus dans les cellules est assurée par l'hémagglutinine (HA), une glycoprotéine trimérique de surface hypervariable dont chaque protomère est constitué de deux sous-unités (HA<sub>1</sub>, HA<sub>2</sub>)<sup>1</sup>. La HA se lie à son récepteur à la surface de la cellule, ce qui induit l'endocytose de la particule virale. L'exposition de la HA au pH acide lors de la maturation de l'endosome déclenche un changement conformationnel (CC) drastique, impliquant notamment une transition de la boucle-b en hélice alpha, qui permet la fusion entre les membranes endosomale et virale<sup>1</sup>.

La structure et la fonction des protéines dépendent principalement des propriétés physicochimiques conférées par des réseaux complexes d'interactions entre acides aminés qui initient et maintiennent un repliement approprié et de potentielles propriétés métastables<sup>2</sup>. Des études antigéniques, phylogénétiques, et de comparaisons structurales ont été réalisées sur la HA et servent de base à la classification en sous-types, groupes et clades des virus grippaux de type A. Le groupe 1 est composé du clade 1 (H1, H2, H5, H6), du clade 9 (H8, H9, H12), du clade 11 (H11, H13, H16) et du clade 17 (H17, H18), tandis que le groupe 2 est composé du clade 3 (H3, H4, H14) et du clade 7 (H7, H10, H15)<sup>3-5</sup>. Cependant, aucune analyse complète des arrangements des réseaux d'interactions de la HA n'a été réalisée. Il a également été suggéré que le CC implique une large distribution de réarrangements moléculaires en réponse à l'acidification, mais aucune description du comportement des interactions lors de l'exposition au pH acide n'est disponible<sup>6</sup>.

En outre, la variabilité et la diversité des virus influenza rend l'identification d'antiviraux et de vaccins à large spectre très difficile, et implique la possibilité d'émergence de variants résistants aux antiviraux ou échappant à la réponse immunitaire conférée par l'infection naturelle ou la vaccination<sup>7,8</sup>. L'identification d'antiviraux et de vaccins à large spectre pour lesquels les virus ne pourraient pas générer de résistance permettrait de contrôler les épidémies annuelles et serait d'une importance majeure dans le cadre de la préparation à l'émergence d'une future pandémie. Ainsi, exploiter la variabilité de la HA pour identifier des interactions largement conservées et essentielles sur le plan fonctionnel, qui ne seraient donc pas enclines à muter, apparaît comme une approche d'intérêt.

Sur la base de 281 structures trimériques de HA, les interactions entre acides aminés (AA) à l'échelle atomique ont été étudiées par analyse computationnelle. Nous avons considéré six types d'interactions : interactions hydrophobes, interactions hydrogènes chaîne latérale/squelette carboné (SD-C), interactions hydrogènes entre chaînes latérales (SD-SD), ponts salins, répulsions anion/anion (-/-) et répulsions cation/cation (+/+). Ces interactions ont été sous-divisées selon leur disposition au sein ou entre les protomères et les sous-unités. Ainsi, douze conditions concernant la distribution des interactions ont été analysées : interactions intra-protomères (ITA-p), inter-protomères (ITR-p) et globales au sein ou entre les sous-unités (HA<sub>1</sub>-HA<sub>1</sub>, HA<sub>1</sub>-HA<sub>2</sub>, HA<sub>2</sub>-HA<sub>2</sub>) et pour la HA dans son ensemble.

Pour étudier les corrélations entre la composition/distribution des réseaux d'interactions et l'évolution des HA, nous avons effectué des analyses en composantes principales (ACP). Les ACP ont montré que les compositions et distributions d'interactions varient selon les clades donnant lieu à des profils physico-chimiques distincts permettant de conserver la structure pré-fusion (Annexe 1 - Fig. 1d).

La variabilité de la HA influence donc la composition et la distribution des réseaux d'interactions, le clade 1 étant clairement distinct du clade 3 tandis que le clade 7 partage des caractéristiques avec les deux autres clades. Ces profils physico-chimiques des réseaux d'interactions permettent le repliement de structures trimériques très similaires (Annexe 1 -

Données étendues Fig.1b) mais préservent également les propriétés métastables et permettent l'exécution du CC déclenché par l'exposition au pH acide. Par conséquent, les possibilités de mutations devraient être limitées en fonction du profil physicochimique sous-jacent, et les moteurs moléculaires du CC pourraient également différer.

Sur la base de la numérotation universelle que nous avons implémentée (voir Méthodes Annexe 1), le degré de conservation des interactions a été calculé. Malgré la variabilité intrinsèque et la dynamique de la HA, des interactions conservées ont pu être identifiées au sein des sous-types mais aussi plus largement au sein des clades, des groupes et globalement (Annexe 1 - Données étendues Tableau 2).

On s'attend à ce que les interactions conservées impliquées dans des fonctions multiples aient une importance cruciale et une faible tolérance aux mutations. Ainsi, nous nous sommes concentrés sur les ponts salins ITR-p HA<sub>2</sub>-HA<sub>2</sub> en raison de: (i) leur influence potentielle sur le repliement et le maintien de la structure quaternaire des trimères de HA ; (ii) leurs propriétés de senseur de pH leur permettant de répondre à une exposition à un pH acide (ces interactions pouvant être rompues à pH acide) ; (iii) leur rôle dans la restriction du mouvement des protéines<sup>9-12</sup>.

Nous avons identifié un complexe, à proximité de la boucle-b, constitué de deux ponts salins ITR-p connectés que nous appellerons " le complexe SB " (Annexe 1 - Fig. 2a). Le premier pont salin est le pont salin ITR-p HA<sub>2</sub>-HA<sub>2</sub> (E74<sub>2</sub>-R76<sub>2</sub>), strictement conservé parmi toutes les HA à l'exception des HA du clade 7 et de la H17 (Annexe 1 - Fig. 2b). Le deuxième pont salin est le pont salin ITR-p HA<sub>1</sub>-HA<sub>2</sub> (E104<sub>1</sub>-R76<sub>2</sub>), conservé parmi presque tous les sous-types de HA du groupe 1, tandis que le pont salin ITR-p HA<sub>2</sub>-HA<sub>2</sub> (R76<sub>2</sub>-E81<sub>2</sub>) est strictement conservé parmi les HA du clade 3 (Annexe 1 - Fig. 2b). Les AA impliqués étaient strictement conservés dans les séquences des virus circulants (annexe 1 - figure 2b, tableau supplémentaire 1).

De plus, pour les HA du clade 1, du clade 9 et du clade 11, l'arrangement géométrique du complexe dans l'espace était strictement identique, avec des contacts optimisés entre les groupes d'atomes ionisables (Annexe 1 - Fig. 2a). L'arrangement spatial du complexe était identique entre les HA du clade 3 mais légèrement différent par rapport aux autres clades en raison de l'absence de lien avec HA<sub>1</sub> (Annexe 1 - Fig. 2a). En revanche, le complexe SB n'était pas formé dans les HA du clade 7 en raison de la présence d'une glutamine (Q) en position 76<sub>2</sub> plutôt que d'une arginine (R) (Annexe 1 - Fig. 2a, 2b).

L'importance fonctionnelle du complexe SB a été évaluée expérimentalement dans différents contextes de profils physico-chimiques. Nous avons introduit des mutations dans la HA de virus de clade 1 et de clade 3 : A/Puerto Rico/8/1934 (H1N1) et A/Centre/1003/2012 (H3N2). Les substitutions ont été introduites à chacune des trois positions impliquées dans le complexe SB : 104<sub>1</sub>, 74<sub>2</sub>, 76<sub>2</sub> pour H1N1 et 74<sub>2</sub>, 76<sub>2</sub>, 81<sub>2</sub> pour H3N2 (Annexe 1 - Fig. 2c). Pour chacune des trois positions étudiées, cinq mutants différents ont été produits par mutagenèse dirigée afin de générer un large éventail d'effets physico-chimiques (Annexe 1 - Fig. 2c). Un double mutant (R74<sub>2</sub>-E76<sub>2</sub>) a également été produit.

Les mutations ont donné lieu à des variations concernant la capacité de la HA à être transportée à la surface cellulaire et à être activée par clivage protéolytique. Néanmoins, aucune des mutations n'a aboli l'expression de la protéine ni sa sensibilité au clivage par une protéase exogène (Annexe 1 - Données étendues Fig. 6).

Les tests de fusion ont indiqué que le complexe SB n'est pas essentiel pour déclencher le processus de fusion dans le contexte H3 alors qu'il est vital dans le contexte H1 (Annexe 1 - Fig. 3). En effet, toutes les mutations ont entraîné une efficacité de fusion comparable à celle de la protéine sauvage (wild-type, WT) dans le contexte H3 (Annexe 1 - Fig. 3a, 3c). En

revanche, toute mutation perturbant le complexe a entraîné une perte de fonction dans le contexte H1, tandis que les mutations conservatrices E104<sub>1</sub>D et E74<sub>2</sub>D ont altéré le potentiel de fusion (Annexe 1 - Fig. 3b, 3d).

Des expériences de génétique inverse ont permis de valider ces résultats. Le large éventail de mutations a été toléré dans le contexte H3 et les virus mutants ont été produits à des titres similaires à ceux du virus WT (Annexe 1 - Fig. 4a). Au contraire, l'importance vitale du complexe SB pour la viabilité virale a été confirmée dans le contexte H1. Les virus mutants portant des mutations non conservatrices n'ont pas pu être produits, tandis que les virus mutants E104<sub>1</sub>D et E74<sub>2</sub>D ont été obtenus à des titres réduits par rapport au WT (Annexe 1 - Fig. 4b).

Pour mieux comprendre les différences expérimentales obtenues, nous avons représenté les ponts salins ITR-p HA<sub>2</sub>-HA<sub>2</sub> sur les structures de représentants des différents sous-types (Annexe 1 - Données étendues Fig. 7). Les ponts salins ITR-p HA<sub>2</sub>-HA<sub>2</sub> étaient situés uniquement au sommet des hélices-a dans les HA du clade 1 (Annexe 1 - Données étendues Fig. 7). Ceci, avec la conservation stricte du complexe SB dans les structures partageant le profil physico-chimique H1, l'arrangement géométrique identique du complexe SB dans l'espace, ainsi que la conservation stricte des résidus impliqués dans les séquences, a souligné que la nature essentielle du complexe pour H1 est susceptible d'être extrapolée à d'autres HA partageant le même profil physico-chimique, y compris celles de virus à haut potentiel pandémique tels que les virus H5N1.

En revanche, pour les HA du clade 3, quelques ponts salins ITR-p HA<sub>2</sub>-HA<sub>2</sub> étaient identifiés au sommet des hélices-a, mais la plupart d'entre eux étaient situés au niveau du domaine proche de la membrane, tandis que pour les HA du clade 7, ils se trouvaient exclusivement au niveau du domaine proche de la membrane (Annexe 1 - Données étendues Fig. 7). Ces ponts salins ITR-p HA<sub>2</sub>-HA<sub>2</sub> proches de la membrane comprennent un complexe ITR-p HA<sub>2</sub>-HA<sub>2</sub> constitué des ponts salins R127<sub>2</sub>-E131<sub>2</sub> et E131<sub>2</sub>-R163<sub>2</sub> hautement conservés, respectivement dans 97% et 94% des structures de HA du groupe 2.

Les analyses de la modulation des réseaux d'interactions au cours du CC, basées sur l'analyse des structures des intermédiaires de fusion (H3) ont révélé le rôle central de certaines interactions (Annexe 1 - Données étendues Tableau 3, Annexe 7). La rupture des ponts salins ITR-p HA<sub>2</sub>-HA<sub>2</sub> ainsi que la formation de répulsions ITR-p HA<sub>2</sub>-HA<sub>2</sub> +/+, expliquent le redressement des hélices-c initialement courbées et leur éloignement, induisant la libération du peptide de fusion initialement enfoui dans sa cavité à l'interface HA<sub>2</sub> (Annexe 1 - Données étendues Fig. 8). Avec l'analyse de la distribution des ponts salins ITR-p HA<sub>2</sub>-HA<sub>2</sub> et les données expérimentales sur le complexe SB, ces résultats suggèrent que les moteurs moléculaires du CC diffèrent selon les profils physico-chimiques.

Dans l'ensemble, ce travail a montré que la variabilité intrinsèque de la HA est reflétée au niveau des réseaux d'interactions et se traduit par des profils physico-chimiques distincts préservant l'état métastable de pré-fusion et remplissant les conditions pour l'exécution du CC. Nous avons confirmé que des mutations identiques au niveau de positions impliquées dans des interactions conservées et introduites dans des HA aux profils physico-chimiques distincts, ont des impacts différents sur la fonction protéique. Les réseaux d'interactions se comportent différemment et limitent la tolérance aux mutations et les voies d'évolution. De plus, nos résultats ont souligné l'intérêt de combiner l'évaluation de la conservation des interactions et les données expérimentales de mutagenèse pour l'identification de cibles thérapeutiques et d'épitopes vaccinaux à large spectre et ne pouvant muter.

## Abbreviations

AA: amino acid

ADCC: antibody-dependent cell cytotoxicity

ADCP: antibody-dependent cell phagocytosis

CC: conformational change

CDC: complement-dependent cytotoxicity

ER: endoplasmic reticulum

IAV: Influenza A virus

ITA-p: intra-protomer

ITR-p: inter-protomer

HA: hemagglutinin

HA<sub>0</sub>: HA precursor

HA<sub>1</sub>: HA sub-unit 1 (head)

HA<sub>2</sub>: HA sub-unit 2 (stem)

M1: matrix protein 1

M2: matrix protein 2 (proton pump)

MMTI: median number of total interactions

NA: neuraminidase

NP: nucleoprotein

NS1: non-structural protein

PA: polymerase acidic protein (endonuclease)

PB1: polymerase basic protein 1 (RNA dependent RNA polymerase)

PB2: polymerase basic protein 2 (cap-binding protein)

PCA: principal component analysis

PDB: protein data bank

pH: potential hydrogen

cRNA: complementary positive-sense RNA

(m)RNA: (messenger) ribonucleic acid

crRNA: complementary positive-sense ribonucleic acid

RBS: receptor binding site

SD-C: sidechain-corechain

SD-SD: sidechain-sidechain

The SB complex: highly conserved ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges complex

VE: vestigial esterase domain

cRNP: complementary positive-sense ribonucleoprotein

vRNP: viral ribonucleoprotein

-/- : anion/anion

+/+ : cation/cation

X<sub>1</sub>: Residue X on HA<sub>1</sub>

X<sub>2</sub>: Residue X on HA<sub>2</sub>

# Table of contents

<b>SUMMARY</b>	<b>6</b>
<hr/>	
<b>CHAPTER 1. GENERAL INTRODUCTION</b>	<b>16</b>
<hr/>	
<b>1.1 INFLUENZA A VIRUSES: A GLOBAL ISSUE</b>	<b>17</b>
<b>1.2 INFLUENZA A VIRUSES: THE VIRAL PARTICLE</b>	<b>18</b>
<b>1.3 INFLUENZA A VIRUSES: THE VIRAL MULTIPLICATION CYCLE</b>	<b>20</b>
<b>1.4 INFLUENZA A VIRUS VARIABILITY AND DIVERSITY</b>	<b>21</b>
<b>1.5 INFLUENZA VACCINES</b>	<b>28</b>
<b>1.6 ANTIVIRALS</b>	<b>30</b>
<b>1.7 THE HEMAGGLUTININ (HA)</b>	<b>32</b>
<b>1.8 INTER-AMINO ACIDS INTERACTIONS</b>	<b>40</b>
<b>CHAPTER 2. OBJECTIVES</b>	<b>44</b>
<hr/>	
<b>CHAPTER 3. RESULTS</b>	<b>47</b>
<hr/>	
<b>3.1 RESULTS SUMMARY</b>	<b>48</b>
3.1.1 DISTINCT PHYSICOCHEMICAL PROFILES PRESERVING HA METASTABLE PREFUSION STATE	48
3.1.2 CORRELATION BETWEEN INTERACTIONS BALANCE AND HA SUBTYPES, HOST-ORIGIN AND MODULATION OF HA STABILITY	50
3.1.3 CHARACTERIZATION OF HIGHLY CONSERVED INTERACTIONS ON PRE-FUSION STRUCTURES	51
3.1.4 IDENTIFICATION AND SPATIAL ARRANGEMENT OF A BROADLY AND STRICTLY CONSERVED ITR-P SALT-BRIDGES COMPLEX	51
3.1.5 FUNCTIONAL RELEVANCE OF THE SB COMPLEX DEPENDS ON THE PHYSICOCHEMICAL PROFILE	54
3.1.6 CLADE-DEPENDENT DISTRIBUTION OF ITR-P HA <sub>2</sub> -HA <sub>2</sub> SALT-BRIDGES AND ROLE DURING THE CC	58
<b>3.2 ANNEX 1: FUNCTIONAL RELEVANCE OF CONSERVED INTERACTIONS IN INFLUENZA A VIRUS HEMAGGLUTININ DEPENDS ON PHYSICOCHEMICAL PROFILE (MANUSCRIPT IN PREPARATION)</b>	<b>62</b>
3.2.1 ABSTRACT	62
3.2.2 INTRODUCTION	63
3.2.3 RESULTS	65
3.2.4 DISCUSSION	75
3.2.5 FIGURES	81
3.2.6 EXTENDED DATA FIGURES/TABLES	88
3.2.7 SUPPLEMENTARY TABLES/FIGURES	112
3.2.8 MATERIAL AND METHODS	120
<b>3.3 ANNEX 2: OVERALL DIFFERENCES IN INTERACTION NETWORKS BETWEEN CLADES</b>	<b>136</b>
<b>3.4 ANNEX 3: CORRELATION BETWEEN INTERACTIONS BALANCE AND SUBTYPE- OR HOST-SPECIFIC HA STABILITY</b>	<b>140</b>
<b>3.5 ANNEX 4: ITR-P INTERACTIONS ARE LESS TOLERANT TO MUTATIONS THAT PRESERVE THE INTERACTION TYPE</b>	<b>155</b>
<b>3.6 ANNEX 5: INFLUENCE OF THE CONSERVED SALT-BRIDGES ON POSITIONING OF THE B-LOOP</b>	<b>157</b>
<b>3.7 ANNEX 6: INFLUENCE OF THE SB COMPLEX ON HA STABILITY</b>	<b>160</b>
<b>3.8 ANNEX 7: TRACKING OF AMINO-ACID INTERACTIONS UPON CONFORMATIONAL CHANGE</b>	<b>162</b>
<b>CHAPTER 4. DISCUSSION</b>	<b>166</b>
<hr/>	
<b>BIBLIOGRAPHY</b>	<b>184</b>
<hr/>	

# **LIST OF FIGURES AND TABLES**

## **CHAPTER 1 AND 3.1**

FIGURE 1: INFLUENZA A VIRUS PARTICLE. ....	18
FIGURE 2. ORGANIZATION OF THE GENOME OF INFLUENZA A VIRUSES.....	19
FIGURE 3: INFLUENZA A VIRUS MULTIPLICATION CYCLE (FROM KRAMMER ET AL <sup>19</sup> ).....	21
FIGURE 4: PHYLOGENETIC TREE REPRESENTING HA AND NA DIVERSITY .....	22
FIGURE 5: INFLUENZA A VIRUS HOSTS (FROM LONG ET AL <sup>23</sup> ) .....	24
FIGURE 6: INFLUENZA A VIRUS VARIABILITY (FROM KRAMMER ET AL <sup>19</sup> AND YAMAYOSHI ET AL <sup>48</sup> ).....	25
FIGURE 7: PANDEMICS CAUSED BY INFLUENZA A VIRUSES (FROM WELDEL ET AL <sup>53</sup> ).....	26
FIGURE 8: ORGANIZATION OF THE HEMAGGLUTININ GENE.....	32
FIGURE 9: STRUCTURAL DESCRIPTION OF THE HA (H1 AND H3).....	34
FIGURE 10: VIRAL ENTRY PROCESS AND CONFORMATIONAL CHANGE OF THE HA.....	35
FIGURE 11: THE FUSION BETWEEN THE VIRAL AND ENDOSOMAL MEMBRANE IS TRIGGERED UPON THE TRANSITION OF THE HA FROM ITS PRE-FUSION STATE TO ITS POST-FUSION STATE .....	36
FIGURE 12: STRUCTURE OF H3 FUSION INTERMEDIATES .....	38
FIGURE 13: RESIDUES INVOLVED IN COMMON INTERACTIONS.....	42
FIGURE 14: DIFFERENT PROFILES OF INTERACTION NETWORKS ENABLE THE FOLDING OF SIMILAR STRUCTURES CARRYING IDENTICAL FUNCTION .....	49
FIGURE 15: SPATIAL ARRANGEMENT OF THE ITR-P HA <sub>2</sub> -HA <sub>2</sub> SALT-BRIDGES COMPLEX IN DIFFERENT SUBTYPES .....	53
FIGURE 16 : FUSION POTENTIAL OF MUTANT HAS.....	55
FIGURE 17: VIABILITY AND FITNESS OF MUTANT VIRUSES.....	57
FIGURE 18: ITR-P HA <sub>2</sub> -HA <sub>2</sub> SALT-BRIDGES DISTRIBUTION THROUGHOUT THE STEM.....	58
FIGURE 19: ITR-P HA <sub>2</sub> -HA <sub>2</sub> SALT-BRIDGES DISRUPTION AND ITR-P HA <sub>2</sub> -HA <sub>2</sub> +/+ REPULSIONS FORMATION DURING THE CONFORMATIONAL CHANGE .....	60


## **CHAPTER 3.2 – 3.8**

### **3.2**

ANNEX 1 - FIGURE 1. DIVERGENCES IN PHYSICOCHEMICAL INTERACTION NETWORKS ORGANIZATION. ....	81
ANNEX 1 - FIGURE 2. CONSERVATION OF THE ITR-P HA <sub>2</sub> -HA <sub>2</sub> SALT-BRIDGES COMPLEX AND MUTATIONAL STRATEGY TO ASSESS ITS FUNCTIONAL IMPORTANCE. ....	83
ANNEX 1 - FIGURE 3. FUSION EFFICIENCY OF MUTANT HAS. ....	85
ANNEX 1 - FIGURE 4. VIABILITY AND FITNESS OF MUTANT VIRUSES. ....	87
ANNEX 1 - EXTENDED DATA FIGURE 1. HA TRIMER AND STRUCTURAL ALIGNMENTS. ....	88
ANNEX 1 - EXTENDED DATA FIGURE 2. COMPARISON OF MEDIAN INTERACTION VALUES BETWEEN CLADES. ....	90
ANNEX 1 - EXTENDED DATA FIGURE 3. HA PHYSICOCHEMICAL PROFILES DO NOT SEGREGATE ACCORDING TO HOST OF ORIGIN BY PRINCIPAL COMPONENT ANALYSIS (PCA). ....	91
ANNEX 1 - EXTENDED DATA FIGURE 4. INSTABILITY OF THE ITR-P HA <sub>2</sub> -HA <sub>2</sub> (E74 <sub>2</sub> -Q76 <sub>2</sub> ) HYDROGEN BONDS IN THE H7 CONTEXT. ....	92
ANNEX 1 - EXTENDED DATA FIGURE 5. SOLVENT EXPOSURE OF POSITIONS INVOLVED IN THE CONSERVED COMPLEX. ....	93
ANNEX 1 - EXTENDED DATA FIGURE 6. EXPRESSION OF MUTANT HAS. ....	94

ANNEX 1 - EXTENDED DATA FIGURE 7. ITR-P HA <sub>2</sub> -HA <sub>2</sub> SALT-BRIDGES DISTRIBUTION THROUGHOUT THE STEM. _____	95
ANNEX 1 - EXTENDED DATA FIGURE 8. ITR-P HA <sub>2</sub> -HA <sub>2</sub> SALT-BRIDGES DISRUPTION AND ITR-P HA <sub>2</sub> -HA <sub>2</sub> +/- REPULSIONS FORMATION DURING THE CONFORMATIONAL CHANGE. _____	96
ANNEX 1 - EXTENDED DATA TABLE 1. DATASET OF INFLUENZA A VIRUS HEMAGGLUTININ TRIMERIC STRUCTURES _____	98
ANNEX 1 - EXTENDED DATA TABLE 2. STRICT CONSERVATION PERCENTAGE OF INTERACTION CONDITIONS ACROSS CLASSIFICATIONS. _____	108
ANNEX 1 - EXTENDED DATA TABLE 3. INTERACTIONS TRACKING DURING CONFORMATIONAL CHANGE. _____	110
ANNEX 1 - SUPPLEMENTARY FIGURE 1. INTERACTION TYPES AND DISTRIBUTIONS VALUES PLOTTED ACCORDING TO HA GROUPS. _____	112
ANNEX 1 - SUPPLEMENTARY TABLE 1. PERCENTAGE OF CONSERVATION OF RESIDUES INVOLVED IN THE SB COMPLEX IN SEQUENCES OF CIRCULATING VIRUSES AND NUMBER OF SEQUENCES WITH ALTERNATIVE AMINO ACID. _____	119
<b>3.3</b>	
ANNEX 2 - TABLE 1. COMPARISONS OF THE MEDIAN VALUES (BETWEEN CLADES) OF INTERACTION NUMBERS ACCORDING TO TYPES AND LOCATION AND COMPARISONS OF THE MEAN OF CLADES MEDIAN VALUES COMPARED TO THE MEAN OF CLADES MEDIAN NUMBER OF TOTAL INTERACTIONS _____	138
<b>3.4</b>	
ANNEX 3 - FIGURE 1. RELATION BETWEEN INTERACTIONS AND HA STABILITY. _____	142
ANNEX 3 - FIGURE 2. INTERACTION TYPES AND DISTRIBUTIONS VALUES PLOTTED ACCORDING TO HA SUBTYPES. _____	144
ANNEX 3 - TABLE 1. INFLUENZA A WILD-TYPE VIRUSES PH OF FUSION. _____	151
<b>3.5</b>	
ANNEX 4 - FIGURE 1. AMINO-ACIDS VARIABILITY AMONGST CONSERVED INTERACTIONS. _____	156
<b>3.6</b>	
ANNEX 5 - FIGURE 1. VISUALIZATION OF THE B-LOOP TURN AND POSITIONING OF THE B-LOOP ACCORDING TO CLASSIFICATIONS. _____	159
<b>3.7</b>	
ANNEX 6 - FIGURE 1. STABILITY OF VIABLE MUTANTS. _____	161
<b>3.8</b>	
ANNEX 7 - TABLE 1. COMPARISONS OF THE INTERACTION MODULATION TRENDS BETWEEN THE TWO SETS OF CC INTERMEDIATES STRUCTURES _____	165





Chapter **1.** General Introduction

## **| 1.1 Influenza A viruses: a global issue**

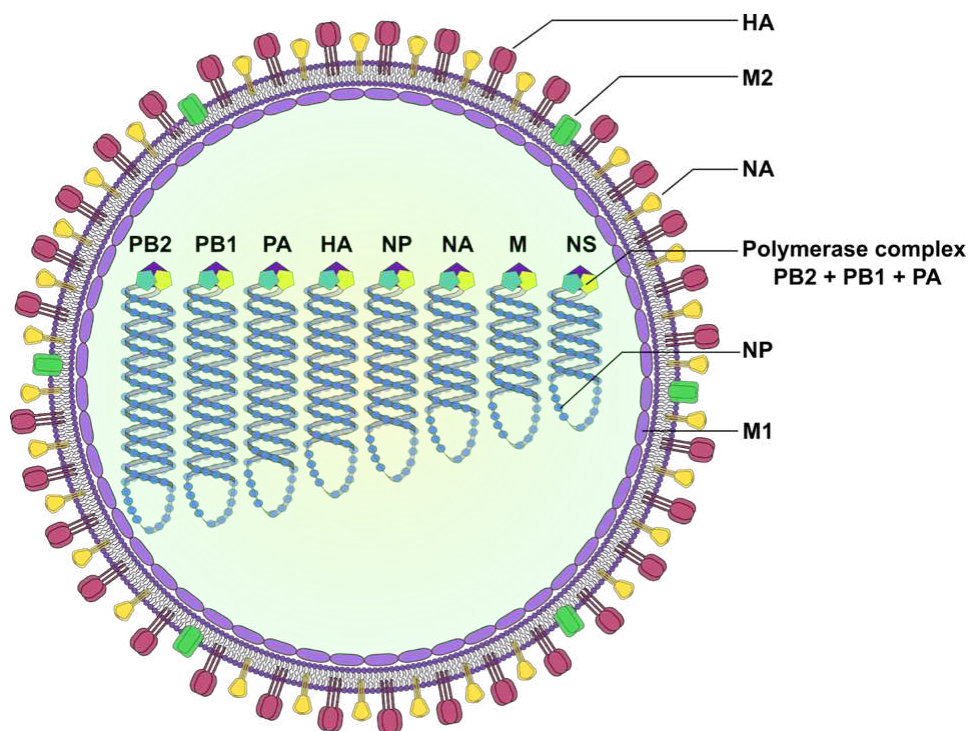
Influenza A viruses (IAVs) are responsible for annual epidemics and constitute a severe burden on morbidity and mortality for humans. According to the World Health Organization, recurring epidemics are responsible for 1 billion infections each year causing 3 to 5 million cases of severe illness worldwide and 290 000 to 650 000 deaths<sup>13</sup>. Moreover, IAVs with zoonotic potential represent a worldwide public health threat of major concern as they are at the source of deadly pandemic viruses<sup>14,15</sup>.

The transmission of IAVs between humans is highly effective. Viruses can be transmitted via respiratory droplets generated by coughing and sneezing of infected individuals, by direct-contact transmission from infected persons, by indirect-contact transmission through contacts with contaminated surfaces or by airborne transmission via inhalation of aerosols containing infectious particles that can remain suspended in the air for hours<sup>16</sup>.

Human IAVs target epithelial cells of the upper and lower respiratory tract. The pathogenicity of IAVs is primarily due to direct damage of infected cells of the respiratory epithelium and to the deleterious impact of lung inflammation generated by immune responses<sup>17</sup>. The invasion of alveolar epithelial cells in the lower respiratory tract and their destruction allow viruses to gain access to endothelial cells that are a major source of pro-inflammatory cytokines driving the magnitude of immune responses<sup>18</sup>. Excessive inflammation combined with the death of alveolar cells result in the inability of the lungs to ensure their function of gas exchange, leading to acute respiratory distress and potentially death in severe cases.

## 1.2 Influenza A viruses: the viral particle

Influenza A viruses (IAVs) are members of the *Orthomyxoviridae* family. They are enveloped particles with a genome made of eight single-stranded negative-sense viral RNAs<sup>19</sup> (**Fig. 1**).



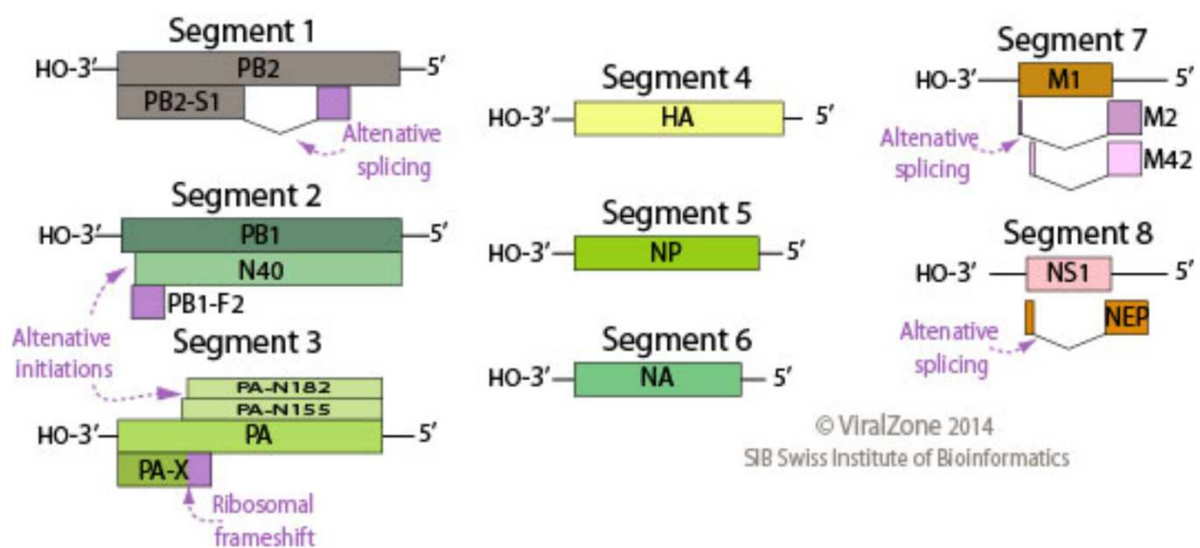
*Figure1: Influenza A virus particle.*

Schematic illustration of the influenza A virus particle. The HA, NA and M2 are the surface proteins. The M1 protein covers the interior of the viral envelope. NP proteins are associated to each of the viral RNA segments, forming ribonucleoproteins. PB2, PB1 and PA form the heterotrimeric viral polymerase complex, attached to each viral ribonucleoprotein.

The RNA segments are associated with the nucleoprotein (NP) in the form of ribonucleoproteins (vRNPs). The viral polymerase is a complex formed by the 3 subunits PB1,

PB2 and PA, associated to each vRNP. The matrix 1 (M1) protein covers the interior of the cellular-origin lipidic envelop and the matrix 2 (M2) protein is a pH dependent proton-pump inserted in the viral envelope. The hemagglutinin (HA) and the neuraminidase (NA) are the viral surface glycoproteins, anchored in the viral envelope. They allow for viral entry into cells and efficient spreading of non-aggregated particles, respectively.

Finally, the genome encodes additional proteins such as NS1, which plays a crucial role in inhibition of host syntheses and is a major antagonist of host antiviral defenses, and NS2/NEP, the nuclear export protein which allows export of newly synthesized vRNPs from the nucleus to the cytoplasm, as well as accessory proteins such as PB2-S1, PB1-S2, PA-X that contribute to shut-off of host syntheses and inhibition of host antiviral defenses (**Fig. 2**).



*Figure 2. Organization of the genome of influenza A viruses*

The viral genome comprises 8 negative-stranded RNA segments. Each segment is transcribed in a mRNA which codes for the major proteins. These segments also allow for the synthesis of accessory proteins. Accessory proteins are generated via different mechanisms such as alternative splicing (segments 1, 7 and 8), alternative initiations of translation (segments 2 and 3) and ribosomal frameshift (segment 3).

### 1.3 Influenza A viruses: the viral multiplication cycle

The IAVs multiplication cycle starts with binding of the HA to sialic acid receptors at the surface of target cells (**Fig. 3**). This causes endocytosis of the viral particle. The lumen of the endosome acidifies as the endosome matures and acidic pH triggers the activation of the M2 proton-pump, which acidifies the core and thus facilitates dissociation between M1 and vRNPs. Acidification also triggers a drastic conformational change of the HA that engenders the fusion of viral and endosomal membranes. It results in the release of viral RNPs into the cytoplasm<sup>1</sup> (**Fig. 3**). RNPs are transported to the nucleus, where transcription and replication of the genome is performed by the viral polymerase. Transcription occurs through a cap-snatching mechanism and capped and polyadenylated viral mRNAs are exported to the cytoplasm, and translated into proteins of which M2, HA and NA are directed toward the plasma membrane while PB1, PB2, PA and NP are re-imported into the nucleus to increase the rate of transcription and replication. Replication occurs through the synthesis of a complementary positive-sense RNA (cRNA) which is assembled in the form of a complementary positive-sense ribonucleoprotein (cRNP) and serves as basis for the synthesis of new viral RNAs assembled as vRNPs (**Fig. 3**). Meanwhile, NS1 attenuates antiviral responses, M1 is imported into the nucleus and binds viral RNPs to enable their export to the cytoplasm upon the binding of NEP<sup>20</sup>. The vRNPs then associate to form groups corresponding to the eight viral segments. These bundles migrate toward the plasma membrane where budding of neo-formed virions take place, enabling the viral particles to be covered by a lipid envelope containing M2, NA and HA surface proteins (**Fig. 3**). Finally, efficient spreading of virions is promoted by the sialidase activity of the NA, that prevents neo-formed virions to aggregate and to bind infected cells.

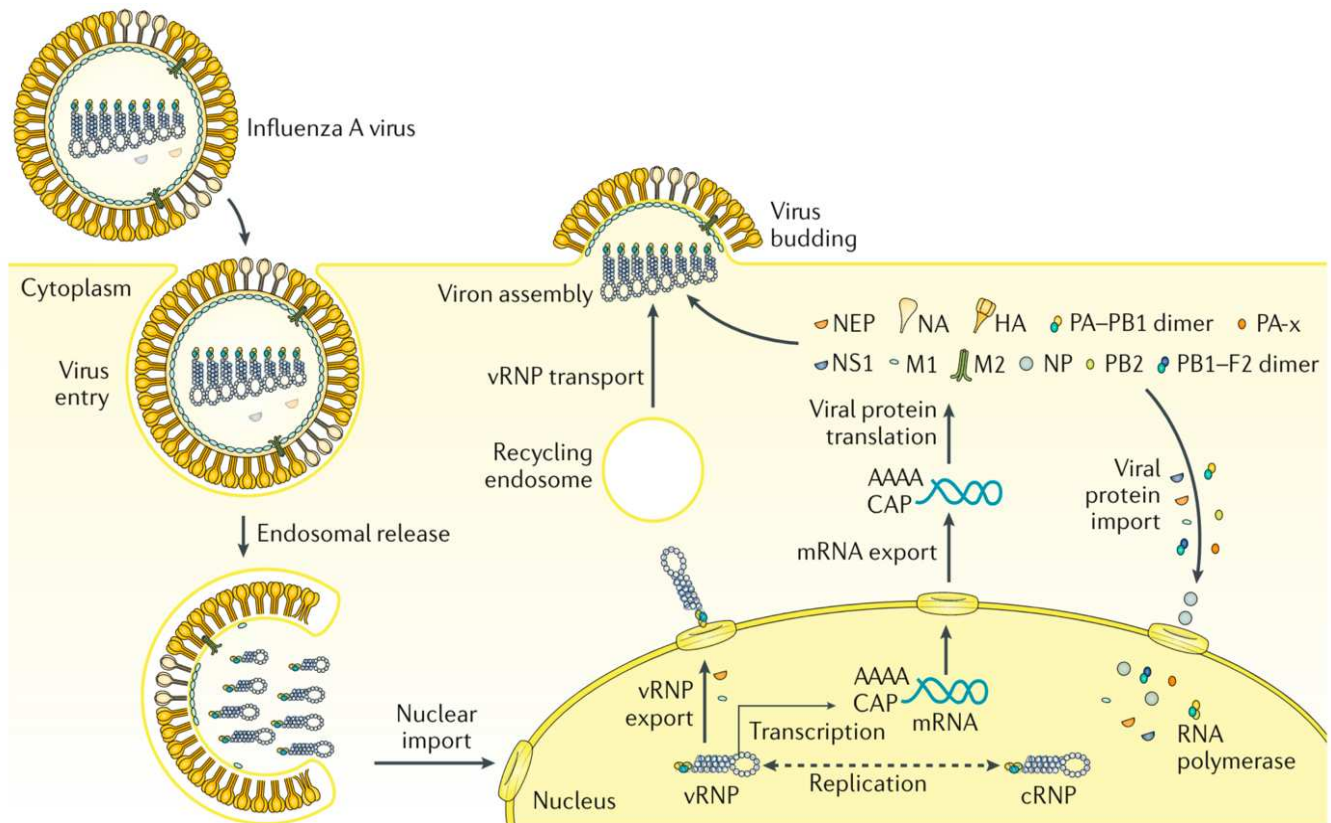


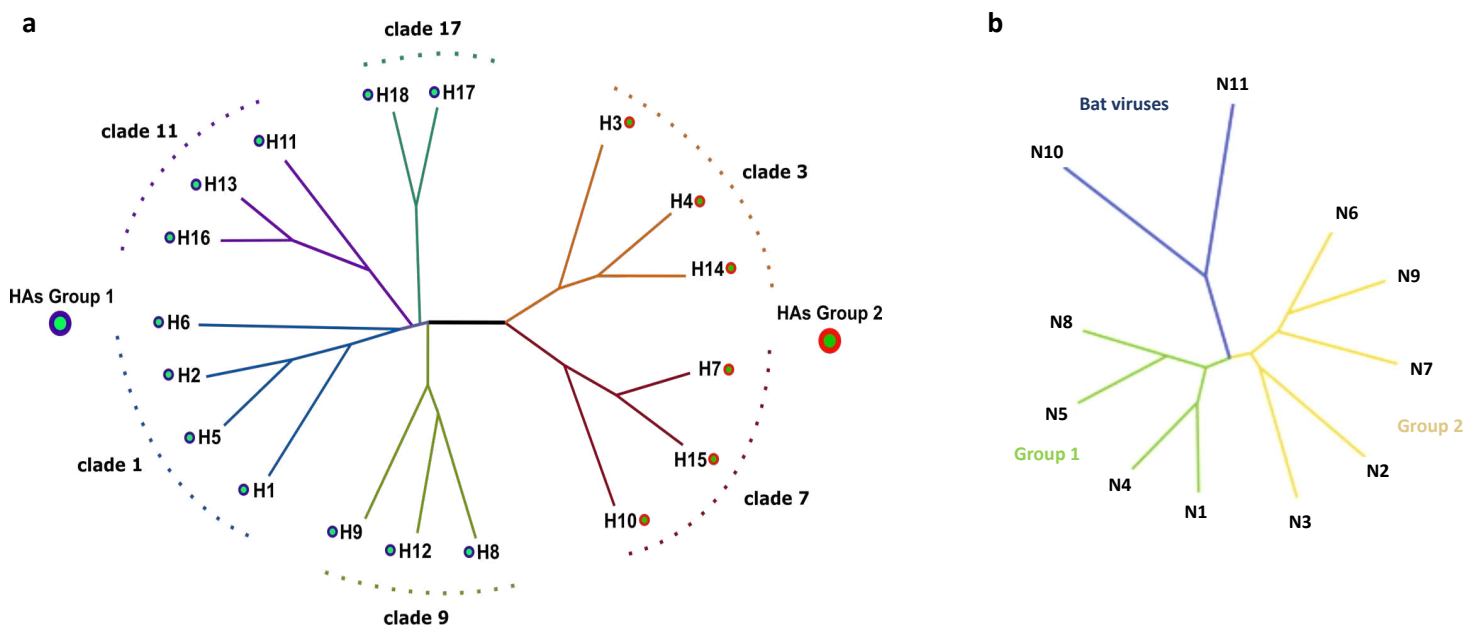
Figure 3: *Influenza A virus multiplication cycle* (From Krammer et al<sup>19</sup>)

Schematic representation of the multiplication cycle of influenza A viruses. The cycle starts by the entry of the virus within the cell through endocytosis. Upon fusion of the viral envelope with the endosomal membrane, vRNPs are released into the cytoplasm and imported in the nucleus where transcription takes place. The mRNAs are exported to the cytoplasm and translated into viral proteins. Upon import into the nucleus of the nucleoprotein and viral polymerase, replication takes place through the synthesis of cRNPs that in turn serve for the synthesis of new vRNPs that are exported to the cytoplasm. The vRNPs assemble and are transported to the budding sites where envelope proteins are inserted in the plasma-membrane. Finally, neo-synthesized infectious virions are released from the cells.

## 1.4 Influenza A virus variability and diversity

IAVs are extremely variable. The HA is the most variable protein and is used as a basis for IAV classification. Eighteen subtypes of HA (H1-H18) have been differentiated based on antigenic properties and refined via gene sequence analysis<sup>3,4</sup>. Two main groups of HAs were

identified by phylogenetic analyses<sup>5</sup>, and structural comparisons led to the description of six clades<sup>3,4</sup>. Group 1 is composed of clade 1 (H1, H2, H5, H6), clade 9 (H8, H9, H12), clade 11 (H11, H13, H16) and clade 17 (H17, H18) while group 2 is composed of clade 3 (H3, H4, H14) and clade 7 (H7, H10, H15) (**Fig. 4**). The HA amino acid sequences display about 90% identity within a given subtype, 60% between subtypes within a same phylogenetic group and 40% between subtypes that belong to different phylogenetic groups<sup>21</sup>. Moreover, the NA can be classified into three groups and eleven different subtypes (N1-N11) (**Fig. 4**). The combination of HA and NA is used to name the different IAV subtypes, i.e. H1N1, H3N2, etc.



*Figure 4: Phylogenetic tree representing HA and NA diversity*

**a.** The HA is highly variable and the resulting diversity is represented. The 2 groups are differentiated by the double circles (group 1: blue and cyan, group 2: red and green). Each group can be subdivided in clades: clade 1 (H1, H2, H5, H6), clade 9 (H8, H9, H12), clade 11 (H11, H13, H16) and clade 17 (H17, H18) for group 1 and clade 3 (H3, H4, H14) and clade 7 (H7, H10, H15) for group 2. The tree was drawn with FigTree.

**b.** Phylogenetic tree of NAs adapted from *Wu et al*<sup>22</sup> that represents the three NA groups. Group 1 comprises (N1, N4, N5, N8), group 2 comprises (N2, N3, N6, N7, N9) and group 3 comprises N10 and N11 from bat viruses.

Variability translates in a wide range of infected host species, including aquatic birds, which constitute the major reservoir, but also poultry, pigs and humans, among others<sup>23</sup> (**Fig. 5**).

While most of HA subtypes (H1-H16; N1-N9) circulate among aquatic birds, only the H1N1, H3N2 and H2N2 subtypes have been known to be well adapted to humans since the beginning of the 20<sup>th</sup> century (**Fig. 5**). The dominant circulating subtypes are limited in pigs (H1-3; N1-3). However, pigs are susceptible to infection by several other subtypes, including avian viruses. Viruses circulating in pigs thus poses a serious threat as pigs can serve as an intermediate host for avian viruses transmission to humans.

Although zoonotic IAV outbreaks in humans resulting from direct-transmission are infrequent, several avian IAV subtypes such as H5N1, H5N6, H5N8, H7N7, H7N9, H7N3, H9N2, H10N8 and have been reported to occasionally infect humans and even to cause death<sup>24-29</sup> (**Fig. 5**). Likewise, swine influenza A viruses (swIAVs) have been shown to occasionally transmit to humans. Cases of human infection with swIAVs of the H1N1, H1N2, and H3N2 subtypes have been recurrently reported<sup>30</sup> (**Fig. 5**). Importantly, reverse zoonotic events with infection of pigs by human seasonal IAVs have also been identified<sup>31</sup>.

Multiple molecular barriers prevent IAVs to infect different hosts regardless of their species<sup>32,33</sup>. Receptor specificity, polymerase adaptation, and the pH of activation of the HA are among the main viral factors to be considered in surveillance and risk assessment tools<sup>34-40</sup>. Host factors such as body temperature, pre-existing adaptive immunity or anti-viral innate immunity can also affect inter-species transmission<sup>41,42</sup>.



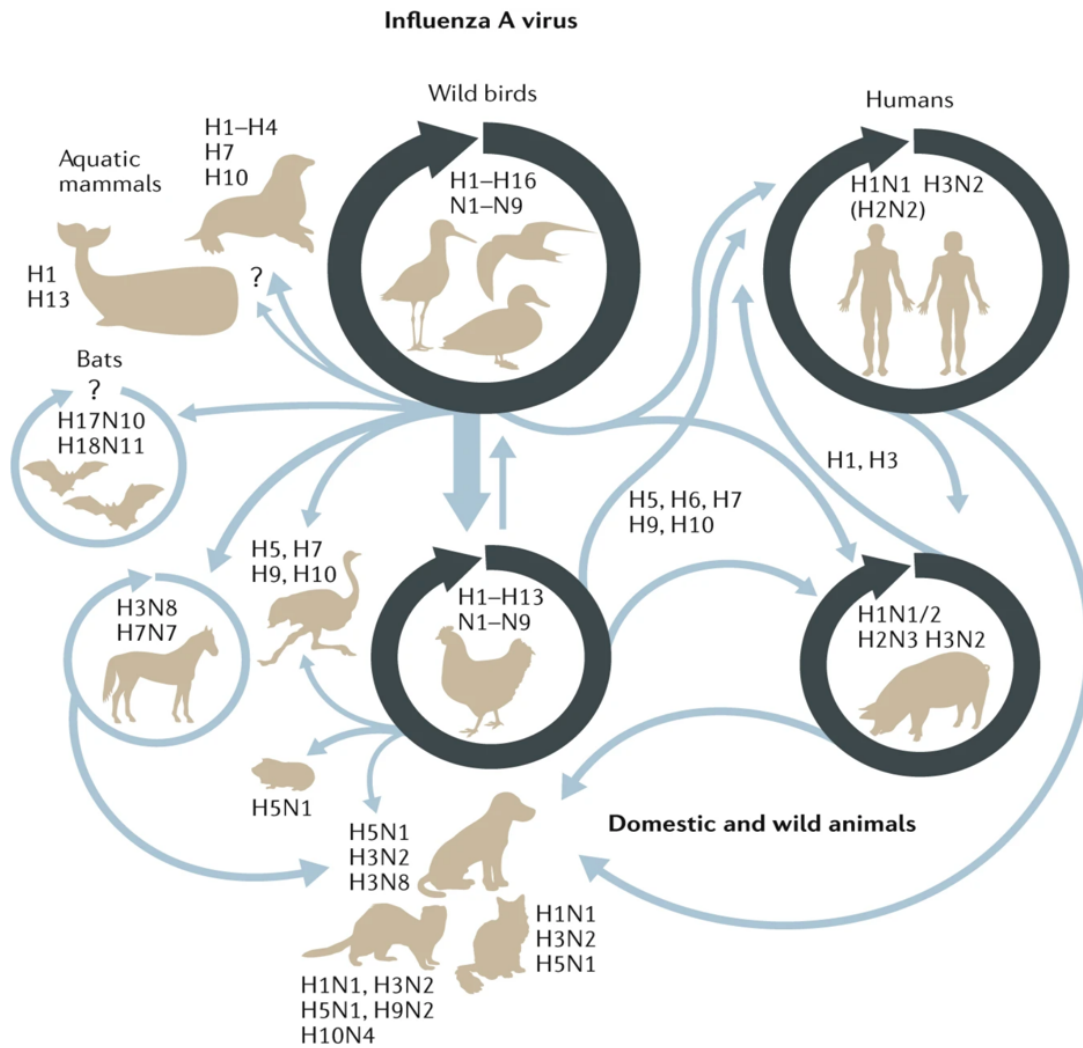


Figure 5: Influenza A virus hosts (From Long et al<sup>23</sup>)

Aquatic birds in which 16 subtypes (H1 to H16) of IAVs circulate form the major reservoir. The H17N10 and H18N11 subtypes circulate exclusively in bats. Some subtypes of IAVs also circulate and are established in poultry, pigs, and humans among others (dark blue circles). Various subtypes are able to cross species-barrier and infect various other species (light blue arrows) in which they eventually establish. Zoonotic events can involve an intermediate host and/or require adaptive mutations.

The viral variability of IAVs results from the RNA and segmented nature of their genome<sup>43-45</sup> (Fig. 6). It is mainly due to the high error rate of the viral RNA-dependent RNA polymerase that lacks proofreading activity, generating  $2.0 \times 10^{-6}$  genomic mutations per site per replication cycle, also referred to as genetic drift<sup>46</sup>. The two transmembrane glycoproteins

HA and NA evolve even faster than the other viral proteins because they intrinsically better tolerate and accumulate mutations, especially the HA<sup>47</sup>. Hence, the yearly recurrence of influenza epidemics is mainly due to continuous genetic variations that translate into HA and NA antigenic variations, also called antigenic drift<sup>48</sup> (**Fig. 6**).

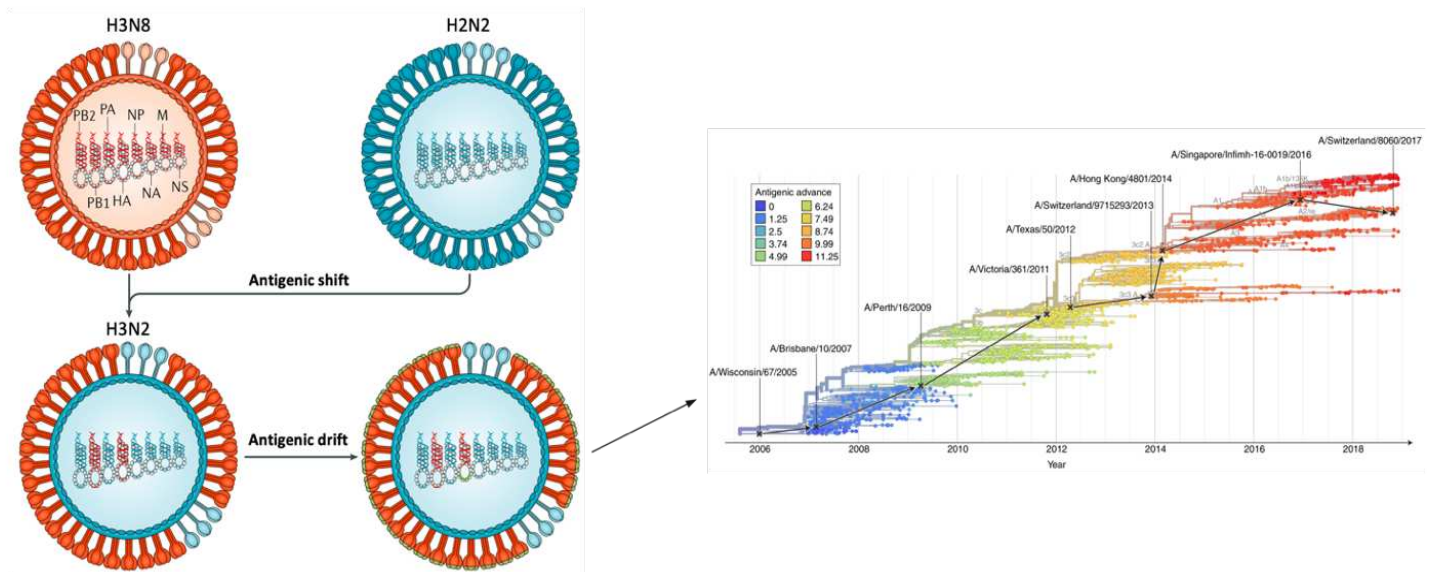


Figure 6: Influenza A virus variability (From Krammer et al<sup>19</sup> and Yamayoshi et al<sup>48</sup>)

Two mechanisms give rise to the antigenic diversity of influenza A viruses: antigenic shift and antigenic drift. Antigenic shift results from segment (coding the major antigens HA or NA) exchanges or reassortment during co-infection events by viruses from different subtypes. Antigenic drift results from the accumulation of mutations upon virus replication and immune selection. Evolution of antigenic characteristics of A(H3N2) viruses in humans over the 2006–2019 period is shown.

In addition, the segmented nature of the genome of IAVs provides an additional advantage in terms of evolutionary potential. It confers the capacity, upon co-infection of a single cell by multiple viruses, to produce reassortant viruses that harbor combinations of genomic segments derived from different parental viruses, a process named genetic reassortment or genetic shift<sup>49</sup> (**Fig. 6**). Moreover, upon reassortment between human and animal viruses,

antigenic shift can happen, generating new IAVs that might contain HA and/or HA/NA combinations to which most of the human population has no preexisting immunity and that might have the ability to spread amongst humans, therefore potentially triggering a pandemic.

Widespread circulation of IAVs with zoonotic potential enhances the probability of introduction of new IAVs in the human population, which upon adaptation and establishment of sustained human-to-human transmission, could potentially give rise to a disastrous situation across the globe<sup>50–52</sup>. In the last century, three influenza pandemics occurred<sup>14,15,53</sup> (Fig. 7).

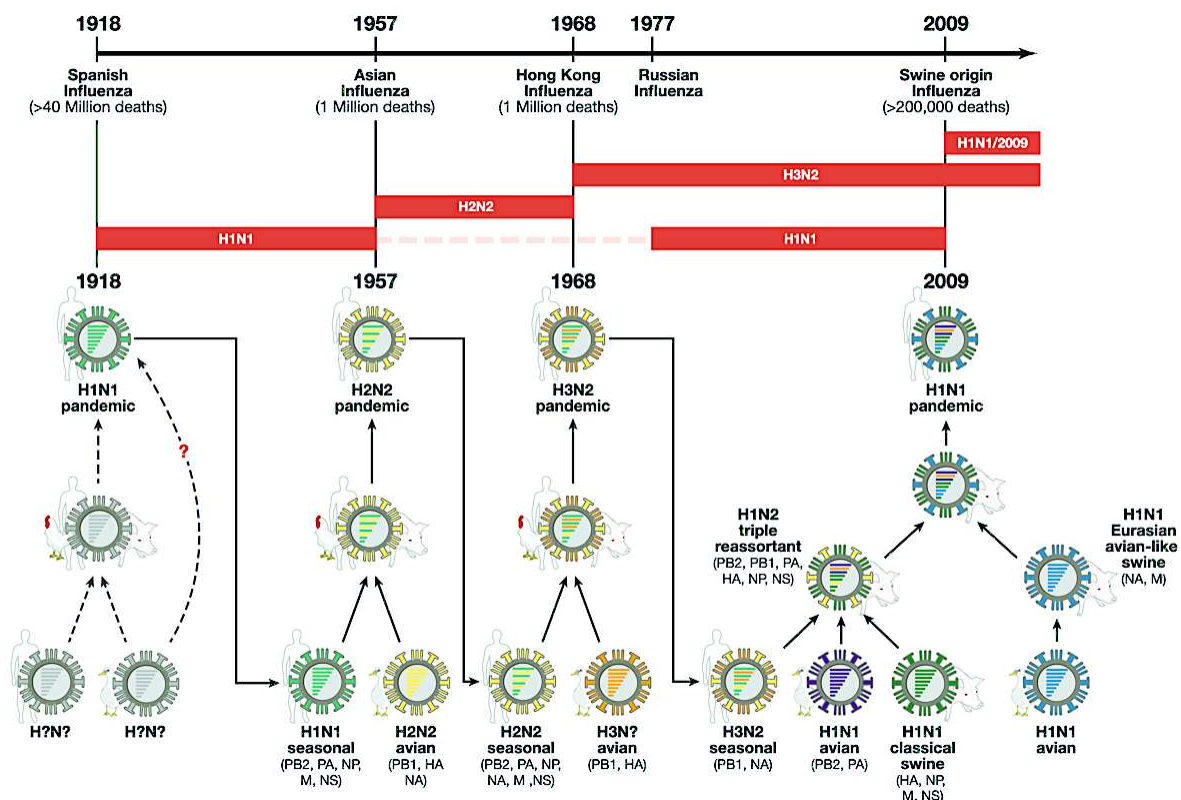


Figure 7: *Pandemics caused by influenza A viruses* (from *Weldel et al*<sup>53</sup>)

Three influenza pandemics occurred during the 20<sup>th</sup> century and, for the moment, one occurred during the 21<sup>st</sup> century. The Spanish influenza pandemic was by far the most devastating event. The timeline depicting pandemic events, the subtype of pandemic strains, the reassortment events and host of origin of the viruses involved are specified.

The 1918 so-called “Spanish flu” pandemic emerged due to a H1N1 virus<sup>54</sup>. This pandemic caused more than forty million deaths and remains as the most devastating infectious event of the 20<sup>th</sup> century<sup>54</sup>. The 1957 Asian flu and 1968 Hong-Kong flu were the result of reassortment events between human and avian viruses in pigs, that generated the outbreak of the pandemic strains of H2N2 and H3N2 subtypes, respectively<sup>14,55</sup>. In 2009, a novel H1N1 virus (H1N1pdm09), was responsible for the first flu pandemic of the 21<sup>th</sup> century<sup>56,57</sup>. The H1N1pdm09 strain resulted from several reassortments between H1N1 avian viruses, H1N1 swine viruses and H3N2 human viruses that occurred in swine over at least 30 years.

The pandemic potential of a strain depends on the ability of the virus to acquire mutations conferring efficient inter-human transmission capacity, as well as on whether the population is, or not, immunologically naïve to the virus of concern. Viruses circulating in pigs, which are considered as mixing vessels, constitute a serious pandemic threat due to the broad susceptibility of pigs to infection by several other subtypes. Nevertheless, future pandemics could potentially emerge through direct inter-species transmission, without passing through an intermediate host, and emergence of mutations conferring further host adaptation. Avian viruses have occasionally infected humans, causing severe disease with high fatality rates (53% and 38% for H5N1 and H7N9, respectively)<sup>58-61</sup>. Fortunately, although human-to-human transmission has been reported in some cases (e.g. H5N1, H7N9, H7N7), no efficient and sustained human-to-human transmission has been acquired following these events<sup>62</sup>. The same holds true for swine influenza viruses which are also associated with high morbidity when directly transferred to humans<sup>30</sup>.

However, the genetic drift of such viruses could confer adaptation to humans, making them a major concern of pandemic emergence risk<sup>63</sup>. Indeed, by infecting ferrets with avian H5N1

or H5 HA/pdm09 reassortant viruses, studies have shown that efficient airborne transmissibility between ferrets was enabled by different mutations, including mutations that conferred enhanced polymerase activity, modified the receptor binding site preferences from  $\alpha$ 2,3 linked to  $\alpha$ 2,6 linked sialic acids and decreased the pH of activation of the HA<sup>64–66</sup>.

## **| 1.5 Influenza vaccines**

The most efficient way to acquire protection against IAVs is through vaccination. It helps to limit the spread of circulating viruses and to considerably reduce the burden of illness, and so the number of hospitalizations and deaths each year. Owing to the variability of influenza viruses, and especially the antigenic drift of the HA, vaccine composition results from extensive and complex global surveillance efforts and is revised bi-annually. The seasonal quadrivalent vaccine is designed to provide protective immunity against 4 circulating strains including H1N1 and H3N2 (IAVs) as well as two viruses from each of the two lineages of influenza B viruses. Three vaccine categories are currently licensed in various countries around the globe: inactivated, live-attenuated and recombinant<sup>48</sup>.

Inactivated vaccines are the most widely used because of both high safety and low production cost associated. Viral seeds are usually propagated in embryonated hen's eggs, which may result in the selection of unwanted egg-adaptation mutations, altering the antigenicity of the vaccine. Alternatively, viral seeds are produced in qualified cell lines, and could further reduce influenza-related medical encounters compared to vaccination by egg-based vaccines<sup>67,68</sup>. The vaccine is derived from chemically inactivated virus, either in the form of a split vaccine obtained upon detergent treatment, or a sub-unit vaccine essentially

composed of the further purified envelope proteins HA and NA. It is administered intramuscularly.

Live attenuated vaccines are based on cold-adapted, temperature-sensitive and attenuated virus strains that harbor adequate HAs according to the chosen vaccine composition. The attenuated viruses are administered by the intranasal route and infect the vaccinated patient, but virus multiplication is restricted to the upper respiratory tract thus ensuring asymptomatic or pauci-symptomatic infection.

Recombinant vaccines based on the HA produced in insect cells using baculovirus expressing systems are administered by the intra-muscular route. This approach has the advantage of not being dependent on eggs supply and does not result in unwanted mutations that could be acquired during replication in eggs or cells.

Currently accepted vaccines confer poor protection to antigenically drifted strains and no protection against different subtypes. Indeed, for optimal effectiveness of influenza vaccines, they have to antigenically match the strains that are expected to circulate during the next epidemic season<sup>69</sup>. In anticipation of the potential introduction in humans of zoonotic IAVs with pandemic potential, candidate vaccine strains that match the avian or swine IAVs identified to be of highest concern are being prepared and regularly updated. Nonetheless, research efforts for the development of universal or broadly effective influenza vaccines are thus of paramount importance to tackle annually recurring epidemics but also for pandemic preparedness<sup>7</sup>. Therefore, new vaccine strategies are being developed to broaden the spectrum of vaccine efficacy. The investigated targets of these so-called universal vaccine candidates are the HA stalk domain, the HA receptor binding site, the M2 ectodomain and conserved cytotoxic T lymphocyte epitopes in M1 and NP<sup>48</sup>.

Some broadly-neutralizing antibodies bind the globular head domain of the HA, but escape mutations are likely to emerge<sup>70,71</sup>. Rare monoclonal antibodies directed against conserved epitopes in the stalk domain of the HA confer protection against several subtypes either by stabilizing the pre-fusion state and thereby inhibiting fusion of viral and endosomal membranes or by activating Fc-dependent immune responses such as antibody-dependent cell cytotoxicity (ADCC), antibody-dependent cell phagocytosis (ADCP), and complement-dependent cytotoxicity (CDC) that enable clearance of infected cells<sup>72</sup>. Hence, multiple vaccine strategies are under investigation to elicit stem-reactive antibodies with large breadth of reactivity such as immunization with headless HA<sup>73-75</sup>, sequential chimeric HAs<sup>4,76,77</sup>, peptides corresponding to parts of the stem<sup>78,79</sup> and hyperglycosylated HA<sup>80</sup>. However, it is noteworthy that mutations involved in resistance against several stem-reactive antibodies have already been reported<sup>72</sup>.

## **| 1.6 Antivirals**

Therapeutics that inhibit virus multiplication are complementary to vaccination for the control of influenza.

With respect to antivirals, approved drugs are limited to inhibitors of the M2 proton-pump (not used anymore), inhibitors of the NA sialidase enzymatic activity and inhibitors of the polymerase which target the endonuclease activity of PA<sup>81,82</sup>. No drug targeting the HA is currently globally approved, although arbidol is used in China and Russia<sup>83</sup>. Similarly to *tert*-butyl hydroquinone TBHQ, it targets the HA stem and prevents its CC at acidic pH<sup>84</sup>.

The use of antivirals as therapies against IAVs also gave rise to the emergence of resistance mutations, hence M2 inhibitors are not recommended anymore. Natural resistance to M2 inhibitors (*adamantanes*) was rather scarce before the start of the 21st century but the

emergence of *adamantane*-resistant viruses rose later due to widespread use of these inhibitors<sup>16,82</sup>. Almost 100% of circulating H3N2 strains around the world and at least 25% of H1N1 strains, including the pandemic strain from 2009, acquired resistance<sup>82</sup>. Reported cases of avian H5N1 and H7N9 strains infecting humans were also resistant to *adamantanes* independently of exposure to treatment<sup>82</sup>.

Moreover, the emergence of viruses naturally resistant to the neuraminidase inhibitor *oseltamivir* and their subsequent worldwide spread have already happened in the past<sup>16,82</sup>. Very high rates of circulating H1N1 viruses that were resistant to *oseltamivir*, were observed from 2007, before the 2009 pandemic, but the pandemic strain H1N1pdm09, which is sensitive to *oseltamivir*, replaced previously circulating H1N1 strains. However, the potential for emergence and spread of viruses naturally resistant to NA inhibitors remains a major concern.

Finally, reports of mutations (I38X) have been shown to considerably reduce the inhibitory effect of the PA endonuclease inhibitor *baloxavir-marboxil* and emergence of these mutations *in vivo* following treatment have been documented<sup>85,86</sup>.

Multiple human monoclonal antibodies have been characterized and can broadly neutralize IAVs. Most of them bind the HA stem and some are in early to mid-stage clinical trials<sup>87</sup>. Antibody-based therapies could become widely used for the treatment and prevention of IAVs in the future, nonetheless mutations conferring resistance to broadly-neutralizing antibodies directed against the HA stem have been reported<sup>72</sup>.

To tackle these issues, a purine nucleoside analogue favipiravir, which inhibits the viral RNA polymerase, has been approved in Japan under strict regulation. It is intended to be used for specific cases of concern such as a pandemic or the emergence of a viral strain resistant to the other approved antivirals<sup>16</sup>. However, an *in vitro* study has reported the emergence of a



mutation that significantly reduces the effectiveness of the antiviral and a second one that compensates for the induced impact on viral fitness, raising concerns about the potential emergence of viruses naturally resistant to this drug<sup>88</sup>. Other approaches to restrict the emergence of resistance mutations consist in combined therapies. Antiviral drugs coupled with complementary therapies, some of which target cellular components, are the subject of pre-clinical and clinical developments and could restrict the emergence of resistance mutations<sup>81</sup>. In addition, *in silico* studies based on HA sequences of the H1N1 and H3N2 subtypes led to the identification of invariance groups that could potentially be targeted<sup>89</sup>. Docking of peptides to an invariance group enriched zone suggested that such peptide could potentially inhibit the CC and limit the emergence of resistance<sup>90</sup>.

## 1.7 The hemagglutinin (HA)

The HA plays a critical role in the multiplication cycle of IAVs by enabling virus entry into cells<sup>1</sup>.

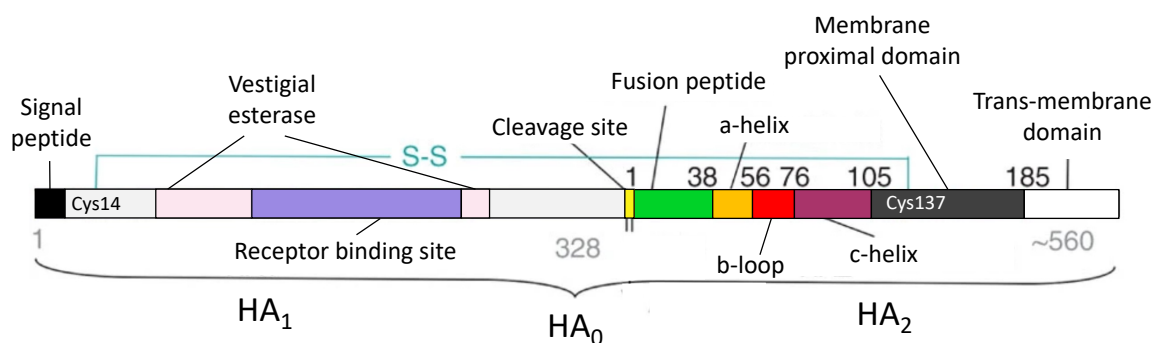


Figure 8: Organization of the hemagglutinin gene

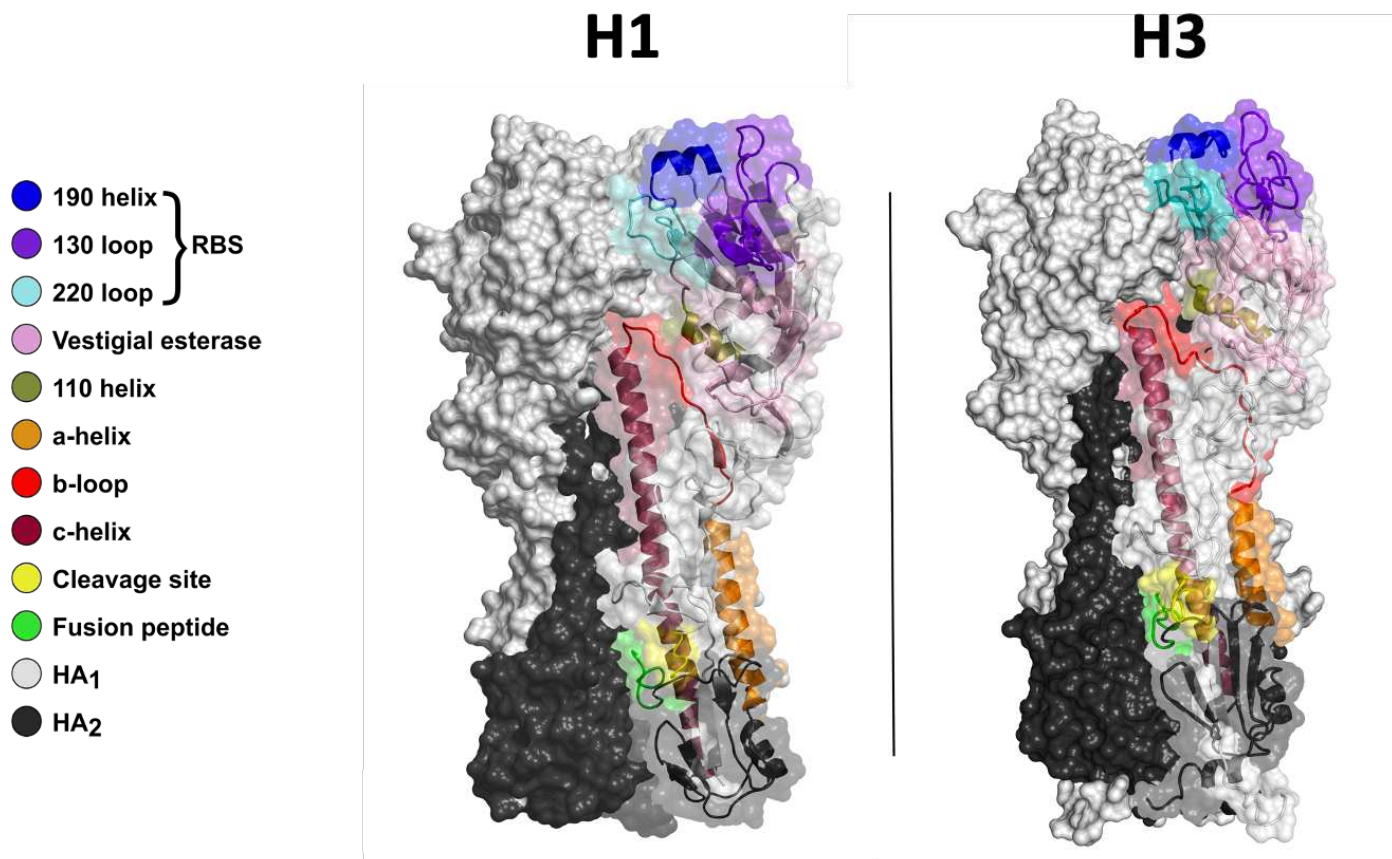
The HA is synthesized as a precursor HA<sub>0</sub>. It comprises a signal peptide at its N-terminus which ensures translocation into the ER, and a transmembrane domain at its C-terminus through which the protein is anchored. The HA<sub>0</sub> precursor is cleaved by trypsin-like proteases at a cleavage site located upstream of the hydrophobic fusion peptide. The cleaved HA encompasses the disulfide bonded HA<sub>1</sub> and HA<sub>2</sub> sub-units. HA<sub>1</sub> comprises the vestigial esterase domain and the receptor binding site while HA<sub>2</sub> comprises the fusion peptide, the  $\alpha$ -helix, the b-loop, the c-helix, the membrane-proximal and trans-membrane domains.

The HA is a type 1 transmembrane protein, characterized by the presence of a transmembrane anchor and an N-terminal cleavable signal sequence allowing insertion in the endoplasmic reticulum (ER) membrane<sup>91</sup>. The HA is synthesized as an HA<sub>0</sub> precursor which associates into a trimeric form in the ER through non-covalent inter-protomer interactions<sup>92</sup>.

Correctly assembled HA<sub>0</sub> homotrimers leave the ER and reach the extracellular side of the plasma membrane via the trans-Golgi network while misfolded HA<sub>0</sub> trimers stay in the ER prior to degradation<sup>93–95</sup>. The trimeric HA precursors (HA<sub>0</sub>) are cleaved by cellular trypsin-like proteases, resulting in a metastable complex of three protomers constituted of disulfide-bonded HA<sub>1</sub> and HA<sub>2</sub> subunits<sup>1</sup> (**Fig. 8, Fig.9**). This cleavage results in the exposure of fusion peptide at the N-terminus of HA<sub>2</sub> which is relocated into an hydrophobic cavity at the trimer interface<sup>92</sup>.

The HA of the pandemic strain *A/Hong Kong/1/1968* (H3) in its pre-fusion form was the first HA structure to be resolved by Wilson, Skehel and Wiley in 1981<sup>96,97</sup>. The description of the quaternary structure was pivotal to understand the biological activities of the HA. Since then, many other trimeric HAs have been resolved and most of the subtypes are represented (**Annex 1 – Extended data table 1**).

The HA<sub>1</sub> trimeric complex forms the so-called membrane-distal domain (or globular head domain) encompassing the vestigial esterase (VE) and the receptor binding site (RBS), which comprises the 130-loop, the 220-loop and the 190-helix<sup>92,98</sup> (**Fig. 8, Fig.9**).

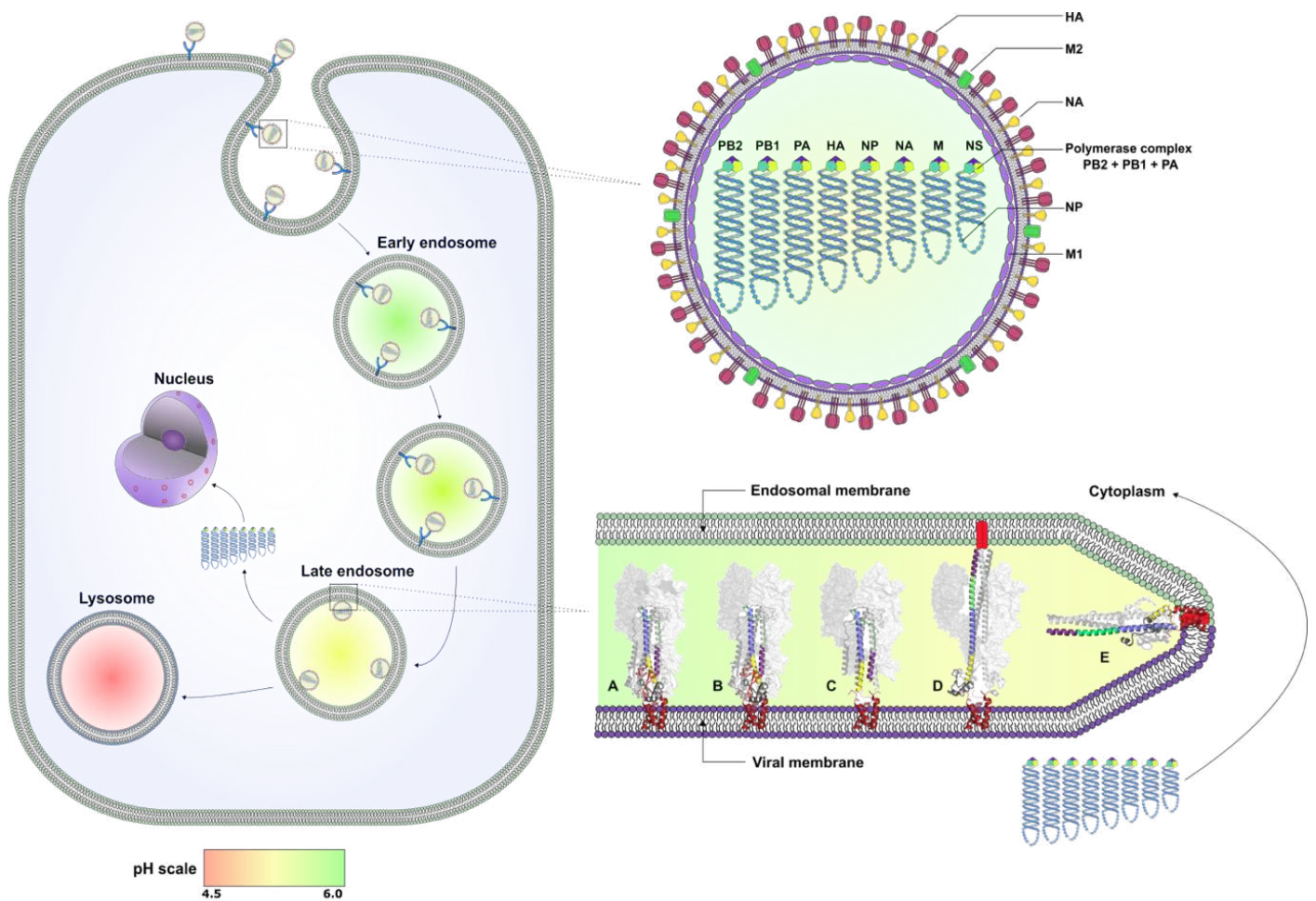


*Figure 9: Structural description of the HA (H1 and H3)*

The HA from H1 and H3 representatives (PDB accession codes: 1RU7, 4FNK) are illustrated in surface mode. The white areas represent the HA<sub>1</sub> head sub-units while the black areas represent the HA<sub>2</sub> stem sub-units. One protomer is illustrated in cartoon mode and colored as follows: blue, purple and cyan areas constitute the receptor binding site, pink represent the vestigial esterase domain, olive represents the 110-helix. The membrane-proximal domain, fusion peptide and cleavage site are represented in black, green and yellow, respectively. The a-helix, b-loop and c-helix are shown in orange, red and raspberry, respectively.

The three HA<sub>2</sub> subunits assembly constitute the stem of the HA. It includes the fusion domain that comprises the a- and c-helices which are linked by the inter-helical b-loop (**Fig. 8, Fig.9**). The three 80Å-long c-helices, form a core coiled-coil structure. The fusion peptide, enabling viral and cellular membranes fusion, is located at the N-terminus of the a-helix and is buried at the HA<sub>2</sub> interface of the trimer<sup>99</sup> (**Fig. 8, Fig.9**).

The trimeric protein binds sialic acids moieties of glycoconjugates present at the cellular surface, mostly through hydrophobic interactions and hydrogen bonds with the RBS, inducing the endocytosis of the virus (**Fig. 10**)<sup>100,101</sup>. The HA is a metastable protein. At neutral pH the trimer remains in the pre-fusion form. However, the protein undergoes a drastic and irreversible conformational change (CC) triggered by the exposure to acidic pH of the maturing endosome and eventually adopt its post-fusion form (**Fig. 10**).



*Figure 10: Viral entry process and conformational change of the HA*

Viral entry starts by the binding of the HA to its receptor located at the cell surface. It induces endocytosis of the particle. During maturation of the endosome, acidification of the lumen triggers a series of conformational changes of the HA (intermediate states A to C) leading to a drastic conformational change (state D) that enables the fusion peptide to merge with the endosomal membrane. The relocation of the fusion peptide towards the transmembrane anchor (post-fusion state E) lead to the fusion between the viral and endosomal membrane and the release of the vRNPs into the cytoplasm.

The first HA post-fusion structure, resolved after exposure to acidic pH, was that of the *A/Aichi/2/1968* (H3) strain and was described in 1994<sup>102</sup>. The determination of this structure was a milestone in understanding of how the HA induces the fusion between the viral and endosomal membrane. It illustrated the drastic CC of the protein, which relocates the fusion peptide 100 Å from its original location (Fig. 11).

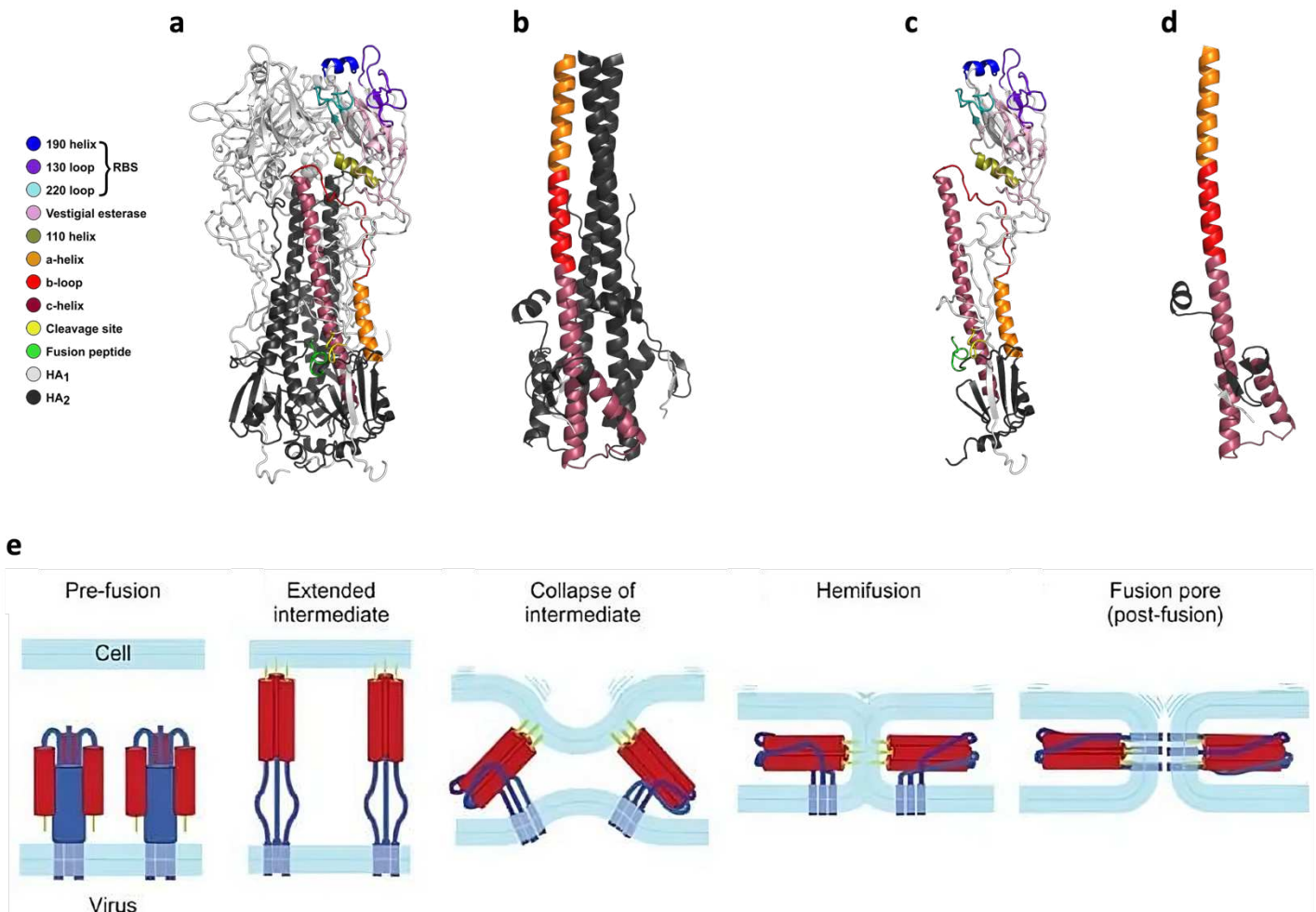
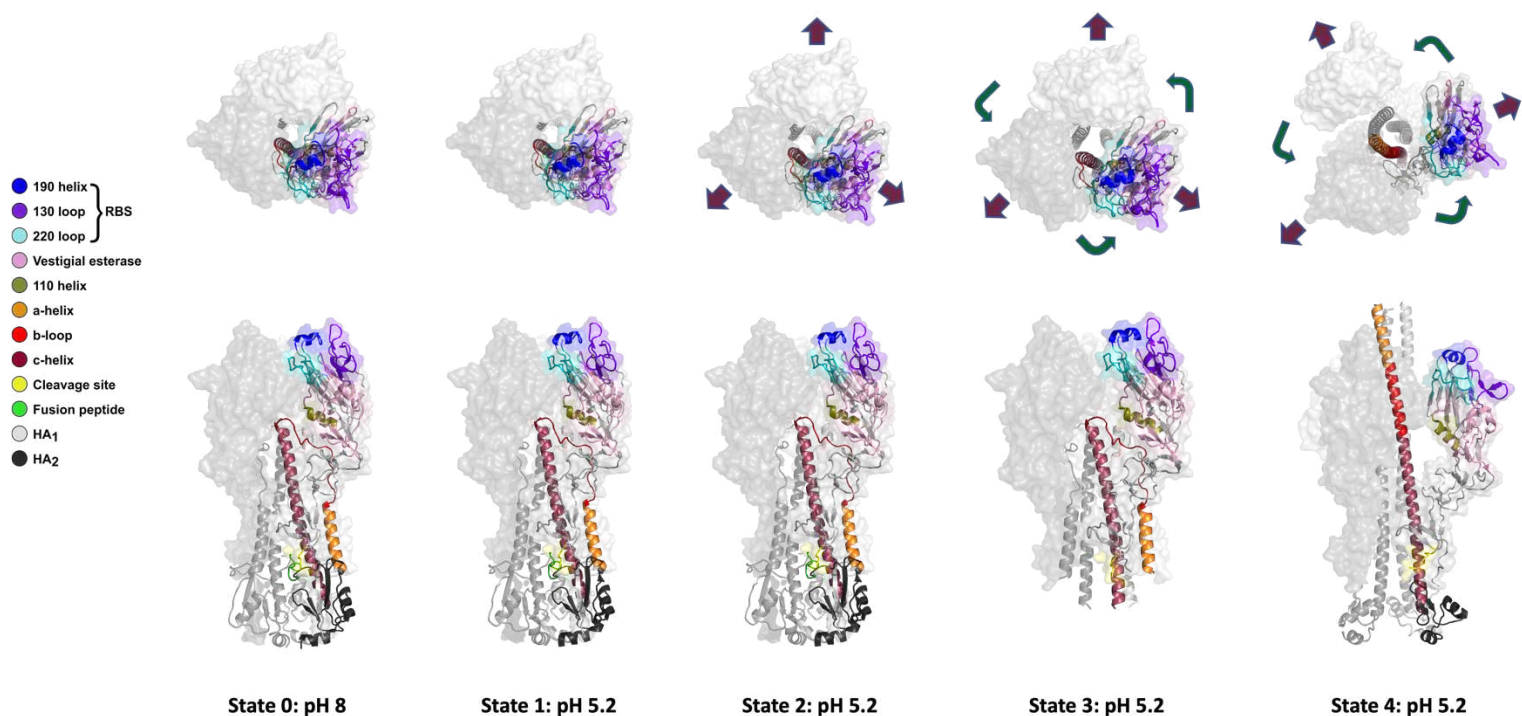


Figure 11: The fusion between the viral and endosomal membrane is triggered upon the transition of the HA from its pre-fusion state to its post-fusion state

The pre-fusion state of the HA trimer (a) and single protomer (c) as well as the post-fusion state of the HA trimer (b) and single protomer (d) are represented in cartoon mode (PDB accession codes: 1HGE and 1HTM, respectively, adapted from<sup>102,103</sup>). The conformational change mainly involves a loop-to-helix transition of the b-loop (colored in red), located in the HA<sub>2</sub> sub-unit (colored in black). This conformational change enables the fusion peptide, initially buried in its cavity, to be relocated toward the endosomal membrane. Fusion between the viral and endosomal membranes is then triggered and involves the collapse of a fusion intermediate, which allows the hemifusion of the membranes and the formation of a fusion pore (e, adapted from<sup>104</sup>).

Recently, it has been demonstrated that the cleaved HA goes through several fusion intermediate states to complete the CC (**Fig. 10, Fig. 12**). The first noticeable change of the pre-fusion structure is the slight dilation of HA<sub>1</sub> sub-units<sup>105,106</sup> (**Fig. 12**). Then, a rotation of HA<sub>1</sub> sub-units relative to HA<sub>2</sub> sub-units is observed together with further HA<sub>1</sub> dilation (**Fig. 12**). Next, a significant disorder is observed in the transmembrane-proximal area of the HA<sup>105,106</sup> (**Fig. 12**). It results in the straightening of the initially curved c-helices and their distancing, causing the release and exposure of the fusion peptide initially buried in its cavity at the HA<sub>2</sub> interface. Subsequently, the HA undergoes the drastic CC, with the loop-to-helix transition of the b-loops which leads to the extension of the initial coiled-coil formed by the three c-helices<sup>105,107</sup> (**Fig. 10, Fig. 11, Fig. 12**). This stabilized structure allows to direct the fusion peptides toward the area bound by HA<sub>1</sub> and enables their anchorage into the endosomal membrane<sup>108,109</sup>. As the last intermediate state collapses, the fusion peptide is relocated towards the transmembrane anchor and HA adopts its post-fusion conformation (**Fig. 10, Fig. 11**).

Membrane perturbations lead to the hemifusion, the formation of stalk structures and the final merge of the viral and endosomal lipid bilayers the release of the vRNPs into the cytoplasm<sup>110–113</sup> (**Fig. 11**).



*Figure 12: Structure of H3 fusion intermediates*

The different H3 fusion intermediate states known to date are represented<sup>114</sup>. State 0 is the structure at neutral pH, the pre-fusion structure. Upon exposure to acidic pH, the conformational change is triggered. In state 1, no major change is noticeable. However, in state 2 the dilation of HA<sub>1</sub> sub-units is observed. In state 3, HA<sub>1</sub> sub-units are even more diluted and rotate relative to the stem and disorder at the membrane-proximal area prevents structural resolution of this section. In state 4, HA<sub>1</sub> sub-units are even more diluted and further rotated relative to the stem. The membrane-proximal part of the c-helices are straightened and have moved away from each other. Moreover, the loop-to-helix transition of the b-loop enables the anchorage of the fusion peptide into the endosomal membrane. One protomer is colored, as depicted, according to the different areas of the protein.

Mass spectrometry, molecular dynamic and cryo-electron microscopy experiments indicated that the fusion peptide could break away prior to the globular head engaging and to the spring-loaded mechanism set off<sup>106,115–117</sup>. However, structural studies have shown that dilation of the three HA<sub>1</sub> subunits remains as the first crucial step in the HA conformational change, enabling the solvent to make contacts with initially buried AAs<sup>105–107,118,119</sup>. Nevertheless, the molecular mechanisms allowing the transition of HA to the fusogenic state are currently not fully understood<sup>37</sup>.

In addition to its main function for viral entry into cells, characteristics of the HA have other implications for the biology of IAVs. The HA is involved in host adaptation and inter-species transmission. Avian and equine viruses preferentially bind to receptors with  $\alpha$ 2,3-linked sialic acids, human viruses prefer receptors with  $\alpha$ 2,6-linked sialic acids while swine viruses have no preference between those two types of links<sup>34,120–123</sup>. Among viruses that are well adapted to distinct environments, such as human and avian viruses, differences in binding specificity thus correspond to distinct attachment capabilities that result in different targeted cells<sup>124</sup>. Mutations leading to variations in receptor binding can therefore contribute to species barrier crossing<sup>35,36,101,125</sup>.

In addition, HA<sub>0</sub> cleavability depends on the AAs located at the C-terminus of HA<sub>1</sub>, which precedes the fusion peptide sequence (**Fig. 8**). Mammalian and low pathogenic avian IAVs mostly hold, on each protomer of the HA<sub>0</sub> precursor, a monobasic cleavage site with a single arginine that can solely be cleaved by trypsin-like proteases expressed by epithelial cells, thereby locally restricting the infection<sup>126,127</sup>. By contrast, highly pathogenic avian viruses carry a multibasic cleavage site susceptible to ubiquitously expressed proteases, such as furin, thereby enabling systemic spread of those viruses<sup>128–130</sup>. Hence, the distribution of proteases in hosts and tissues also contribute to defining the tropism as well as pathogenicity of a IAV strains<sup>131</sup>.

Stability of the HA is defined as the pH at which the conformational change leading to the fusion between the viral and cellular membranes occurs, with lower pH required for more stable HAs. Stability varies depending on subtype and host of origin<sup>132</sup>. For instance, the HAs of human IAVs are more stable than HAs of avian IAVs<sup>133</sup>. The HA stability has been demonstrated to affect inter-species transmission, host specific adaptation, pathogenicity, and airborne transmission, thereby influencing pandemic potential<sup>38,134</sup>. Typically, pH values



comprised between 6.0-6.5, 5.0-5.5 and 4.6-5.0 are described for the early endosomes, late endosomes and lysosomes, respectively<sup>135</sup>. However, differences in endosomal pH are described depending on cell types<sup>136-139</sup>. Such variations of endosomal acidification impact the susceptibility of cells to be infected by IAVs<sup>137,139-144</sup>.

## **| 1.8 Inter-amino acids interactions**

Interactions between amino acids (AAs) constitute the structural basis of protein function. Different types of bonds between AAs play a role in order to ensure adequate secondary, tertiary and quaternary protein structure. Mainly stemming from the sequence composition, sidechain interactions between AAs and their location within the sequence play a fundamental role in protein folding. While two AAs, close in the primary sequence can interact and take part in loops, alpha-helices and beta-sheets formation, interactions between distant AAs in the sequence can shape or modify the overall tertiary structure of a protein. Interactions between protomers lead to the quaternary structure which can be a simple dimer or a homo/hetero oligomer.

Biologically active proteins can be a monomer or a hetero/homo-multimer and each protomer can comprise different sub-units. Therefore, specific composition and disposition of interaction networks throughout the protein are essential. Interactions can be arranged in different fashions such as intra- or inter-subunits while also being intra- or inter-protomer.

Regarding non-covalent bonds involving the side chain of AAs, the most common ones are hydrophobic interactions, sidechain-corechain hydrogen (SD-C) interactions, sidechain-sidechain hydrogen (SD-SD) interactions as well as salt-bridges, which involve hydrophobic AAs, polar AAs and charged AAs, respectively<sup>145-147</sup> (**Fig. 12**). Nonetheless, other type of

bonds like anion-anion (-/-) repulsions, cation-cation (+/+) repulsions,  $\pi$ - $\pi$ , cation- $\pi$ , anion- $\pi$  interactions can also contribute to the folding and maintenance of the structure<sup>148-150</sup>.

Hydrophobic interactions are considered to be the major force contributing to folding and stability of soluble globular proteins. It is a force resulting from nonpolar sidechains clustering, excluding water molecules in solution and driving the burial of hydrophobic AA in the core of the protein (**Fig. 12**). Hydrophobic interactions occur when hydrophobic carbons of AA side chains are located at less than 4Å from each other<sup>145,147,151</sup>.

An hydrogen bond is an attractive force between a hydrogen from the covalently bound electronegative proton donor and the electronegative proton acceptor atom holding a lone pair of electrons<sup>147</sup> (**Fig. 12**). Donor and acceptor atoms are generally located at 2.5 to 3.5Å from each other for strong hydrogen bonds<sup>151</sup>.

A ionic bond is formed when a positively charged atom is located at less than 4Å from a negatively charged one, resulting in an attractive force<sup>152</sup>. The charge can be shared between a few atoms of a protonable group due to the resonance stabilization of the charge. A combination of hydrogen bond and ionic bond occur when oppositely charged atoms of charged AAs are located at a maximum of 4Å from each other, forming a salt-bridge<sup>152</sup> (**Fig. 12**). On the contrary, the close proximity, 5Å or less, of two identical charges gives rise to a repulsive force<sup>148</sup> (**Fig. 12**).

Regarding the contribution of interactions to the free energy of folding of a protein, each of these interactions is stabilizing, except for repulsions. Each close range hydrophobic interaction contributes up to  $1.1 \pm 0.5$  kcal/mol and hydrophobic interactions make the major contribution to the free energy of folding<sup>145,147,151</sup>. Hydrogen bonds seem to contribute up to  $1.1 \pm 0.8$  kcal/mol to the free energy of folding, a considerable and favorable contribution to the structural stability<sup>147,153</sup>. Moreover, depending on the

environment of the interaction, some studies show that, in specific cases, salt-bridges do not contribute to the stability of proteins while others show that a single salt-bridge can contribute up to 5 kcal/mol to the free energy of folding<sup>146,154,155</sup>. However, although the effect of salt-bridges on structural stability remains unclear, it has been assumed that the surface salt-bridges are stabilizing and that the buried ones also remain stabilizing provided that: i) the involved side-chains are positioned in a favorable geometrical orientation and ii) amphiphilic residues, charged or not, interact with them<sup>156–160</sup>. In contrast, electrostatic repulsions are destabilizing interactions with a negative effect on the free energy of folding of around 0.5 kcal/mol per repulsion<sup>161</sup>.

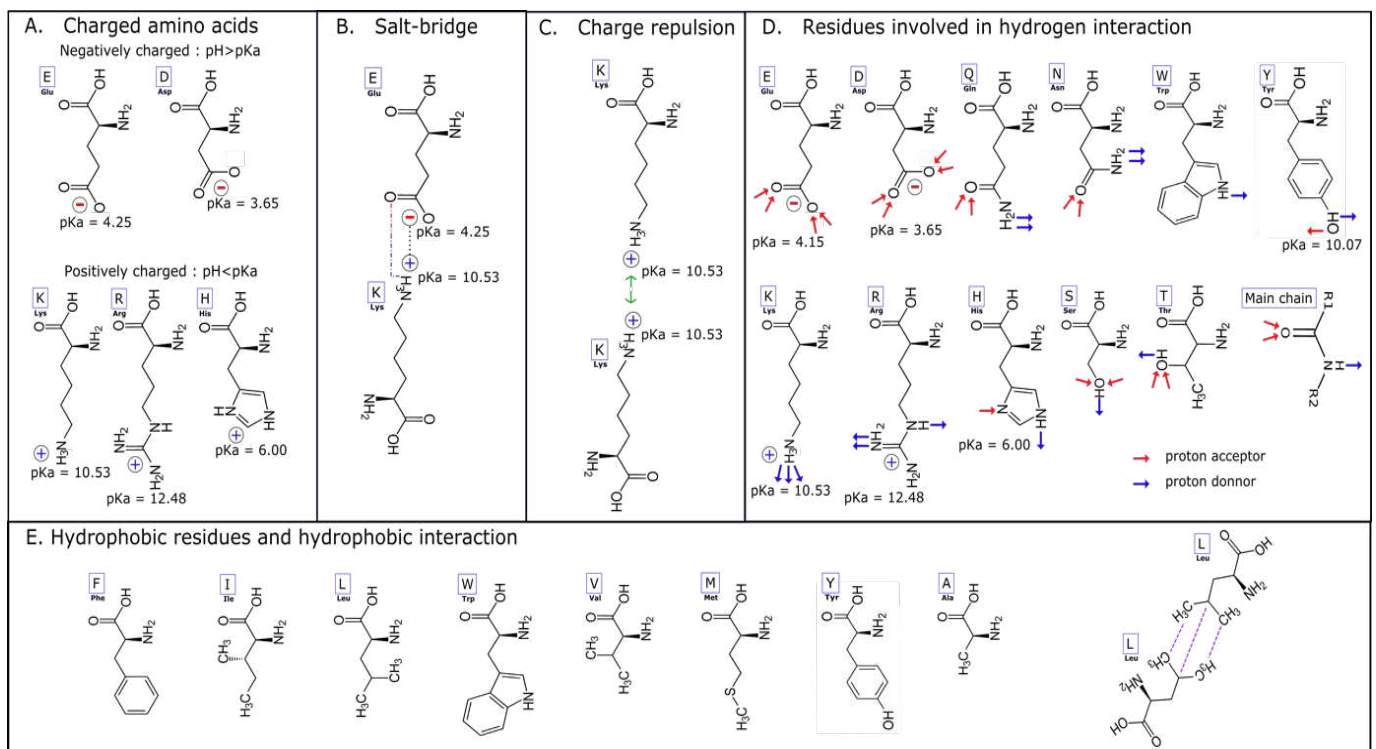


Figure 13: Residues involved in common interactions

Representation of amino acids involved in common interaction types. Charged amino acids (A) can be engaged in salt-bridges (B), which involve at least one negatively charged and one positively charged amino acids or in repulsions (C), which involve two negatively or positively charged AAs. Proximity between a proton donor and a proton acceptor (polar AAs) engenders the formation of a hydrogen interaction (D) while the proximity between hydrophobic carbons gives rise to the formation of hydrophobic interactions.

Five residues are charged at neutral pH: glutamate (Glu; pKa  $\approx$  4.1 ) and aspartate (Asp; pKa  $\approx$  3.4) are negatively charged; arginine (Arg; pKa  $\approx$  12) and lysine (Lys; pKa  $\approx$  10.7) are positively charged; histidine (His; pKa  $\approx$  6.4) is partially positively charged<sup>10,162</sup>. Histidine residues are widely accepted as biological pH sensors since their pKa is close to neutral pH. This makes His partially charged at neutral pH and positively charged at acidic pH. Moreover, Arg is always charged regardless of the environment whereas the pKa of Lys can undergo an important shift, especially when buried, and can thus, in some instances, be uncharged at neutral pH<sup>163–165</sup>. Indeed, substantial pKa fluctuations occur depending on the micro-environment<sup>10,160,162,166,167</sup>. The pKa of the negatively charged carboxyl group of Asp and Glu (pKa  $\approx$  4 and 4.5) can also vary and be shifted closer to pH values reached upon endosomal maturation<sup>135</sup>. These become protonated and lose their negative charge if  $\text{pH} < \text{pKa}_{\text{charged group}}$ . Hence, as the pH drops, for instance in maturing endosomes, salt-bridges and +/- repulsions of accessible proteins can be disrupted due to protonation of negatively charged carboxyl groups<sup>10,11,135,148,168</sup>.



# Chapter 2 Objectives

Since the first HA structure was resolved in 1981, many other pre-fusion HA proteins from different subtypes have been resolved in their monomeric or trimeric forms and gathered in the Protein Data Bank (PDB)<sup>96,97,169</sup>. A few intermediate and post-fusion structures have been documented for the H3 subtype<sup>102,105,106,170</sup>.

Protein structure and function mostly depend on physicochemical properties conferred by highly intricate networks of amino acid interactions that initiates and maintains appropriate folding and potential metastable properties<sup>2</sup>. Phylogenetic, antigenic and structural resemblance studies have been performed on the HA and serve as a basis for IAV classification, however, no comprehensive analysis of interaction networks arrangement across HA classifications has been performed. In addition, it has been suggested that the CC involves a wide distribution of molecular rearrangements in response to acidification but no extensive description of interactions behavior upon acidic pH exposure has been available to date<sup>6</sup>.

Furthermore, the variability and diversity of IAVs constitute a worldwide health threat of major concern that makes the identification of broadly effective antivirals and vaccines very challenging, but also implies the potential for emergence of variants that are resistant to therapeutics or evade immune responses<sup>7,8</sup>. Identifying broad-spectrum and non-escapable therapeutics and vaccines that would help to control annually recurring epidemics, to avoid the emergence of resistance and especially be of major importance for pandemic preparedness, is a challenge that must be addressed.

Therefore, exploiting HA variability to identify broadly conserved targets of paramount functional importance, which would not be inclined to mutate is an approach of interest. Invariant positions within a highly variable environment are globally accepted as markers of functional importance. However, invariance of a position does not necessarily imply that the

residue cannot mutate without disrupting the function<sup>171,172</sup>. Since interaction networks form the structural basis of protein function, investigating conserved interactions, rather than simply focusing on conserved residues across sequences, could provide better insights to identify non-mutable targets of functional importance.

During my thesis, I aimed at extensively examining physicochemical interactions across HAs to provide new insights on the relationships between HA evolution pathways and interaction networks composition/distribution. Moreover, I aimed at providing detailed comprehension of the molecular mechanisms driving the CC. Finally, I sought to identify conserved physicochemical interactions across HA classifications, that could serve for the development of broad therapeutic antivirals as well as universal vaccines, and to validate their biological functional importance for fusion and viral viability.



# Chapter **3.** Results



## 3.1 Results summary

The results are summarized below and a detailed description is available in Annexes 1 to 7.

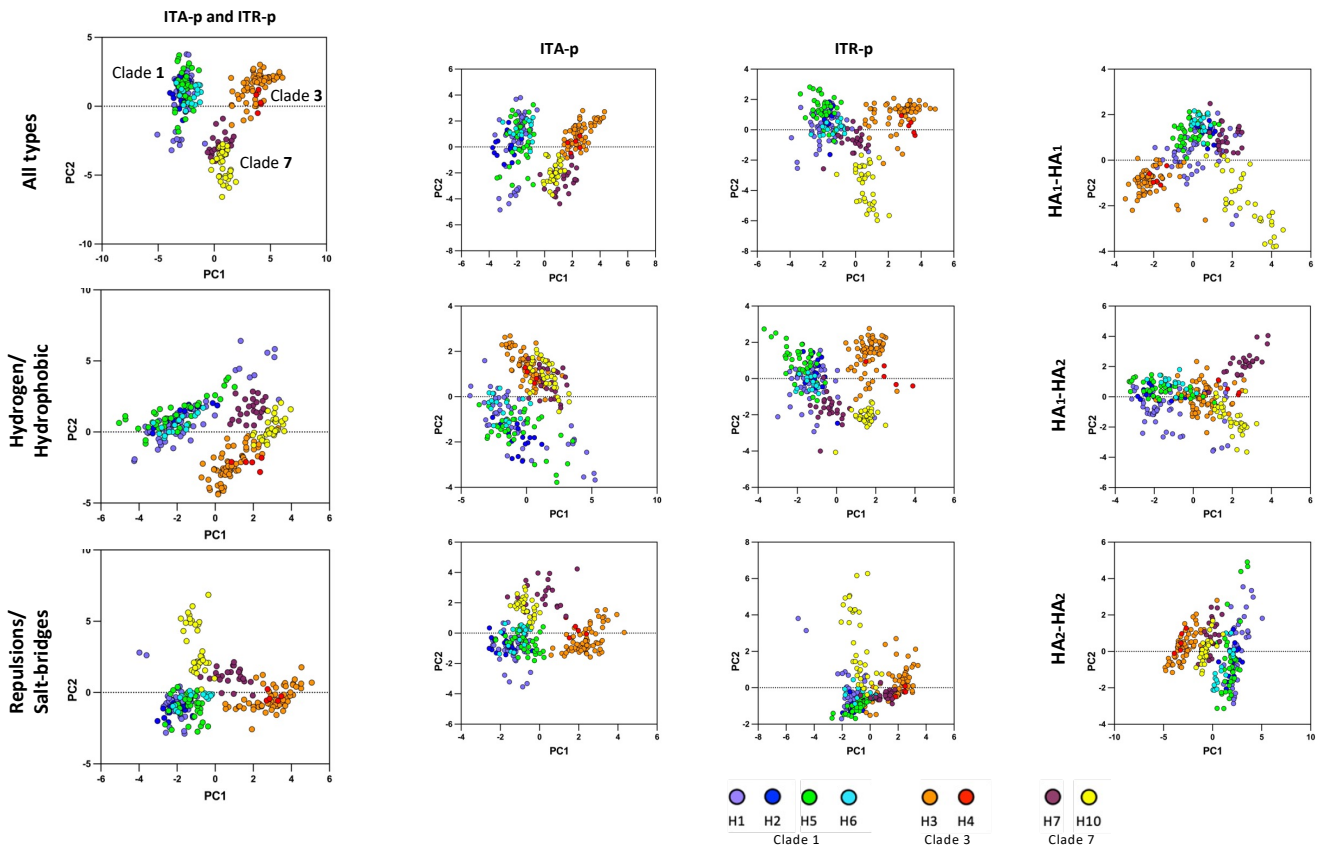
### 3.1.1 Distinct physicochemical profiles preserving HA metastable prefusion state

To generate the dataset on which our computational study is based on, we decided to consider HA trimeric structures only as they exhibit substantial information on interactions at the protomer interface and on buried interactions. The dataset (**Annex 1 - Extended Data Table 1**) encompasses the list of resolved trimeric structures, including fusion intermediates structures and post-fusion structures, mostly at less than 3.0 Å spatial resolution, along with the following information: PDB accession code, subtype, strain, host of origin, resolution method, resolution in Å and the reference. Despite variability and subsequent diversity, HA structures remain highly similar (**Annex 1 - Extended data Fig. 1b**)

For each of the 281 trimeric HA structures of our dataset, amino acid (AA) interactions at atomic scale were calculated and categorized according to AA locations and interaction types. We computed 6 interaction types (hydrophobic, SD-SD, SD-c, salt-bridges, -/- repulsions and +/+ repulsions) within and between protomers and sub-units (**Annex 1 - Fig. 1a**). Hence, twelve conditions regarding interactions distribution were considered: intra-protomer (ITA-p), inter-protomer (ITR-p) and global interactions within or between subunits (HA<sub>1</sub>-HA<sub>1</sub>, HA<sub>1</sub>-HA<sub>2</sub>, HA<sub>2</sub>-HA<sub>2</sub>) and for HA as a whole.

The composition and distribution of interaction networks were shown to be similar for HAs from group 1 while being different between HAs from group 2 (**Annex 1 - Fig. 1b, Supplementary fig. 1**). Major differences were observed between clades 1, 3, and 7, especially regarding ITR-p interactions, salt-bridges as well as -/- and +/+ repulsions, and

heatmaps representing HA subtype interactions propensity according to location and type showed patterns which appeared clade dependent (**Annex 2, Annex 1 - Extended data Fig.2, Fig.1c**).



*Figure 14: Different profiles of interaction networks enable the folding of similar structures carrying identical function*

Scatter plots representing PC1 and PC2 coefficients of standardized interaction counts from 258 structures: H1 (n=52), H2 (n=16), H3 (n=70), H4 (n=5), H5 (n=38), H6 (n=17), H7 (n=24), H10 (n=36). Parallel analysis was carried out to select the principal components. Monte Carlo simulations on random data (1000 iterations) were computed and principal components with eigenvalues greater than the 95<sup>th</sup> percentile of the eigenvalues from the simulations were retained. Principle components with the highest variance were selected for plotting. Components represent the standardized interaction number corresponding to considered interaction types and locations conditions as indicated. Unless specified, interaction types include SD-c, SD-SD, hydrophobic, salt-bridges, repulsives -/-, repulsives +/+, and locations include HA<sub>1</sub>-HA<sub>1</sub>, HA<sub>1</sub>-HA<sub>2</sub>, HA<sub>2</sub>-HA<sub>2</sub>.

To unveil correlations between interaction networks composition/distribution and HA evolution, we performed principal component analyses (PCA).

PCAs highlighted clade-dependent compositions and distributions of interactions that were shown to result in distinct overall physicochemical profiles retaining the pre-fusion structure (**Fig.14, Annex 1 - Fig. 1d**). Indeed, for all interaction types and locations altogether, we observed three significantly distinct clusters corresponding to clade 1, 3 and 7 HAs (**Fig.14, Annex 1 - Fig. 1d**). Clade 7 shared similarities with clades 1 and 3 and showed some intra-clade sub-clustering. The clade 7 cluster stood between clade 1 and clade 3 for hydrogen/hydrophobic interactions, with H7 closer to clade 1 and H10 closer to clade 3, as well as for repulsions/salt-bridges, with H7 closer to clade 3 and H10 closer to clade 1 (**Fig.14, Annex 1 - Fig. 1d**).

It therefore revealed that the intrinsic variability influences interaction networks composition and distribution, with clade 1 clearly distinct from clade 3 while clade 7 shared characteristics with both. These distinct physicochemical profiles of interaction networks enable folding of highly similar trimeric structures (**Annex 1 - Extended data Fig.1b**) but also preserve the metastable properties and meet the requirements for the CC process. Therefore, possibilities of mutations should be constrained according to the underlying physicochemical profile, and CC molecular drivers could differ.

### 3.1.2 Correlation between interactions balance and HA subtypes, host-origin and modulation of HA stability

When examining interaction networks according to the hosts of origin, PCAs showed that avian and human viruses were represented in each of the described HA physicochemical

profiles (**Annex 1 - Extended data Fig.3**). The overall HA physicochemical strategy is thus not restrictive for species-barrier crossing. Nonetheless, precise balance between destabilization mediated by repulsions and stabilization brought by ITA-p and ITR-p seems to be involved in regulating HA stability (**Annex 3**).

### 3.1.3 Characterization of highly conserved interactions on pre-fusion structures

Based on the universal numbering that we implemented (**Methods**), conserved interactions were computed. Despite intrinsic variability and protein dynamics in solvent, conserved interactions were successfully identified within subtypes but also more broadly within clades, groups and globally (**Annex 1 - Extended data Table 2**). While no ITR-p HA<sub>1</sub>-HA<sub>1</sub> interaction was conserved, disparities regarding conservation percentages were observed between clades for ITR-p HA<sub>1</sub>-HA<sub>2</sub> and ITR-p HA<sub>2</sub>-HA<sub>2</sub> specific interaction types (**Annex 1 - Extended data Table 2**). Moreover, conserved repulsions were only found in clade 7 and more significantly clade 3, which emphasizes their importance within these clades.

We computed the mean number of detected AA combinations allowing to preserve interaction type among conserved interaction and ITR-p interactions were shown to be less susceptible to conservative substitutions than ITA-p interactions (**Annex 4**).

### 3.1.4 Identification and spatial arrangement of a broadly and strictly conserved ITR-p salt-bridges complex

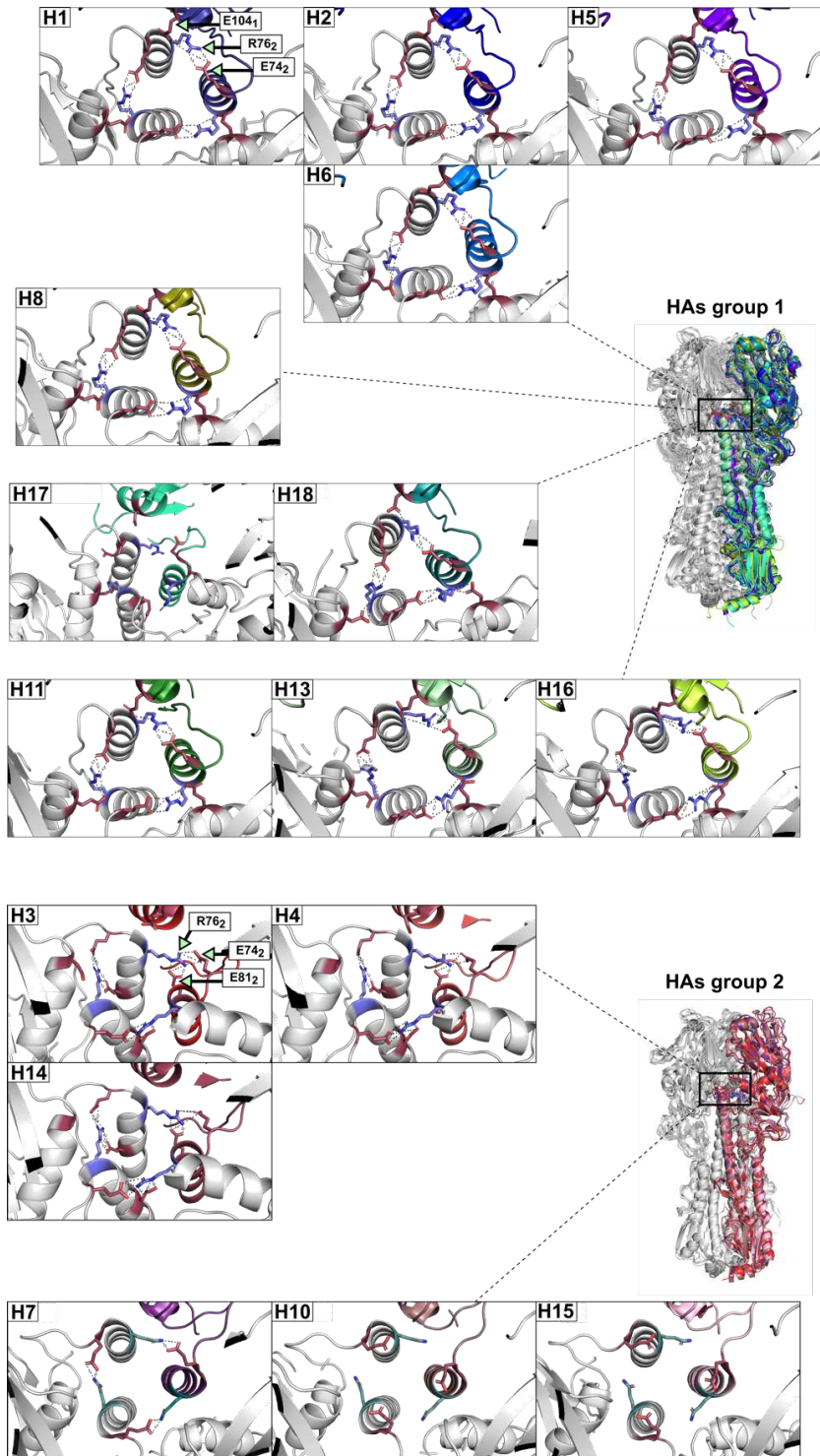
Conserved interactions supporting multiple functions are expected to be of crucial functional significance and involved AAs are thus expected to have low mutational tolerance.

Therefore, we focused on conserved inter-protomer (ITR-p) salt-bridges interactions due to their: (i) potential influence in folding and maintenance of the quaternary structure of HA trimers; (ii) pH sensor properties enabling them to respond to acidic pH exposure; (iii) role in acting as a gatekeeper against partial unfolding and in restricting protein motion<sup>9-12</sup>.

We identified a complex, in the vicinity of the b-loop, consisting of two connected inter-protomer (ITR-p) salt-bridges that we will refer to as “the SB complex” (**Fig.15, Annex 1 - Fig. 2a**). The first salt-bridge is the ITR-p HA<sub>2</sub>-HA<sub>2</sub> (E74<sub>2</sub>-R76<sub>2</sub>) salt-bridge, strictly conserved among all HAs except clade 7 HAs and H17 (**Annex 1 - Fig. 2b**). The second salt-bridge is the ITR-p HA<sub>1</sub>-HA<sub>2</sub> (E104<sub>1</sub>-R76<sub>2</sub>) salt-bridge, conserved among almost every subtype of group 1 HAs, whereas the ITR-p HA<sub>2</sub>-HA<sub>2</sub> (R76<sub>2</sub>-E81<sub>2</sub>) salt-bridge is strictly conserved among clade 3 HAs (**Annex 1 - Fig. 2b**). Accordingly, AAs involved were strictly conserved within sequences of circulating viruses (**Annex 1 - Fig. 2b, Supplementary table 1**).

Moreover, for clade 1, clade 9 and clade 11 HAs, the geometrical arrangement of the complex in space was strictly identical with optimized contacts between opposite ionizable groups of atoms (**Fig.15, Annex 1 - Fig. 2a**). The charge-conservative E104<sub>1</sub>D substitution in clade 17 HAs highlighted the interdependence of the salt-bridges involved in the SB complex. Indeed, for H18, the ITR-p HA<sub>1</sub>-HA<sub>2</sub> (D104<sub>1</sub>-R76<sub>2</sub>) salt-bridge was preserved but the optimal spatial arrangement of the ITR-p HA<sub>2</sub>-HA<sub>2</sub> (E74<sub>2</sub>-R76<sub>2</sub>) salt-bridge was impacted, whereas the complex was not formed in the only H17 trimeric structure resolved.

The spatial arrangement of the salt-bridges complex was identical between clade 3 HAs but slightly different compared to the other clades due to the absence of link with HA<sub>1</sub> (**Fig.15, Annex 1 - Fig. 2a**).



*Figure 15: Spatial arrangement of the ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges complex in different subtypes*

One protomer per structure is colored for clarity. Top-views of the HA<sub>2</sub> interface structure are shown for subtype representatives and grouped by HA groups and clades. When present, the ITR-p salt-bridges complex is represented in the panels of the subtypes in which it is conserved. Otherwise, side chains are nevertheless shown (clade 7 HAs and H17). Side chains of positions 104<sub>1</sub> and 74<sub>2</sub> are colored in red and 76<sub>2</sub> in blue (R) or cyan (Q) and actual bonds are shown with dashes. The positions are indicated in the H1 and H3 panels.

It is noteworthy that the SB complex was not formed in clade 7 HAs due to the presence of a glutamine (Q) in position 76<sub>2</sub> rather than an arginine (R) (**Fig.15, Annex 1 - Fig. 2a, 2b**). The side chains of the involved amino acids adopted various spatial arrangements in H7 structures, indicating the instability of the ITR-p HA<sub>2</sub>-HA<sub>2</sub> hydrogen SD-SD interactions (**Annex 1 - Extended data Fig. 4**).

Moreover, positions involved in the complex were differently exposed to solvent according to clades and the formation of the complex seemed to influence the positioning of the b-loop (**Annex 1 - Extended data Fig. 5, Annex 5**).

### 3.1.5 Functional relevance of the SB complex depends on the physicochemical profile

The functional importance of the SB complex was experimentally assessed in the context of distinct physicochemical profiles. We introduced mutations in the HA of clade 1 and clade 3 viruses: A/Puerto Rico/8/1934 (H1N1) and A/Centre/1003/2012 (H3N2). Substitutions were generated at each of the three positions involved in the SB complex: 104<sub>1</sub>, 74<sub>2</sub>, 76<sub>2</sub> for H1N1 and 74<sub>2</sub>, 76<sub>2</sub>, 81<sub>2</sub> for H3N2, and were chosen so that physicochemical properties vary according to the introduced mutation (**Annex 1 - Fig. 2c**).

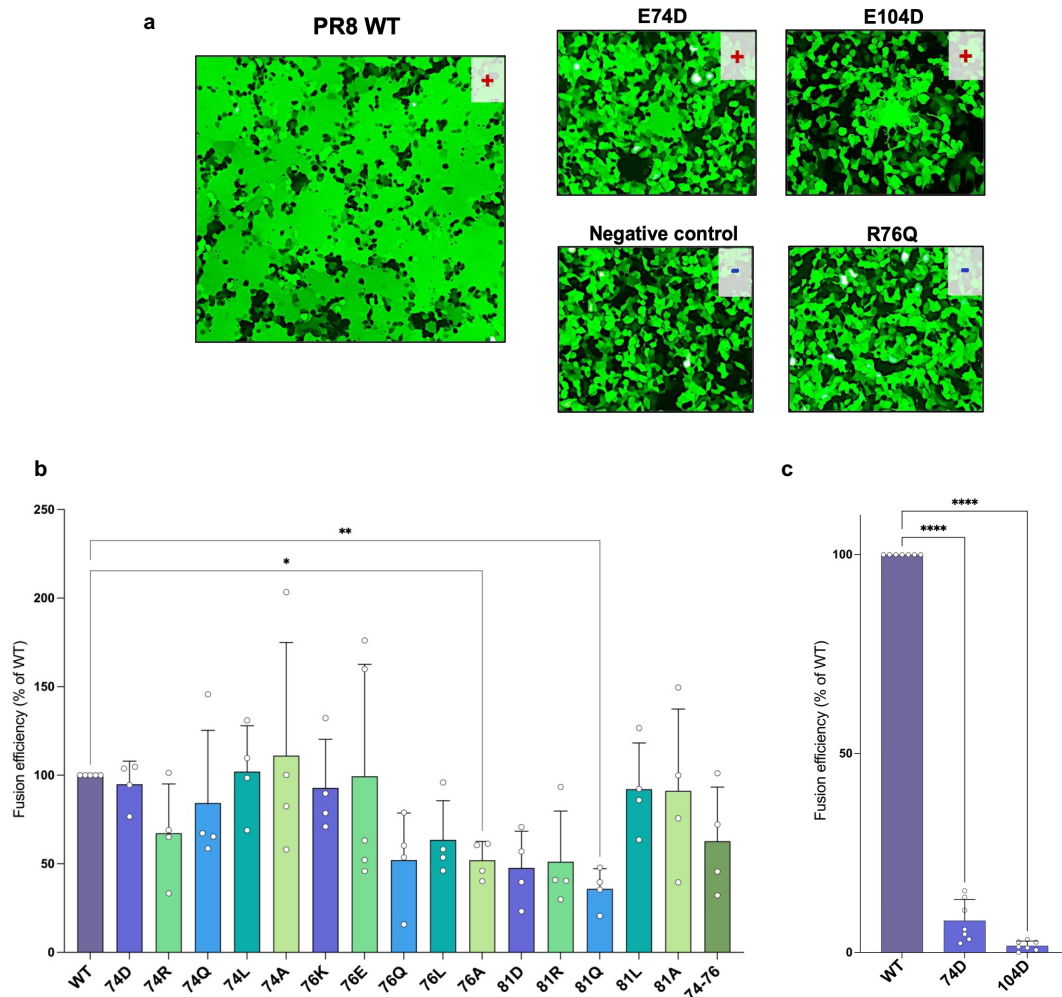


Figure 16 : Fusion potential of mutant HAs

Qualitative fusion assay of HA mutants in the H1 (PR8, A/Puerto Rico/8/1934) context (a). Transfected HEK-293T cells were treated by TPCK-trypsin and exposed to acidic pH (4.6). The ability to induce fusion was validated by the detection of formed syncytia. Photographs of the positive and negative controls are shown. The only functional mutants, 104<sub>1</sub>D and 74<sub>2</sub>D, are shown as well as R76<sub>2</sub>Q, which is one of the non-functional mutants. See Annex 1 - Figure 3 for photographs for the other non-functional H1 mutants as well as the H3 functional mutants.

Quantitative fusion assay of functional HA mutants in the H3 context (b) and in the H1 context (c). Transfected HEK-293T cells transfected with the HA mutant or WT expression plasmids and the pTM1-T7Fvo reporter plasmid were co-seeded with HEK293-T7 target cells, treated by TPCK-trypsin and exposed to acidic pH (4.6) and fusion efficiency was determined by the detection of firefly luciferase activity. Values are expressed as percentage of WT fusion activity. Data were compared to WT values using Brown-Forsythe and Welch one-way ANOVA test followed by Dunnett's multiple comparisons test (\*\*\*\* adjusted P < 0.0001, \*\*\* adjusted P < 0.001, \*\* adjusted P < 0.01, \* adjusted P < 0.05, no symbol means not significant).

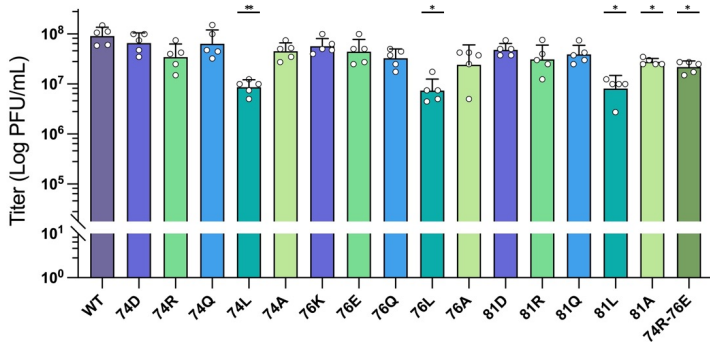


For each of the three positions studied, five different mutants were produced by site-directed mutagenesis to mimic a broad range of physicochemical effects: (i) unchanged charge, can maintain the salt-bridge; (ii) opposite charge, can form a repulsion or prevent formation of the complex; (iii) absence of charge but polar sidechain, can form a hydrogen bond; (iv) apolar and long sidechain, no interaction formed but generates steric hindrance; (v) apolar and short sidechain, no interaction formed nor steric hindrance generated (**Annex 1 - Fig 2.c**). A double-mutant (R74<sub>2</sub>-E76<sub>2</sub>) switching charges, thus preventing the formation of one salt-bridge of the complex was also produced.

The mutations generated variations in the ability of the HA to be transported at the cell surface and to be activated by proteolytic cleavage, nevertheless, none of the mutations abolished the expression of the protein nor its susceptibility to cleavage by exogenous protease (**Annex 1 - Extended data Fig. 6**).

Fusion assays indicated that the SB complex is not essential in the H3N2 context whereas it is vital in the H1N1 context for the HA to undergo the CC and to trigger the fusion process at acidic pH (**Fig. 16, Annex 1 - Fig. 3**). Indeed, all mutations resulted in comparable fusion efficiency as the WT in the H3 context (**Fig. 16, Annex 1 - Fig. 3a, 3c**). Moreover, mutations did not lead to substantial stability differences in HA stability (**Annex 6**). In contrast, any mutation disrupting the complex resulted in a loss of function in the H1 context while the E104<sub>1</sub>D and E74<sub>2</sub>D conservative mutations impaired the fusion potential (**Fig. 16, Annex 1 - Fig. 3b, 3d**).

a Reverse genetic rescue: H3



b Reverse genetic rescue: H1

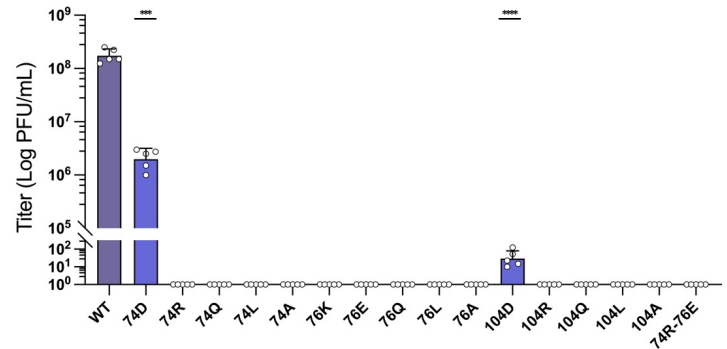


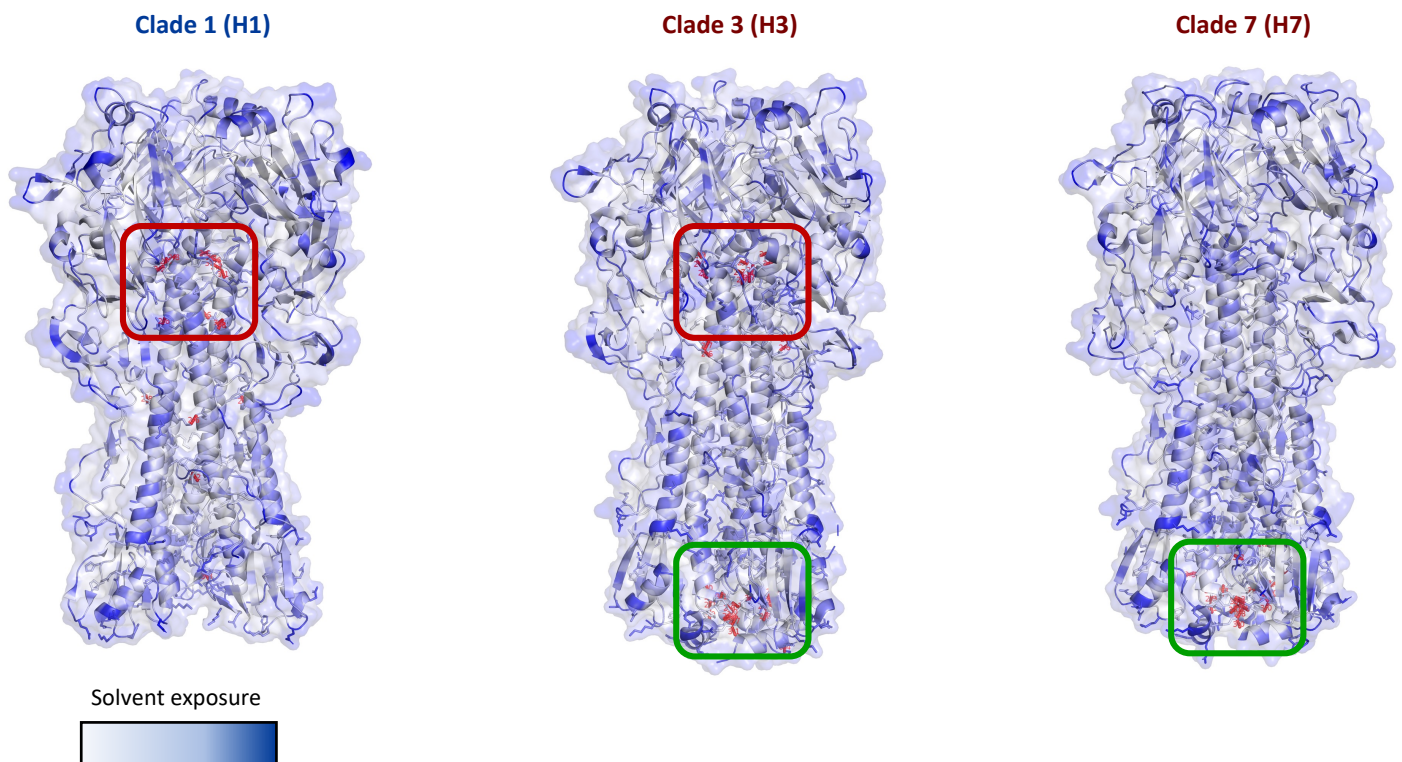
Figure 17: Viability and fitness of mutant viruses

Reverse genetics experiments in the H3 (A/Centre/1003/2012) context (a) and in the H1 (A/Puerto Rico/8/1934) context (b). Co-culture of HEK-293T and MDCK (MDCK-SIAT in the case of H3) cells were co-transfected with eight bi-directional plasmids corresponding to the eight viral segments, including WT or mutated HA sequences, and incubated for 2 days in the presence of TPCK-trypsin to allow for multiple infection cycles. Supernatants were titrated by plaque assays and titers of rescued viruses were expressed in PFU/mL. Data were compared relative to the WT with the RM one-way ANOVA, with the Geisser-Greenhouse correction followed by Holm-Sidak multiple comparisons test (\*\*\*\* adjusted  $P < 0.0001$ , \*\*\* adjusted  $P < 0.001$ , \*\* adjusted  $P < 0.01$ , \* adjusted  $P < 0.05$ , no symbol means not significant).

Furthermore, reverse genetic experiments validated these results. The broad array of mutations was tolerated in the H3 context, at least *in cellulo*, and mutant viruses were rescued at similar titers as the WT (Fig. 17, Annex 1 - Fig. 4a). On the contrary, the vital importance of the SB complex for viral viability was emphasized in the H1 context. Mutant viruses harboring non-conservative mutations could not be rescued while the E104<sub>1</sub>D and E74<sub>2</sub>D mutant viruses were rescued at reduced titers compared to the WT (Fig. 17, Annex 1 - Fig. 4b).

### 3.1.6 Clade-dependent distribution of ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges and role during the CC

To better understand the experimental disparities obtained according to H1 and H3 contexts, we represented ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges on structures of subtype representatives (**Fig. 18, Annex 1 - Extended data Fig. 7**).



*Figure 18: ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges distribution throughout the stem*

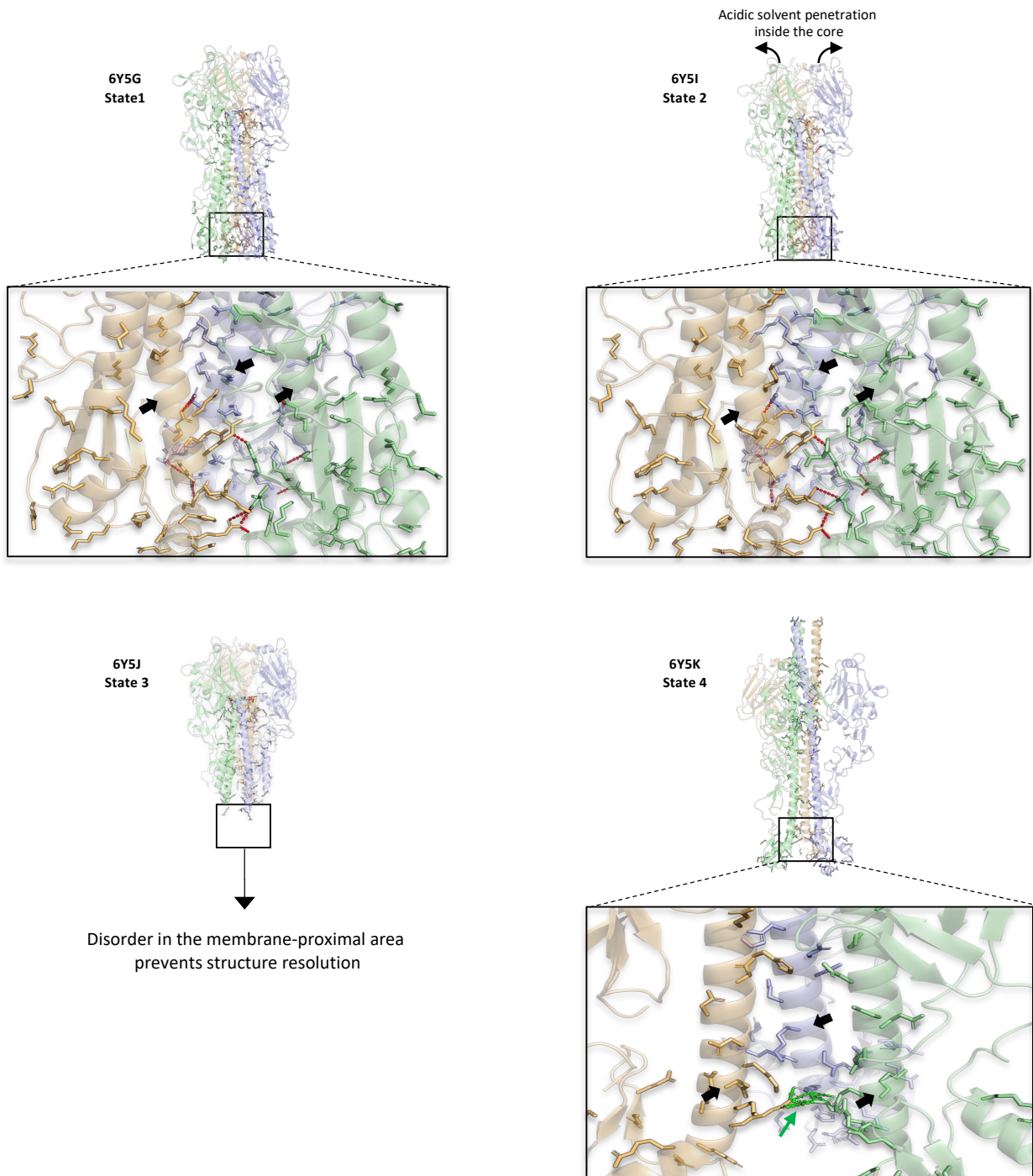
Representatives of subtypes from clades 1, 3 and 7 are represented (see Annex 1 – extended data figure 8 for the representation of all subtypes for which at least one structure is available). All formed ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges within each structure were indicated with red dashes to visualize their distribution. ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges, either located at the top of  $\alpha$ -helices or at the membrane-proximal domain, were boxed in crimson and green, respectively. PDB accession codes of represented structures are the following: H1:1RU7, H3:6MXU, H7:4NL6.

We noticed clear differences in their distribution throughout the stem between HA from different clades, especially between clades 1, 3 and 7 HAs. ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges were

located only at the top of  $\alpha$ -helices in clade 1 HAs (H1, H2, H5 and H6) (**Fig. 18, Annex 1 - Extended data Fig. 7**). This, together with the strict conservation of the SB complex in structures sharing the H1 physicochemical profile, the identical geometric arrangement of the complex in space, as well as the strict conservation of involved residues in sequences, emphasized that the essential nature of the complex in H1 are likely to be extrapolated to other HAs sharing the H1 physicochemical profile, including viruses with high pandemic potential such as H5N1 viruses.

In contrast, clade 3 HAs (H3, H4 and H14) harbored some ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges at the top of  $\alpha$ -helices but most of them were located at the membrane-proximal domain, whereas for clade 7 HAs (H7, H10 and H15), they were exclusively found at the membrane-proximal domain (**Fig. 18, Annex 1 - Extended data Fig. 7**). These membrane-proximal ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges comprise the R124<sub>2</sub>-D132<sub>2</sub> (H3 numbering; conserved in 85% and 67% of clade 3 and H7 structures, respectively), the E128<sub>2</sub>-R170<sub>2</sub> (conserved in 92% and 59% of clade 3 and H7 structures, respectively) and most importantly a highly conserved ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges complex which is constituted of the R127<sub>2</sub>-E131<sub>2</sub> and E131<sub>2</sub>-R163<sub>2</sub> salt-bridges (respectively conserved in 97% of structures and 94% of structures from group 2 HAs).

The analyses of interaction networks modulations during the CC based on the structures of H3 fusion intermediates<sup>106,114</sup> revealed the pivotal role of these conserved ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges located at the membrane-proximal area along with that of ITR-p hydrophobic interactions and repulsions as molecular drivers of the CC process (**Annex 1 - Extended data Table 3, Annex 7**).



*Figure 19: ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges disruption and ITR-p HA<sub>2</sub>-HA<sub>2</sub> +/- repulsions formation during the conformational change*

Cartoon representations of HA trimer structures of state 1, 2, 3 and 4 intermediates during the CC (DOI: [10.1038/s41586-020-2333-6](https://doi.org/10.1038/s41586-020-2333-6)). Protomers are represented in different colors. The insets show views of the membrane proximal domain where most of these interactions are found. Side chains of charged amino acids are represented as sticks. ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges are shown as red dots and ITR-p HA<sub>2</sub>-HA<sub>2</sub> +/- repulsions, only present in state 4 (green arrow), as green dots. Black arrows in insets indicate the C-helix curved in states 1 and 2 (6Y5G, 6Y5I) and straightened in state 4 (6Y5K).

The disruption of ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges and hydrophobic interactions, together with the formation of ITR-p HA<sub>2</sub>-HA<sub>2</sub> +/- repulsions, explained the straightening of the initially curved c-helices and their distancing, inducing the total extrusion of the fusion peptide initially buried in its cavity at the HA<sub>2</sub> interface (**Fig. 19, Annex 1 - Extended data Fig. 8**). Together with the analysis of the distribution of ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges across HA classifications and the experimental data on the conserved SB complex, these results emphasized the differences of molecular CC drivers according to distinct physicochemical profiles of interaction networks. The functional importance of the conserved ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges located at the membrane-proximal nonetheless remains to be experimentally evaluated.

Altogether, this work showed that intrinsic variability of the HA is reflected on inter-amino acids interaction networks and translates into distinct physicochemical profiles preserving the metastable pre-fusion state and fulfilling the requirements for the CC process. We confirmed that identical mutations of residues involved in conserved interactions in HAs with distinct physicochemical profiles have different impacts on function, as interaction networks behave differently and limit mutational tolerance and evolutionary pathways. Moreover, our findings highlighted the significance of combining interaction conservation assessments and functional experimental data for the identification of broad and non-escapable therapeutic targets and vaccine epitopes.

## 3.2 Annex 1: Functional relevance of conserved interactions in influenza A virus hemagglutinin depends on physicochemical profile (Manuscript in preparation)

### Functional relevance of conserved interactions in influenza A virus hemagglutinin depends on physicochemical profile

Valentin Ozeel<sup>1\*</sup>, Anne Vanet<sup>2</sup>, Sylvie van der Werf<sup>1,3,\*</sup>

1. Institut Pasteur, Université Paris Cité, CNRS UMR 3569, Molecular Genetics of RNA Viruses Unit, F-75724 Paris, France.
2. Institut Jacques Monod, Université Paris Cité, UMR7592, CNRS, épôle de génoinformatique, F-75013 Paris, France.
3. Institut Pasteur, Université Paris Cité, National Reference Center for Respiratory Viruses, F-75724 Paris, France.

\* Co-corresponding authors :

Institut Pasteur, Paris, France. [valentin.ozeel.pro@gmail.com](mailto:valentin.ozeel.pro@gmail.com) / [valentin.ozeel@pasteur.fr](mailto:valentin.ozeel@pasteur.fr)

Institut Pasteur, Paris, France. [sylvie.van-der-werf@pasteur.fr](mailto:sylvie.van-der-werf@pasteur.fr)

#### 3.2.1 Abstract

Influenza A viruses are of major global public health concern. The highly variable hemagglutinin (HA) undergoes a drastic conformational change (CC) at acidic pH, necessary for virus entry.

Based on 281 trimeric prefusion and fusion intermediates structures, we computed six interaction types and their distribution within and between protomers and HA<sub>1</sub>/HA<sub>2</sub> subunits. We revealed distinct physicochemical profiles of interaction networks and strictly conserved interactions across HA classifications.

The functional importance of an inter-protomer salt-bridges complex, strictly conserved in HAs from clades 1, 3, 9 and 11, was characterized in clade 1 and 3 viruses through mutagenesis and reverse genetics. Mutations were tolerated in the H3N2 context, *in vitro*, while any mutation disrupting the complex resulted in a loss of function in the H1N1 context. In addition, we revealed crucial interactions rearrangements during the CC (H3), especially the disruption of another inter-protomer salt-bridges complex highly conserved in HAs from clades 3 and 7.

The unveiled clade-dependent physicochemical profiles suggest different CC molecular drivers and apply evolutive constraints on the HA. Moreover, our study highlights the value of combining interaction conservation assessments and functional validation to provide insights for the research of broad and non-escapable therapeutic targets and vaccine epitopes.

### 3.2.2 Introduction

The hemagglutinin (HA) plays a critical role in the influenza A virus (IAV) multiplication cycle by enabling virus entry into cells<sup>1</sup>. The HA precursor (HA<sub>0</sub>) is cleaved by cellular trypsin-like proteases, resulting in a metastable complex of three protomers constituted of HA<sub>1</sub> and HA<sub>2</sub> subunits. The receptor-binding site, located on HA<sub>1</sub>, recognizes sialic acids on the cell surface, thereby inducing endocytosis of the virus. As endosome matures, acidification triggers a drastic conformational change (CC) that mainly involves the fusion domain consisting of the a- and c-helices and the inter-helical b-loop of HA<sub>2</sub><sup>1</sup> (Extended Fig. 1a). It enables the fusion peptide, initially buried at the HA<sub>2</sub> interface, to induce viral and



endosomal membranes fusion and the release of viral ribonucleoproteins into the cytoplasm<sup>1</sup>.

The HA is extremely variable and is used as a basis for IAV classification. Eighteen subtypes of HA (H1-H18) have been differentiated and further classified in groups and clades<sup>3-5</sup>. Group 1 is composed of clade 1 (H1, H2, H5, H6), clade 9 (H8, H9, H12), clade 11 (H11, H13, H16) and clade 17 (H17, H18) while group 2 is composed of clade 3 (H3, H4, H14) and clade 7 (H7, H10, H15).

Folding and maintenance of the trimeric metastable HA structure depend on a complex network of amino acids interactions, involving various interaction types<sup>148,152,173-175</sup>. Substantial pKa fluctuations occur depending on the micro-environment<sup>10,160,162,166,167</sup>. Therefore, as the pH drops in maturing endosomes, salt-bridges and anion-anion repulsions are likely disrupted due to protonation of negatively charged carboxyl groups<sup>10,11,135,148,168</sup>.

Despite extreme variability, with amino acid sequence identities of around 90% intra-subtype, 60% intra-group and 40% between subtypes from different groups, HA structures remain highly similar and support identical functions<sup>1,5,21,176</sup>. Interactions constitute the structural basis of protein function, however no data on interaction networks arrangement across HA classifications and subsequent generated evolutive constraints is available nor on their behavior during the CC. Moreover, IAVs constitute a worldwide health threat of major concern and their variability implies the need for broad-spectrum and non-escapable therapeutics and vaccines.

Our study aimed at examining physicochemical interactions across HAs to provide new insights on the relationships between evolution pathways and interaction networks composition/distribution. In addition, we sought to identify conserved physicochemical interactions and to assess their functional importance and mutational tolerance. We

computed inter-atomic interactions within and between protomers and sub-units, based on 281 available HA trimeric structures.

Overall, we uncovered distinct clade-dependent physicochemical profiles of interaction networks that preserve the metastable trimeric structure of the HA and its function. Possibilities in terms of viable mutations adapting to interaction networks are thus limited depending on profiles.

In addition, we identified highly conserved interactions within pre-fusion structures across HA classifications and focused on a broadly conserved inter-protomer salt-bridges complex. Any mutation disrupting the complex resulted in nonfunctional HA and non-viable virus in the H1 context whereas they were tolerated in the H3 context. Our experimental results confirmed that conserved interactions across HAs with distinct physicochemical profiles have different functional influence and mutational tolerance.

Furthermore, we analyzed the behavior of interaction networks during the CC based on the structures of H3 fusion intermediates<sup>106,114</sup>. We showed the pivotal role of another ITR-p salt-bridges complex, highly conserved in group 2 HAs, as molecular driver of the CC, which indicated various CC drivers in the context of distinct physicochemical profiles.

Importantly, our findings highlighted the value of combining interaction conservation assessments and functional experimental data for the identification of broad-spectrum and non-escapable therapeutic and vaccine targets.

### 3.2.3 Results

#### **Distinct physicochemical profiles preserving HA metastable prefusion state**

A dataset of 269 trimeric pre-fusion HA structures was assembled (**Annex 1 - Extended Data Table 1**). For each structure, amino acid interactions at atomic scale were computed. We

considered six interaction types: hydrophobic, hydrogen sidechain-corechain (SD-C) and sidechain-sidechain (SD-SD), salt-bridges, cation-cation (+/+) and anion-anion (-/-) repulsions (**Annex 1 - Fig. 1a**). Moreover, we sub-divided these interactions in twelve categories according to their distribution: intra-protomer (ITA-p), inter-protomer (ITR-p) and global interactions within or between subunits (HA<sub>1</sub>-HA<sub>1</sub>, HA<sub>1</sub>-HA<sub>2</sub>, HA<sub>2</sub>-HA<sub>2</sub>) and for HA as a whole. Although the dataset is, to some extent, heterogeneous regarding subtype representativeness and spatial resolution of structures, we analyzed interactions according to HA classifications. Group 1 HAs displayed similar interactions composition and distribution, even though some variations were found for specific interaction types and locations (**Annex 1 - Fig. 1b, Supplementary Fig. 1**). Within group 2, segregation was observed for total, ITA-p and more markedly for ITR-p interactions, whether considering all interactions or interaction types independently (**Annex 1 - Fig. 1b, Supplementary Fig. 1**). Heatmaps representing HA subtype interactions propensity according to location and type showed patterns which appeared clade dependent (**Annex 1 - Fig. 1c**). Indeed, major differences were observed between clades, especially regarding ITR-p interactions, salt-bridges along with -/- and +/+ repulsions (**Annex 1 - Extended Data Fig. 2**).

Furthermore, we performed principal component analyses (PCA) to unveil correlations between interaction networks composition/distribution and HA evolution. Unless stated otherwise, all locations (HA<sub>1</sub>-HA<sub>1</sub>, HA<sub>1</sub>-HA<sub>2</sub>, HA<sub>2</sub>-HA<sub>2</sub>) were considered. For all interaction types and locations altogether, we observed three significantly distinct clusters corresponding to clade 1, 3 and 7 HAs, indicating distinct overall physicochemical profiles of interaction networks (**Annex 1 - Fig. 1d**). These distinct physicochemical profiles were independent of the host of origin (**Annex 1 - Extended Data Fig. 3**).

The clade 7 cluster stood between clade 1 and clade 3 for hydrogen/hydrophobic interactions, with H7 closer to clade 1 and H10 closer to clade 3, as well as for repulsions/salt-bridges, with H7 closer to clade 3 and H10 closer to clade 1 (**Annex 1 - Fig. 1d**). The PCA of overall ITA-p interactions showed a clade 1 cluster and a second cluster comprising clades 3 and 7, with evident sub-clustering. This pattern was mainly due to ITA-p hydrogen/hydrophobic interactions as three distinct clusters were observed in the PCA of ITA-p repulsions/salt-bridges. Analyses of overall ITR-p and ITR-p hydrogen/hydrophobic interactions also revealed three clusters, with H7 particularly close to clade 1 and segregated from H10 (**Annex 1 - Fig. 1d**). In the PCA of ITR-p repulsions/salt-bridges H7 and H10 were relatively separated and most H10 were found close to clade 1 while H7 laid between clade 1 and 3 clusters. The distribution of interactions throughout the protein also differed between clades, especially for HA<sub>2</sub>-HA<sub>2</sub> interactions with clade 7 situated between clades 1 and 3 (**Annex 1 - Fig. 1d**). Regarding HA<sub>1</sub>-HA<sub>1</sub> and HA<sub>1</sub>-HA<sub>2</sub> interactions, intra-clade sub-clustering was observed, especially for clade 7, with H7 close to clade 1 for HA<sub>1</sub>-HA<sub>1</sub> and to clade 3 for HA<sub>1</sub>-HA<sub>2</sub> interactions.

Overall, computation of inter-atomic interactions within HA pre-fusion trimeric structures revealed that the intrinsic variability influences interaction networks composition and distribution, with clade 1 clearly distinct from clade 3 while clade 7 shared characteristics with both of them. These distinct physicochemical profiles of interaction networks enable folding of highly similar structures, as shown by structural alignments (**Annex 1 - Extended data Fig. 1b**), that preserve the metastable pre-fusion state and fulfill the requirements for the CC to proceed. Therefore, possibilities of mutations should be constrained depending on the underlying physicochemical profile and CC drivers could differ.

### **Characterization of conserved interactions on pre-fusion structures**

We hypothesized that key conserved interactions could be identified across HA pre-fusion structures, despite intrinsic variability and protein dynamics in solvent. Based on the universal numbering we implemented, conserved interactions were computed (**see Methods**). Conservation was identified within subtypes and more broadly within clades, groups and globally (**Annex 1 - Extended Data Table 2**). Some hydrophobic and SD-C interactions but none of the SD-SD interactions nor the electrostatic interactions were conserved globally. Calculated intra-subtype conservation values for total interactions were around 21% to 49%. While no ITR-p HA<sub>1</sub>-HA<sub>1</sub> interaction was conserved, conservation percentages varied between clades for specific ITR-p HA<sub>1</sub>-HA<sub>2</sub> and ITR-p HA<sub>2</sub>-HA<sub>2</sub> interactions (**Annex 1 - Extended Data Table 2**). Conservation of both ITR-p HA<sub>2</sub>-HA<sub>2</sub> hydrophobic and ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges was 25% or more within clades (25%, 25%; 44.7%, 64.3%; and 31.6%, 50% for clades 1, 3 and 7). However, ITR-p HA<sub>1</sub>-HA<sub>2</sub> hydrophobic conservation was solely detected within clade 3 (50%) whereas ITR-p HA<sub>1</sub>-HA<sub>2</sub> salt-bridges were exclusively conserved within clade 1 (50%). Interestingly, conserved repulsions were only found for clade 7 and more significantly for clade 3, which emphasizes their importance within these clades.

### **Identification of a conserved ITR-p salt-bridges complex**

Conserved interactions supporting multiple functions are expected to have low mutational tolerance and to be of crucial functional significance. Therefore, we focused on inter-protomer (ITR-p) salt-bridges interactions due to their: (i) potential influence in folding and maintenance of the quaternary structure of HA trimers; (ii) pH sensor properties enabling them to respond to acidic pH exposure; (iii) role in acting as a gatekeeper against partial

unfolding and in restricting protein motion<sup>9-12</sup>. Among highly conserved interactions, we identified a complex of two connected ITR-p salt-bridges that we will refer to as the SB complex. The first interaction of the complex is an ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridge involving residues E74<sub>2</sub> and R76<sub>2</sub> (subscripts indicate HA<sub>1</sub> or HA<sub>2</sub> residues, H3 numbering). This interaction is strictly conserved within available structures from clade 1 (H1, H2, H5 and H6), clade 3 (H3, H4, H14), clade 9 (H8), clade 11 (H11, H13, H16) and H18 (**Annex 1 - Fig. 2b**). The second salt-bridge of the complex differs between clade 3 HAs and other HAs. For clades 1, 9, 11 and H18, the salt-bridge is an ITR-p HA<sub>1</sub>-HA<sub>2</sub> interaction involving residues E104<sub>1</sub> (D104<sub>1</sub> for H18) and R76<sub>2</sub>. For clade 3, it is an ITR-p HA<sub>2</sub>-HA<sub>2</sub> interaction involving residues E81<sub>2</sub> and R76<sub>2</sub>. Strikingly, for clade 7 HAs, none of the salt-bridges forming the SB complex were observed.

For clades 1 and 3, the most represented clades regarding resolved trimeric structures, the residues involved in the SB complex were strictly conserved within available sequences (43578 and 37945 sequences, respectively). This indicated that involved positions do not mutate in circulating viruses and further strengthened the notion that the SB complex could be crucial to fulfill HA functions (**Annex 1 - Supplementary table 1**). Concerning less represented clade 11, clade 9 and H18, the strict conservation of involved residues within sequences (1759, 8463 and 2 sequences, respectively) suggested that the SB complex should be formed, as observed in resolved structures.

### **Spatial arrangement of the SB complex**

For clade 1, 9 and 11 HAs, the ITR-p HA<sub>2</sub>-HA<sub>2</sub> (E74<sub>2</sub>-R76<sub>2</sub>) salt-bridge linked the top of the  $\alpha$ -helices while the ITR-p HA<sub>1</sub>-HA<sub>2</sub> (E104<sub>2</sub>-R76<sub>2</sub>) salt-bridge linked the 110-helices with the top of the  $\alpha$ -helices (**Annex 1 - Extended Data video 1**). The spatial arrangement of the SB

complex was strictly identical between HAs belonging to these clades, with optimized contacts between opposite ionizable groups of atoms (**Annex 1 - Fig. 2a**). Clade 17 HAs from bat viruses harbored a conservative E104<sub>1</sub>D mutation. For H18, the ITR-p HA<sub>1</sub>-HA<sub>2</sub> (D104<sub>1</sub>-R76<sub>2</sub>) salt-bridge was preserved but the spatial arrangement of the ITR-p HA<sub>2</sub>-HA<sub>2</sub> (E74<sub>2</sub>-R76<sub>2</sub>) salt-bridge was altered. However, the complex did not form in the only H17 trimeric structure (**Annex 1 - Fig. 2a**). The E104<sub>1</sub>D mutation in bat viruses thus suggested some interdependence between the salt-bridges involved in the complex.

For clade 3 HAs, the ITR-p HA<sub>2</sub>-HA<sub>2</sub> (E74<sub>2</sub>-R76<sub>2</sub> and E81<sub>2</sub>-R76<sub>2</sub>) salt-bridges connected the top of the  $\alpha$ -helices and C-terminus of the  $\beta$ -loops (**Annex 1 - Extended Data video 2**). The spatial arrangement of the SB complex was identical between clade 3 HAs but slightly differed from that of other clades due to the absence of linkage with HA<sub>1</sub> (**Annex 1 - Fig. 2a**). Noteworthy, clade 7 is the only clade harboring the R76<sub>2</sub>Q substitution, which results in unstable ITR-p HA<sub>2</sub>-HA<sub>2</sub> (E74<sub>2</sub>-Q76<sub>2</sub>) hydrogen bonds in a minority of H7 while no interaction was formed in H10 and H15 (**Annex 1 - Fig. 2a-b, Extended Data Fig. 4**).

Solvent accessibility of the residues involved in the formation of the SB complex also differed between clades 1, 3 and 7 HAs. It was lowest for H1 and H5, higher for H3 and even higher for H7 (**Annex 1 - Extended Data Fig. 5**) and indicated that, prior to initiation of the CC, residues involved in the SB complex are differently exposed to endosomal acidic pH.

### **Functional relevance of the SB complex depends on the physicochemical profile**

To experimentally assess the functional importance of the ITR-p salt-bridges forming the SB complex in the context of distinct physicochemical profiles, we introduced mutations in the HA of clade 1 and clade 3 viruses: A/Puerto Rico/8/1934 (H1N1) and A/Centre/1003/2012 (H3N2).

Substitutions were generated at each position involved in the SB complex: 104<sub>1</sub>, 74<sub>2</sub>, 76<sub>2</sub> for H1 and 74<sub>2</sub>, 76<sub>2</sub>, 81<sub>2</sub> for H3 (**Annex 1 - Fig. 2c**). For each position, five mutants were produced by site-directed mutagenesis to explore different physicochemical properties: (i) unchanged charge, can maintain the salt-bridge; (ii) opposite charge, can form repulsion or prevent formation of the complex; (iii) no charge but polar sidechain, can form hydrogen bond; (iv) apolar and long sidechain, no interaction but generates steric hindrance; (v) apolar and short sidechain, no interaction but generates slight steric hindrance (**Annex 1 - Fig 2c**). A double-mutant (R74<sub>2</sub>-E76<sub>2</sub>) switching charges and expected to form a simple salt-bridge was also produced. Hence, a total of 16 mutants per strain were generated and validated by sequencing.

As a first step of the functional analysis, expression of the mutant proteins in HEK293T transfected cells was assessed by western-blot (**Annex 1 - Extended Data Fig. 6a-b**). For each mutant, the precursor HA<sub>0</sub> was expressed at levels comparable to wild-type (WT) HA and migrated as an equivalent molecular weight species, indicating no impact of the mutations on the level of glycosylation. Surface expression of the precursor was assessed by flow-cytometry (**Annex 1 - Extended Data Fig. 6e-f**). It was lower for H1 mutants compared to WT, with values ranging from 41% to 78% of WT surface expression. In contrast, surface expression of H3 mutants was overall similar to WT, with values ranging from 79% to 147% of WT surface expression. The ability of mutant HA<sub>0</sub> precursors expressed at the cellular surface to be cleaved and activated was assessed in the presence of exogenous TPCK-trypsin (**Annex 1 - Extended Data Fig. 6c-d**). The H1 and H3 mutants were all susceptible to cleavage as shown by the detection of HA<sub>1</sub>, albeit differences in cleavage efficiencies were observed, in agreement with reports that showed reduced cleavage efficiency of HAs mutated at positions flanking the b-loop in different subtypes<sup>132,177,178</sup>. Overall, despite some variations,



none of the mutations abolished surface expression of the protein nor its susceptibility to cleavage by exogenous protease.

As a second step of the functional analysis, we evaluated the fusogenic potential of mutant HAs using a cell-cell fusion assay. HA-expressing cells were treated by exogenous TPCK-trypsin and subsequently exposed to acidic pH (4.6). Qualitative fusion experiments showed that all H3 mutants were able to induce the formation of syncytia, and quantitative fusion assays showed no major difference of fusion efficiency although mutants that were slightly less expressed at the cell surface showed moderately reduced fusion activity (**Annex 1 - Fig. 3a, 3c**). In contrast, for H1, E104D<sub>1</sub> and E74D<sub>2</sub> were the only mutants able to trigger the fusion process, yet with substantially reduced efficiency at levels of 8% and 2% of WT, respectively (**Annex 1 - Fig. 3b, 3d**). These mutations are charge-conservative, allowing the formation of the ITR-p HA<sub>1</sub>-HA<sub>2</sub> (104<sub>1</sub>-76<sub>2</sub>) and ITR-p HA<sub>2</sub>-HA<sub>2</sub> (74<sub>2</sub>-76<sub>2</sub>) salt-bridges.

Overall, the results indicated that the SB complex is not essential for the HA to trigger the fusion process in the H3 context whereas it is of paramount importance in the H1 context.

### **Functional relevance of the SB complex for virus viability**

We next assessed the biological implications of the mutations on virus viability and infectivity *in cellulo* through reverse genetics. HAs were continuously activated by exogenous TPCK-trypsin to allow for multiple infectious cycles, supernatants were titrated by plaque assay and viable viruses were sequenced to validate the presence of the expected mutations.

Both H1N1 and H3N2 WT viruses were rescued very efficiently with titers around  $1.8 \times 10^8$  and  $1.1 \times 10^8$  PFU/mL, respectively (**Annex 1 - Fig. 4a-b**). All generated H3N2 mutants were highly infectious. They were rescued with titers similar to WT, except for the E74<sub>2</sub>L, R76<sub>2</sub>L

and E81<sub>2</sub>L mutants, which multiplied less efficiently, with lower titers around  $9 \times 10^6$  PFU/mL (**Annex 1 - Fig. 4b**). In contrast, the majority of H1N1 mutants could not be rescued except for the charge conservative E104<sub>1</sub>D and E74<sub>2</sub>D mutants, which were rescued at highly reduced titers of  $2.4 \times 10^2$  and  $2.1 \times 10^6$  PFU/mL, respectively.

To further evaluate their replication efficiency, growth kinetics were performed for the E104<sub>1</sub>D and E74<sub>2</sub>D H1N1 mutants on MDCK and A549 cells (**Annex 1 - Fig. 4c-d**). Both E104<sub>1</sub>D and E74<sub>2</sub>D mutants showed reduced multiplication efficiency in MDCK cells with titers reduced by two and one order of magnitude, respectively, at 12h, 24h, 48h and 72h post-infection compared to WT virus. Similar results were observed for E74<sub>2</sub>D on A549 cells. Surprisingly, on A549 cells, only one of the three E104<sub>1</sub>D triplicates could be amplified and its growth kinetic was substantially impaired with three order of magnitude of titers difference at 72h post-infection compared to the WT virus. Sequence analysis of this virus revealed an additional R76<sub>2</sub>G mutation, further emphasizing the salt-bridges interdependency and indicating that the E104<sub>1</sub>D mutant virus cannot multiply on A549 cells, likely because of its very low fusion efficiency and the less robust endosomal acidification of A549 cells compared to MDCK cells<sup>139</sup>.

Altogether, reverse-genetics results were in accordance with the HA fusion assays and revealed the crucial importance of the broadly conserved ITR-p SB complex in the H1 context for HA function and virus viability and its low mutational tolerance, while it was not essential in the H3 context, at least *in cellulo*, underlying differences in the functional importance of the conserved SB complex according to the distinct physicochemical profiles of interaction networks.

### **Distribution of ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges is clade-dependent**

To better understand the molecular basis of the experimental differences in the functional importance of the SB complex according to distinct physicochemical profiles, we represented ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges on structure of subtype representatives (**Annex 1 - Extended Data Fig. 7**) and noticed clear differences in their distribution along the stem between HAs from different clades, particularly clades 1, 3 and 7 HAs. The vast majority of ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges were these involved in the SB complex for clade 1, 8 and 11 HAs while clade 17 HAs displayed some at the membrane-proximal domain. In contrast, clade 3 HAs harbored ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges at the top of the  $\alpha$ -helices but most of them were located at the membrane-proximal domain, whereas for H7, they were exclusively found at the membrane-proximal domain (**Annex 1 - Extended Data Fig. 7**). These membrane-proximal ITR-p salt-bridges comprise, among others, the highly conserved ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges complex which is constituted of the connected R127<sub>2</sub>-E131<sub>2</sub> and E131<sub>2</sub>-R163<sub>2</sub> salt-bridges, respectively conserved in 97% and 94% of structures from group 2 HAs.

It has been suggested that the CC involves extensive molecular rearrangements in response to acidification<sup>6</sup>. However, no comprehensive description of interactions behavior upon acidic pH exposure has been available to date. By tracking inter-atomic interactions throughout the two available sets of intermediary states structures of the CC (H3)<sup>106,114</sup>, we detected substantial disruption of the ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges located at the membrane-proximal domain, which gave rise to subsequent formation of ITR-p HA<sub>2</sub>-HA<sub>2</sub> +/- repulsions at the same location (**Annex 1 - Extended Data Fig. 8, Extended Table 3**). Together, they were at least involved in the straightening of the  $\alpha$ -helices, their distancing and release of the fusion peptide initially buried at the HA<sub>2</sub> interface (**Annex 1 - Extended Data Fig. 8**).

It emphasized that the disruption of the highly conserved membrane-proximal domain ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges, likely caused by the acidification of the core, thus act as a molecular driver of the CC in the clade 3 and 7 contexts.

### 3.2.4 Discussion

Evolution of IAVs generated tremendous diversity which led to the existing HA classifications. However, no extensive description of physicochemical interaction networks, which constitute the structural basis of protein function, has been available to date nor of their modulation during the CC. We performed a comprehensive analysis of HA physicochemical interactions, based on 269 trimeric pre-fusion structures, detailing the distribution, within and between protomers and HA<sub>1</sub>/HA<sub>2</sub> sub-units, of six interaction types at atomic scale. Substantial differences were identified between clades regarding interaction networks composition and distribution, especially regarding inter-protomer interactions, hydrophobic, salt-bridges and repulsions. Using principal component analyses, we uncovered distinct physicochemical profiles of interactions networks, with clade 1 and clade 3 HAs displaying different interaction networks, and clade 7 HAs exhibiting intermediate features with slight intra-clade segregation. Nevertheless, the unveiled distinct physicochemical profiles are compatible with the folding of highly similar trimeric metastable structures carrying identical function, therefore limiting possibilities in terms of mutation emergence fitting into the networks and suggesting different CC drivers. Possibilities become even more restricted if a selection pressure, such as immune response or treatment with antivirals, is applied. The underlying physicochemical profile of interaction networks applies evolutive constraints, which may allow enhanced prediction of the effect of mutations and better understanding and tracking of the HA evolution.

In the context of high pandemic risks and emergence of resistance against antivirals and immune response conferred through vaccination, we need solutions to efficiently tackle viral variability. Conservation is most relevant when observed in highly variable context because the accumulation of substitutions is limited to mutations that preserve protein function. Invariant residues are not necessarily of pivotal functional importance<sup>171,172</sup>. As interaction networks composition and arrangement form the structural basis of protein function, we computed conserved interactions across classifications. We did not filter out solvent-accessible conserved interactions because initially buried areas can become accessible to antivirals or antibodies following reversible steps of the CC at the cell surface, due to proton release from immune cells or hijacked glycolysis generating extracellular acidification, or within endosomes during intermediate steps of the CC<sup>6,179–185</sup>.

We identified the SB complex which comprises two buried and connected ITR-p salt-bridges located in the vicinity of the b-loop. The first is the ITR-p HA<sub>2</sub>-HA<sub>2</sub> (E74<sub>2</sub>-R76<sub>2</sub>) salt-bridge, strictly conserved among all HAs except clade 7 HAs and H17. The second is the ITR-p HA<sub>1</sub>-HA<sub>2</sub> (E104<sub>1</sub>-R76<sub>2</sub>) salt-bridge, conserved among almost every subtype of group 1 HAs, whereas the ITR-p HA<sub>2</sub>-HA<sub>2</sub> (R76<sub>2</sub>-E81<sub>2</sub>) salt-bridge is strictly conserved among clade 3 HAs.

In the H1 context, non-conservative mutations at any of the positions involved in the complex resulted in loss of function and infectivity, highlighting the interdependence of these ITR-p salt-bridges and the vital role of the complex. The charge-conservative substitutions E104<sub>1</sub>D and E74<sub>2</sub>D were the only viable mutations, but their fusion efficiency was dramatically impaired and the produced viruses showed reduced growth *in vitro*, presumably because of the impact of the shortened sidechain on the spatial arrangement of the complex, as observed for clade 17 HAs. The charge-conservative R76<sub>2</sub>K mutation was

however not viable, highlighting the importance of residue-specific physicochemical properties<sup>9,163–165</sup>.

Remarkably, in the H3 context, all mutants were able to induce cell fusion and yielded infectious viruses.

Our results therefore correlate with the *in vitro* study of the evolutionary pathways of A/WSN/1933 (H1N1) and A/Perth/16/2009 (H3N2) strains<sup>176</sup>. These experiments showed no mutational tolerance of H1 at the positions involved in the ITR-p salt-bridges complex, except for the very minority presence of the R76<sub>2</sub>W mutant which may be a non-functional HA originating from genetic complementation enabled by co-infections<sup>176</sup>. In contrast, in the H3 context, a diversity of residues was tolerated at each of the positions involved in the complex with no preference for charged residues despite their conservation in nature.

Complex salt-bridges stabilize transient folding states by acting as a gatekeeper against unfolding that would disrupt its optimal spatial arrangement while restricting flexibility and movement of the surrounding area<sup>11,12,156</sup>. Hence, the SB complex might contribute to preventing the (b-)loop-to-helix transition at neutral pH in the H1 context. This was supported by the unviable non-conservative mutations and the restoration of some growth potential of the E104<sub>1</sub>D mutant in A549 cells through emergence an additional mutation at the position 76<sub>2</sub>, originally involved in the complex. The R76<sub>2</sub>G mutation disrupted the mutated salt-bridges complex and allowed additional local flexibility due to the absence of sidechain, however at dramatic cost on viral fitness. Most importantly, two independent studies emphasized the role of the ITR-p 74<sub>2</sub>-76<sub>2</sub> salt-bridge in conferring metastable properties<sup>168,186</sup>. By exploring early fusion intermediates of H2 proteins, these studies suggested that further movement of the b-loop would require the disruption of the ITR-p

74<sub>2</sub>-76<sub>2</sub> salt-bridge to enable the (b-)loop-to-helix transition at acidic pH<sup>168,186</sup>. Given the strict conservation of the SB complex in structures sharing the H1 physicochemical profile and that of involved residues in sequences, the identical geometric arrangement of the complex in space, the highly similar distribution of ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges along with the necessity to disrupt the complex to undergo the CC in H2 proteins, we believe our findings regarding the essential nature of the complex in H1 are likely to be extrapolated to other HAs sharing the H1 physicochemical profile, including viruses with high pandemic potential such as H5N1 viruses.

Despite the non-essential nature of the SB complex *in vitro* in H3, its conservation in circulating viruses harboring clade 3 HAs suggests that it could be crucial for infectivity *in vivo*, which remains to be evaluated. Nevertheless, together with our experimental results, the distribution of ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges of different subtypes revealed differences between HAs with different physicochemical profiles, indicating different behaviors of these interactions. This was emphasized by the modulation of interactions during the CC (H3). Following HA<sub>1</sub> dilation and rotation upon acidification of the endosome, buried areas within the trimer become highly solvent accessible<sup>107,114,117</sup> resulting in the protonation of initially negatively charged carboxyl groups<sup>10,160,162,166,167</sup>. Importantly, we detected the total disruption of ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges located at the membrane-proximal domain, and especially that of an ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges complex highly conserved in group 2 HAs. Their disruption lifted their local restrictive effect on unfolding as well as motion, and gave rise to the formation of ITR-p HA<sub>2</sub>-HA<sub>2</sub> +/- repulsions<sup>11,12,156</sup>. It correlates with the observed disorder and structural reorganization of the membrane-proximal area preceding the CC<sup>12,114,187</sup>. Together with the disruption of ITR-p HA<sub>2</sub>-HA<sub>2</sub> hydrophobic interactions, these

electrostatic rearrangements explain the straightening of the membrane-proximal area of HA<sub>2</sub> sub-units and their distancing at acidic pH, enabling the complete extrusion and exposure of the hydrophobic residues of the fusion peptide originally buried at the HA<sub>2</sub>-HA<sub>2</sub> interface<sup>6,106,107,114,117</sup>. The uncovered disruption of the membrane-proximal ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges is thereby in agreement with a previous study demonstrating that an H3 truncated HA<sub>2</sub> trimer, lacking HA<sub>1</sub> sub-units, retains its pre-fusion form at neutral pH but undergoes the CC at acidic pH<sup>188</sup>. Therefore, in the H3 context and probably in the whole group 2 context, the disruption of the membrane-proximal ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges, and especially that of the highly conserved ITR-p salt-bridges complex within group 2 HAs most likely act as CC molecular drivers following HA<sub>1</sub> dilation and subsequent acidification of the core. It will thus be interesting to experimentally assess the functional importance of these salt-bridges.

Overall, our assessment of the functional importance of the conserved SB complex in different environments highlighted that HA evolution is limited by restricted possibilities in terms of viable mutations fitting the interaction networks, according to uncovered physicochemical profiles of interaction networks, and that these profiles influence the molecular drivers of the CC.

Our computational results were limited by heterogeneity in strain diversity among structures of different subtypes and the number of resolved structures. Generalized access to single particle cryo-electron microscopy and artificial intelligence structural predictions (AlphaFold) will provide exponentially more structures<sup>117,189–191</sup>. Therefore, resulting from additional experimental data, enhanced predictions of functional consequences of mutations on the

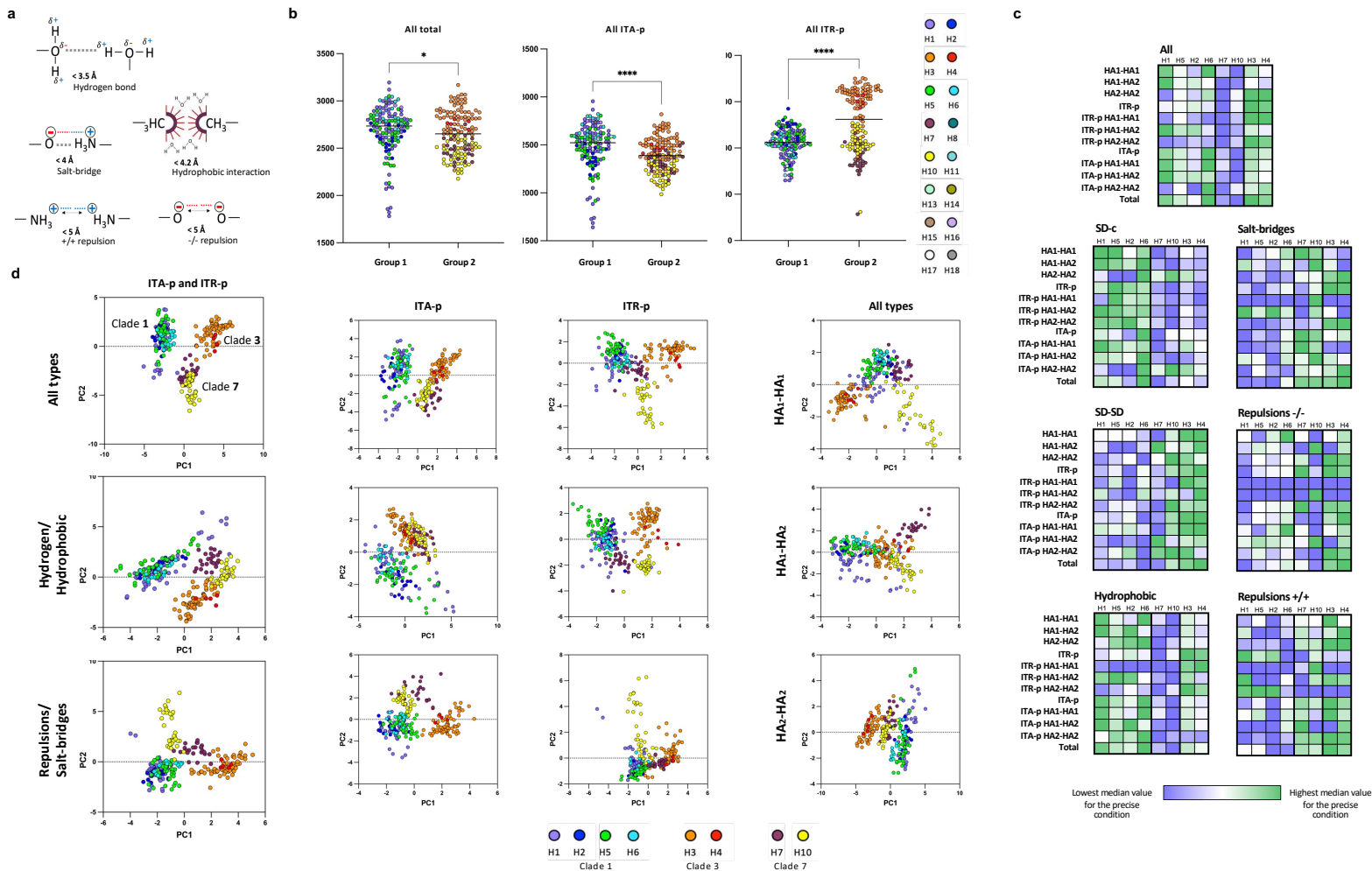


evolutionarily constrained physicochemical profiles of interaction networks should constitute an additional tool for the surveillance of IAVs and monitoring of their evolution.

Given the public health burden of annual influenza epidemics and pandemic threat of IAVs with zoonotic potential, there is a need for effective broad-spectrum antivirals or vaccines that are not prone to viral escape<sup>15,19,82,192,193</sup>. Taking advantage of high variability to identify conserved interactions and validating their functional importance and low mutational tolerance could offer new opportunities for the characterization of broad-spectrum and non-mutable therapeutic targets and vaccine epitopes.

## 3.2.5 Figures

Annex 1 - Figure 1. Divergences in physicochemical interaction networks organization.



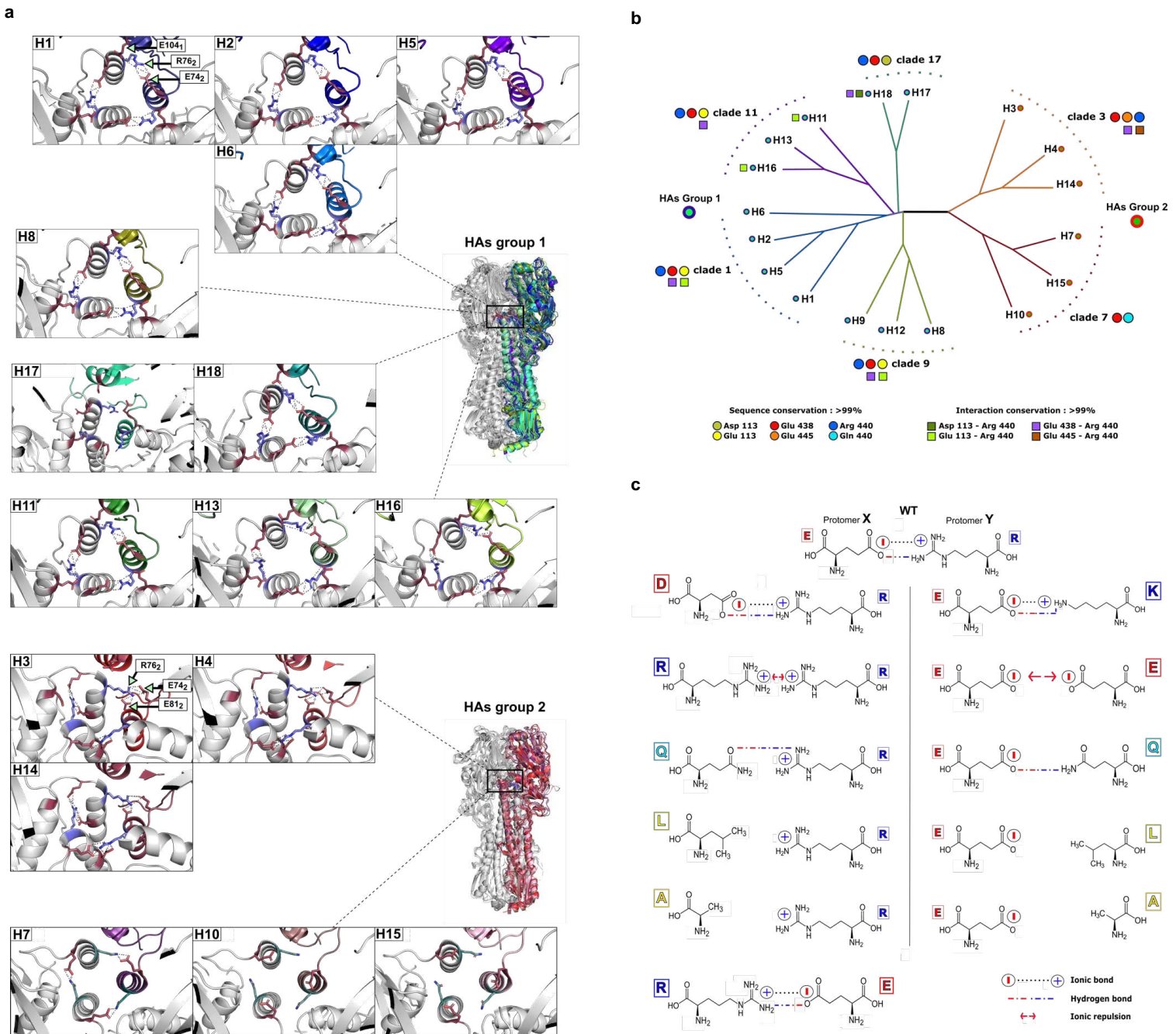
Annex 1 - Figure 1. Divergences in physicochemical interaction networks organization.

- Representation of interaction types included in the analysis: hydrogen bonds (either SD-c or SD-SD), hydrophobic interactions, salt-bridges,  $-/-$  repulsions,  $+/+$  repulsions.
- Dotplots depicting the number of calculated interactions from each structure. Dot colors differentiate subtypes. One dot represents one value for one structure. Values were plotted according to HA group classification to easily identify differences between groups/clades/subtypes. Median value for each group is represented as a black line. Data were compared (group 1  $n = 132$ , group 2  $n = 137$ ) by two-tailed Mann-Whitney test (\*\*\*\*  $P < 0.0001$ , \*\*\*  $P < 0.001$ , \*\*  $P < 0.01$ , \*  $P < 0.05$ , no symbol means not significant).
- Heatmaps representing the median value of the number of specific interaction type/location identified per subtype. The scale is created as follows: for each interaction type/location, the difference between the highest median and the lowest median values among all subtypes are calculated and used as a reference. Then, the

difference between the highest median and the subtype median values are calculated and expressed as percentage of the reference. Deep green represents the highest median values and deep purple corresponds to the lowest median values for each specific condition.

- d. Scatter plots representing PC1 and PC2 coefficients of standardized interaction counts from 258 structures: H1 (n=52), H2 (n=16), H3 (n=70), H4 (n=5), H5 (n=38), H6 (n=17), H7 (n=24), H10 (n=36). Parallel analysis was carried out to select the principal components. Monte Carlo simulations on random data (1000 iterations) were computed and principal components with eigenvalues greater than the 95<sup>th</sup> percentile of the eigenvalues from the simulations were retained. Principle components with the highest variance were selected for plotting. Components represent the standardized interaction number corresponding to considered interaction types and locations conditions as indicated. Unless specified, interaction types include SD-c, SD-SD, hydrophobic, salt-bridges, repulsives -/-, repulsives +/+, and locations include HA<sub>1</sub>-HA<sub>1</sub>, HA<sub>1</sub>-HA<sub>2</sub>, HA<sub>2</sub>-HA<sub>2</sub>.

Annex 1 - Figure 2. Conservation of the ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges complex and mutational strategy to assess its functional importance.



Annex 1 - Figure 2. Conservation of the ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges complex and mutational strategy to assess its functional importance.

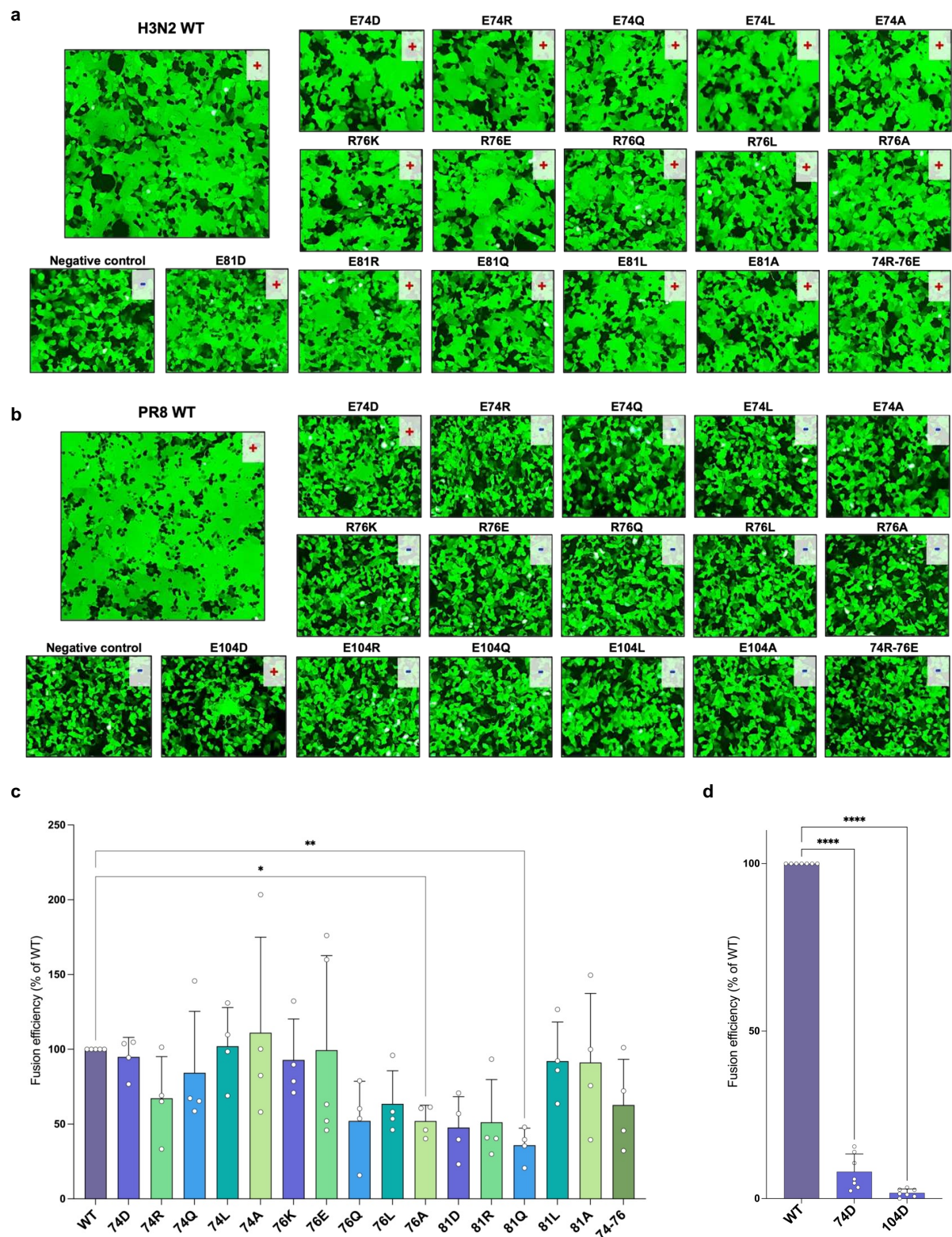
a. Spatial arrangement of the complex. Top-views of the HA<sub>2</sub> interface are shown on structure of subtype representatives and grouped by HA groups and clades. One protomer per structure is colored for clarity.

The ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges complex is represented and the side chain of involved positions are represented to visualize their spatial arrangement. Side chains of positions 104<sub>1</sub> and 74<sub>2</sub> are colored in red and 76<sub>2</sub> in blue (R) or cyan (Q) and actual bonds are shown with dashes. The positions are indicated in the H1 and H3 panels.

b. Conservation of the complex. The sequence alignment and phylogenetic tree were performed with MAFFT (v7) and the phylogenetic tree was visualized with FigTree. Branches are colored according to the six clades. Colored circles and squares indicate residue conservation (>99%) within sequences at positions involved in the complex and interaction conservation (>99%) within structures, respectively. Symbols are indicated for clades, or subtype in the case the conservation within clade is inferior to 99%.

c. Mutagenesis strategy to assess the functional importance of the complex. Mutations are introduced at each of the three positions involved in the complex: E104<sub>1</sub>, E74<sub>2</sub> (protomer X) and R76<sub>2</sub> (protomer Y) for H1 and E74<sub>2</sub>, E81<sub>2</sub> (protomer X) and R76<sub>2</sub> (protomer Y) for H3. A total of 16 mutants per subtype was generated. Five different mutations were introduced to generate specific physicochemical effects regardless of whether a glutamate or an arginine is substituted: charge conservation, opposite charge, absence of charge but possibility to form hydrogen bond, non-polar with some steric-hindrance, non-polar with low steric-hindrance. A charge switch (R74<sub>2</sub>-E76<sub>2</sub>) double mutant was also generated.

Annex 1 - Figure 3. Fusion efficiency of mutant HAs.



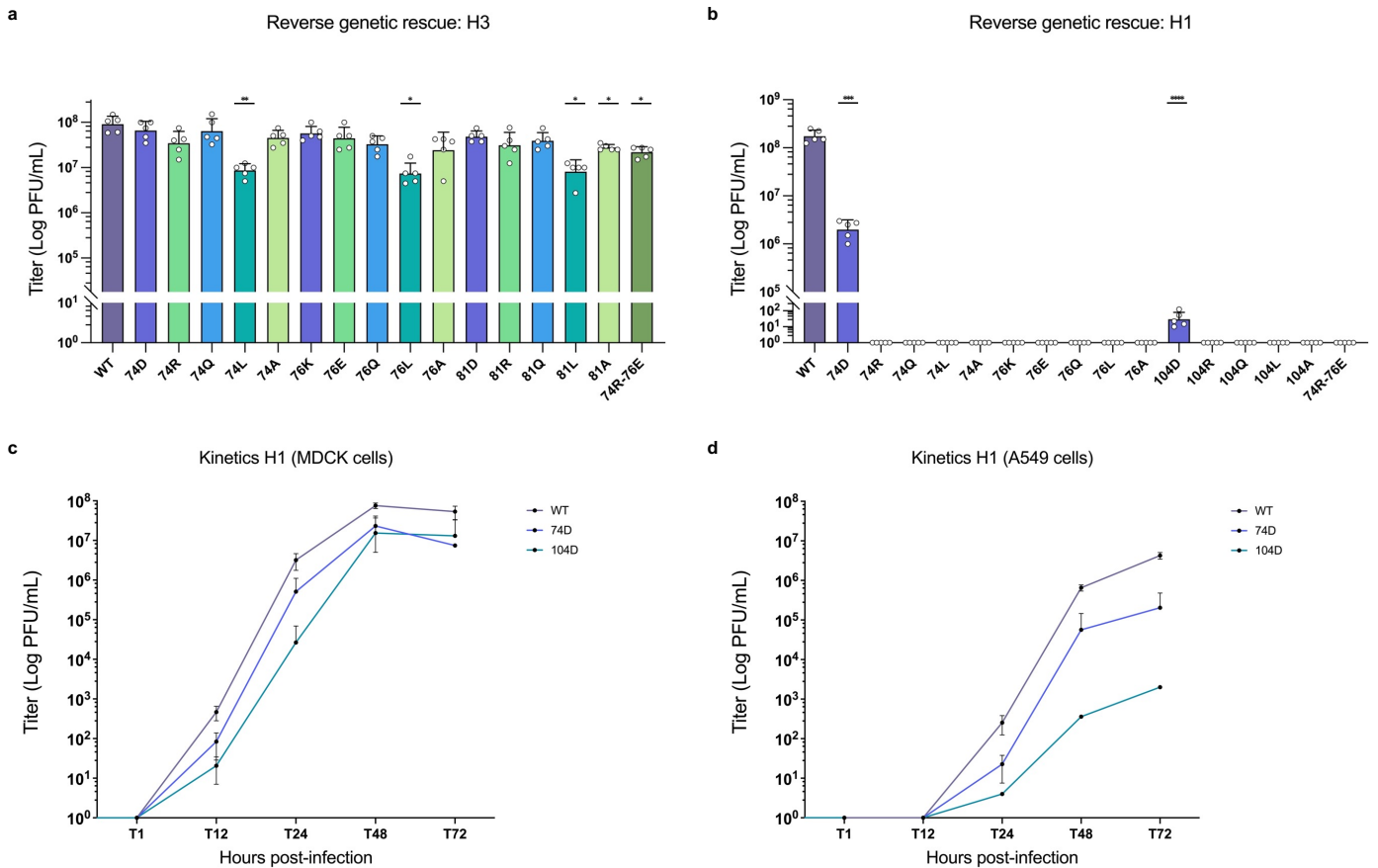
**Annex 1 - Figure 3. Fusion efficiency of mutant HAs.**

Qualitative fusion assay of HAs mutants in the H3 (A/Centre/1003/2012) context (a) and in the H1 (PR8, A/Puerto Rico/8/1934) context (b). Transfected cells were treated by TPCK-

trypsin and exposed to acidic pH (4.6). The ability to induce fusion was validated by the detection of formed syncytia.

Quantitative fusion assay of HAs mutants in the H3 context (c) and in the H1 context (d). Transfected cells were co-seeded with target cells, treated by TPCK-trypsin and exposed to acidic pH (4.6) and fusion efficiency was determined by the detection of firefly luciferase activity. Values are expressed as percentage of WT fusion activity. Data were compared to WT values using Brown-Forsythe and Welch one-way ANOVA test followed by Dunnett's multiple comparisons test (\*\*\*\* adjusted  $P < 0.0001$ , \*\*\* adjusted  $P < 0.001$ , \*\* adjusted  $P < 0.01$ , \* adjusted  $P < 0.05$ , no symbol means not significant).

## Annex 1 - Figure 4. Viability and fitness of mutant viruses.



### Annex 1 - Figure 4. Viability and fitness of mutant viruses.

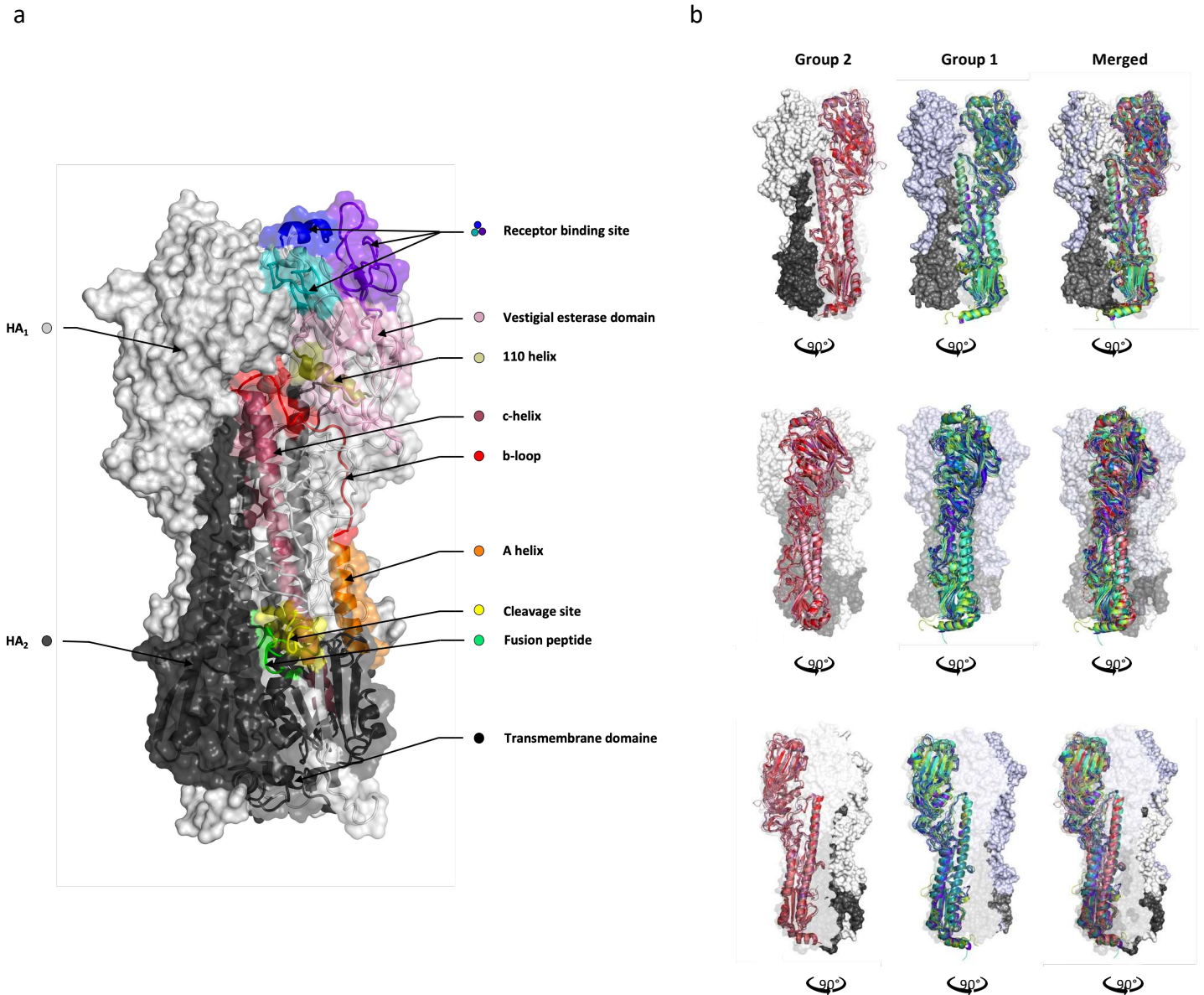
Reverse genetic experiments in the H3 context (a) and in the H1 context (b). Co-culture of HEK293T and MDCK (SIAT in the case of H3) cells were co-transfected with eight bi-directional plasmids corresponding to the eight viral segments, including WT or mutated HAs, treated with TPCK-trypsin to allow for multiple infection cycles. Supernatants were titrated by plaque assays and titers of rescued viruses were expressed in PFU/mL. Data were compared relative to the WT with the RM one-way ANOVA, with the Geisser-Greenhouse correction followed by Holm-Sidak multiple comparisons test (\*\*\*\* adjusted  $P < 0.0001$ , \*\*\* adjusted  $P < 0.001$ , \*\* adjusted  $P < 0.01$ , \* adjusted  $P < 0.05$ , no symbol means not significant).

Growth kinetics of viable mutants (H1) were performed on MDCK (c) and A549 (d) cells. Supernatants were collected at T0-, T12-, T24-, T48- and T72-hours post-infection.



### 3.2.6 Extended data figures/tables

Annex 1 - Extended data figure 1. HA trimer and structural alignments.



**Annex 1 - Extended Data Figure 1. HA trimer and structural alignments.**

a. Trimeric structure of the HA. The trimeric structure of HA (H3 subtype, PDB accession code:3LZG) is represented in surface mode with HA<sub>1</sub> sub-units in white and HA<sub>2</sub> sub-units in black. One protomer is shown in cartoon mode to visualize the different sections. The cleavage site, enabling the processing of precursors by trypsin-like proteases, is colored in yellow. Regarding HA<sub>1</sub> sub-unit, the receptor binding site (RBS) is constituted of the 190-helix, the 130-loop and the 220-loop colored in blue, purple and cyan, respectively. The vestigial esterase domain is colored in pink and the 110-helix in olive. Regarding HA<sub>2</sub>, the a-helix, b-loop and c-helix are represented in orange, red and raspberry, respectively. The fusion

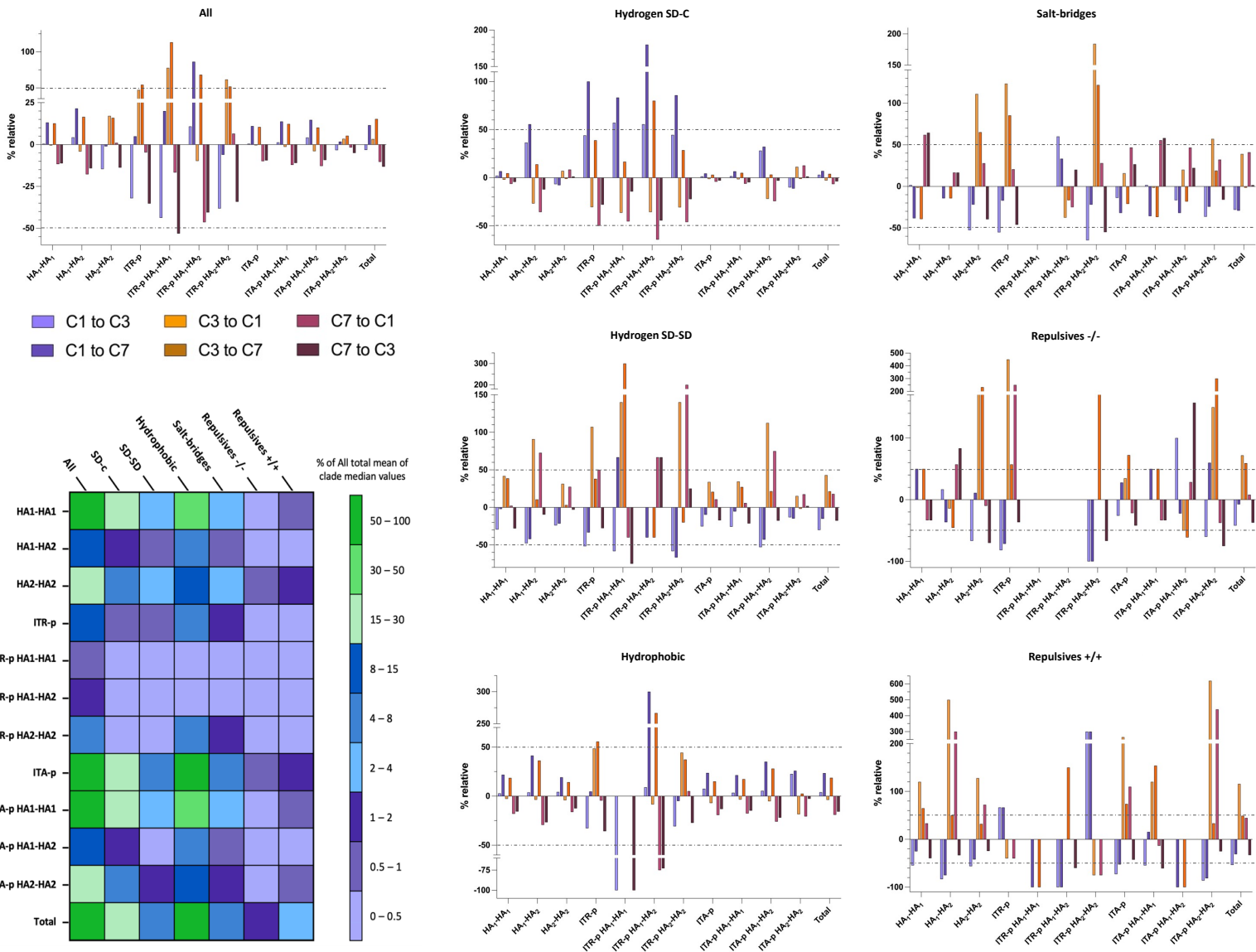
peptide, enabling fusion between cellular and viral membranes, is located at the N-terminus of HA<sub>2</sub> and colored in green. The membrane-proximal domain is represented in black.

b. The figure was made with the CEalign command of PyMOL software (v2.4.1), enabling to carry out a structural alignment based on structural resemblance solely rather than on sequence similarity. One protomer was used to perform the alignment, as aligning the three protomers together resulted in poor alignment quality due to differences in protomer interfaces.

For each subtype with at least one structure gathered in the RSCB PDB (Extended Data Table 1), the structure with the best resolution was chosen to represent the subtype in this alignment. Group 1, group 2 and merged are shown as indicated with 90 degrees rotations. One protomer is shown in cartoon representation. The other two protomers are shown in surface representation.

## Annex 1 - Extended data figure 2. Comparison of median interaction values between clades.

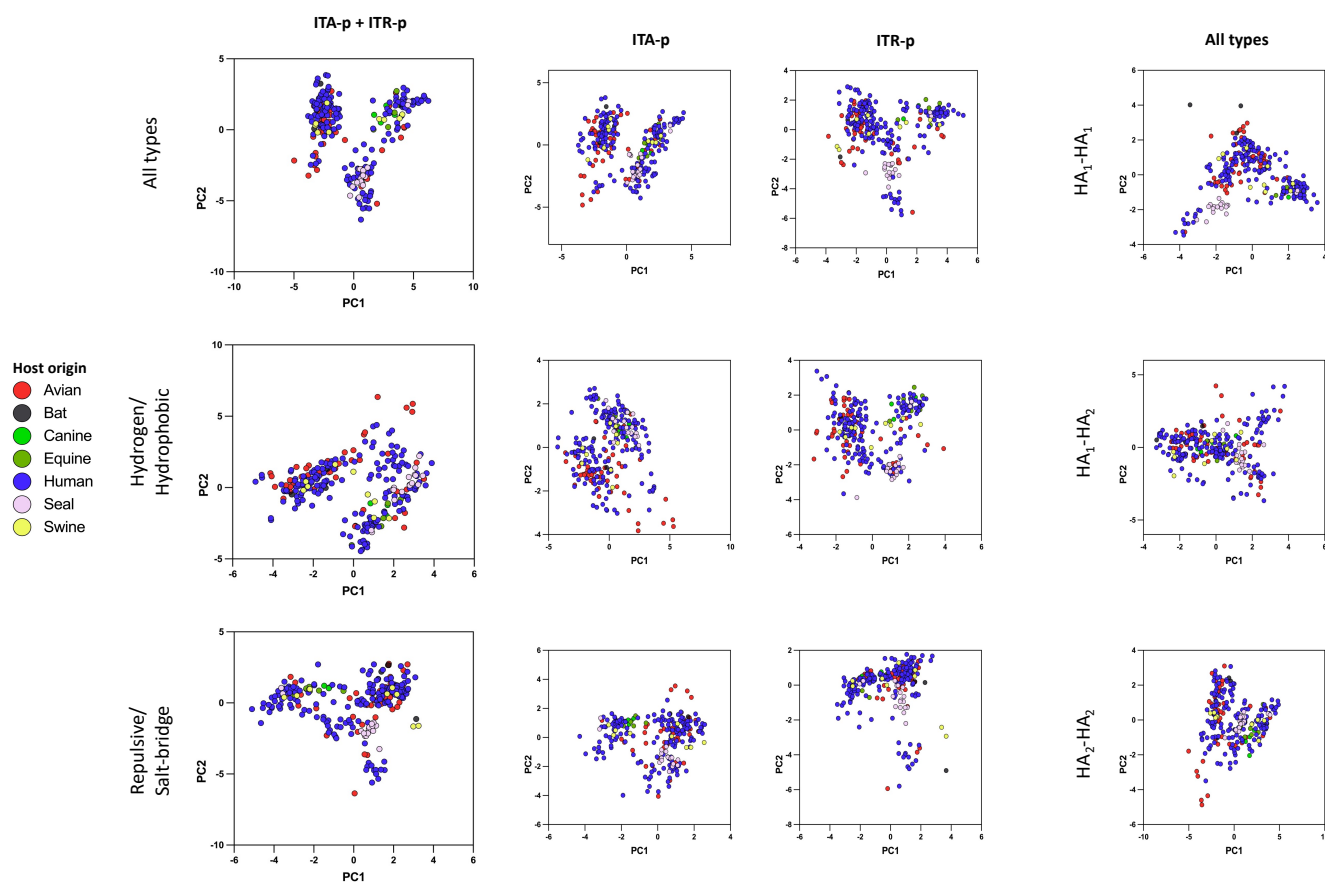
a



## Annex 1 - Extended Data Figure 2. Comparison of median interaction values between clades.

- Grouped bar-plots representing median values of the number of interactions for a given clade relative to other clades expressed as percentage (clade 1 n = 123, clade 3 n = 76, clade 7 n = 61). Plots are represented for the different interaction types as indicated according to the different location possibilities. The following comparisons are represented: clade 1 relative to clade 3 (light purple) and clade 7 (dark purple); clade 3 relative to clade 1 (light orange) and clade 7 (dark orange); clade 7 relative to clade 1 (light burgundy) and clade 3 (dark burgundy). Bars are not represented when one of the compared median value is equal to zero. Horizontal dotted lines indicate +50% and -50% variations.
- Heatmap representing the mean of the median number of interactions, collected from clades, as a percentage of the mean of the median number of total interactions.

Annex 1 - Extended data figure 3. HA physicochemical profiles do not segregate according to host of origin by Principal Component Analysis (PCA).



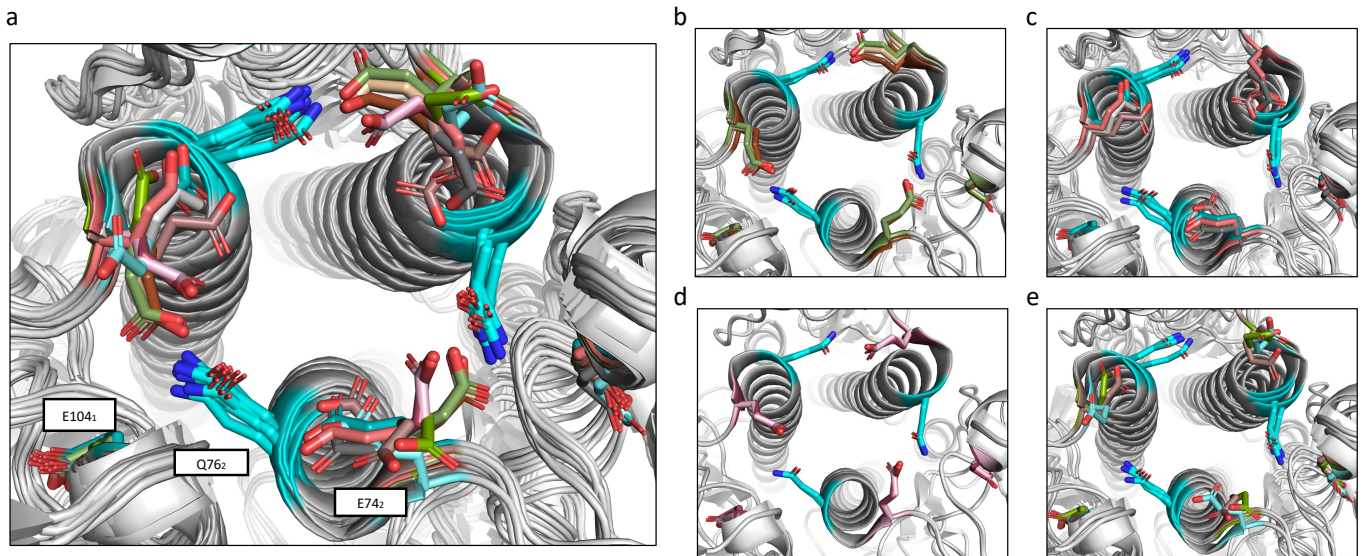
**Annex 1 - Extended Data Figure 3. HA physicochemical profiles do not segregate according to host of origin by Principal Component Analysis (PCA).**

Scatter plots representing PC1 and PC2 coefficients of standardized interaction counts from 269 structures: Human (n=173), avian (n=56), seal (n=18), swine (n=9), equine (n=6), canine (n=4), bat (n=3).

Parallel analysis was carried out to select the principal components. Monte Carlo simulations on random data (1000 iterations) were computed and principal components with eigenvalues greater than the 95<sup>th</sup> percentile of the eigenvalues from the simulations were retained. Principle components with the highest variance were selected for plotting.

Components represent the standardized interaction number corresponding to considered interaction types and locations as indicated. Unless specified, interaction types include SD-c, SD-SD, hydrophobic, salt-bridges, repulsives -/-, repulsives +/-, and locations include ITA-p HA<sub>1</sub>-HA<sub>1</sub>, HA<sub>1</sub>-HA<sub>2</sub>, HA<sub>2</sub>HA<sub>2</sub> and ITR-p HA<sub>1</sub>-H<sub>1</sub>, HA<sub>1</sub>-HA<sub>2</sub>, HA<sub>2</sub>HA<sub>2</sub>.

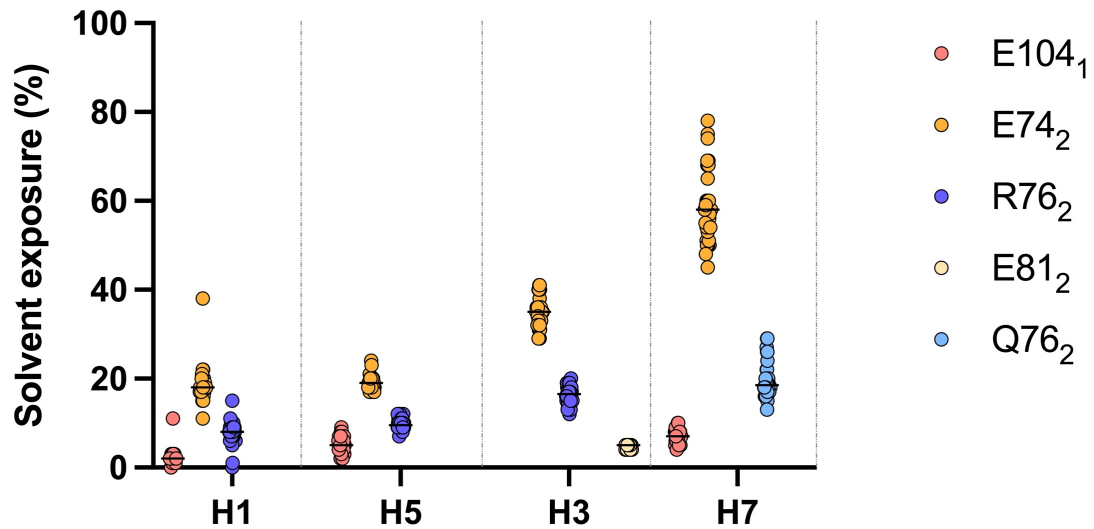
Annex 1 - Extended data figure 4. Instability of the ITR-p HA<sub>2</sub>-HA<sub>2</sub> (E74<sub>2</sub>-Q76<sub>2</sub>) hydrogen bonds in the H7 context.



**Annex 1 - Extended Data Figure 4. Instability of the ITR-p HA<sub>2</sub>-HA<sub>2</sub> (E74<sub>2</sub>-Q76<sub>2</sub>) hydrogen bonds in the H7 context.**

Multiple HA structures the H7 subtypes were aligned to visualize the instability of the ITR-p HA<sub>2</sub>-HA<sub>2</sub> (E74<sub>2</sub>-Q76<sub>2</sub>) hydrogen bonds in the H7 context. Alignments were performed with the *cealign* command of PyMOL software (v2.4.1). Positions 104<sub>1</sub>, 74<sub>2</sub> and 76<sub>2</sub> are represented. 76<sub>2</sub> is colored in cyan, 104<sub>1</sub> and 74<sub>2</sub> are colored according to the different structures aligned. PDB codes accessions: 4NL6, 4NL8, 6D7c (b); 5TS6, 6IDD, 6FYU (c); 4LN3, 4LN4 (d); 6DYU, 6I18, 6I19 (e); all together (a).

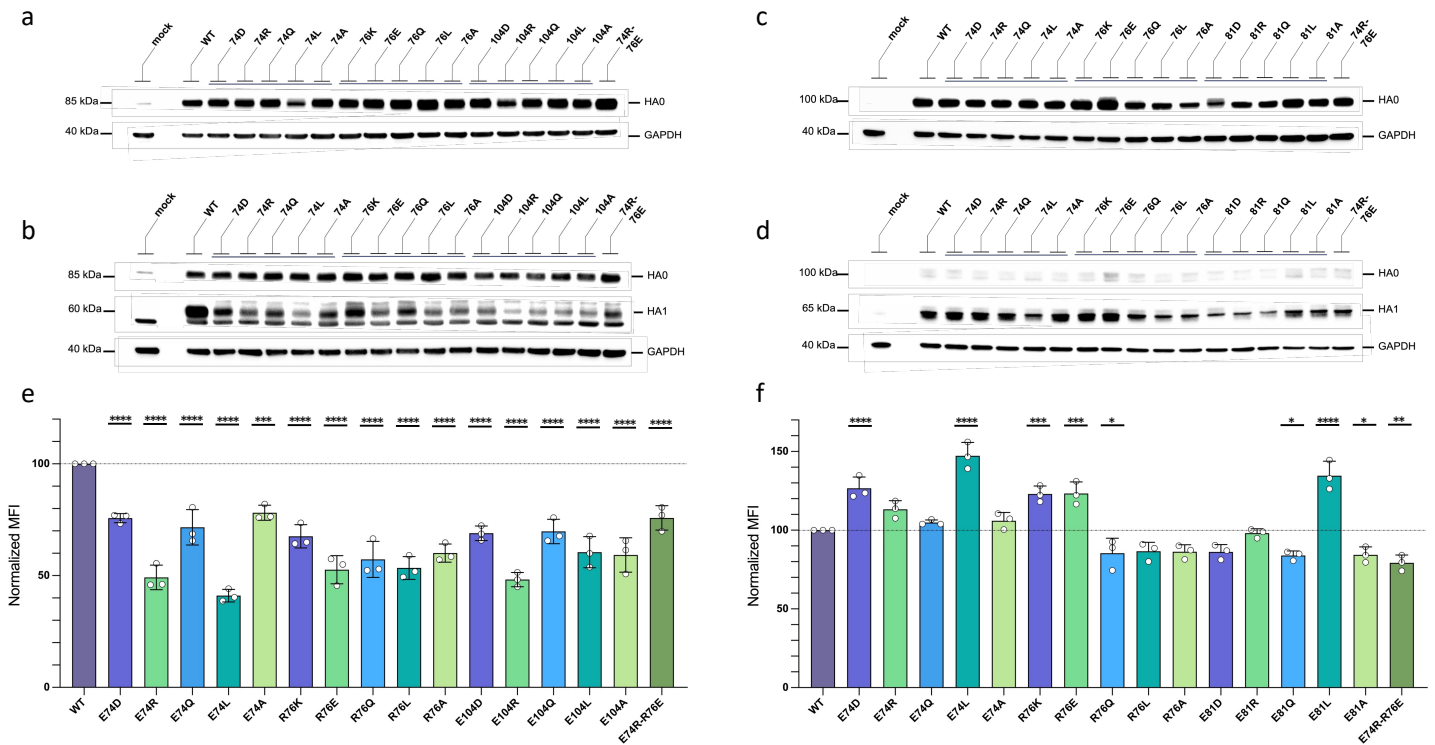
Annex 1 - Extended data figure 5. Solvent exposure of positions involved in the conserved complex.



**Annex 1 - Extended Data Figure 5. Solvent exposure of positions involved in the conserved complex.**

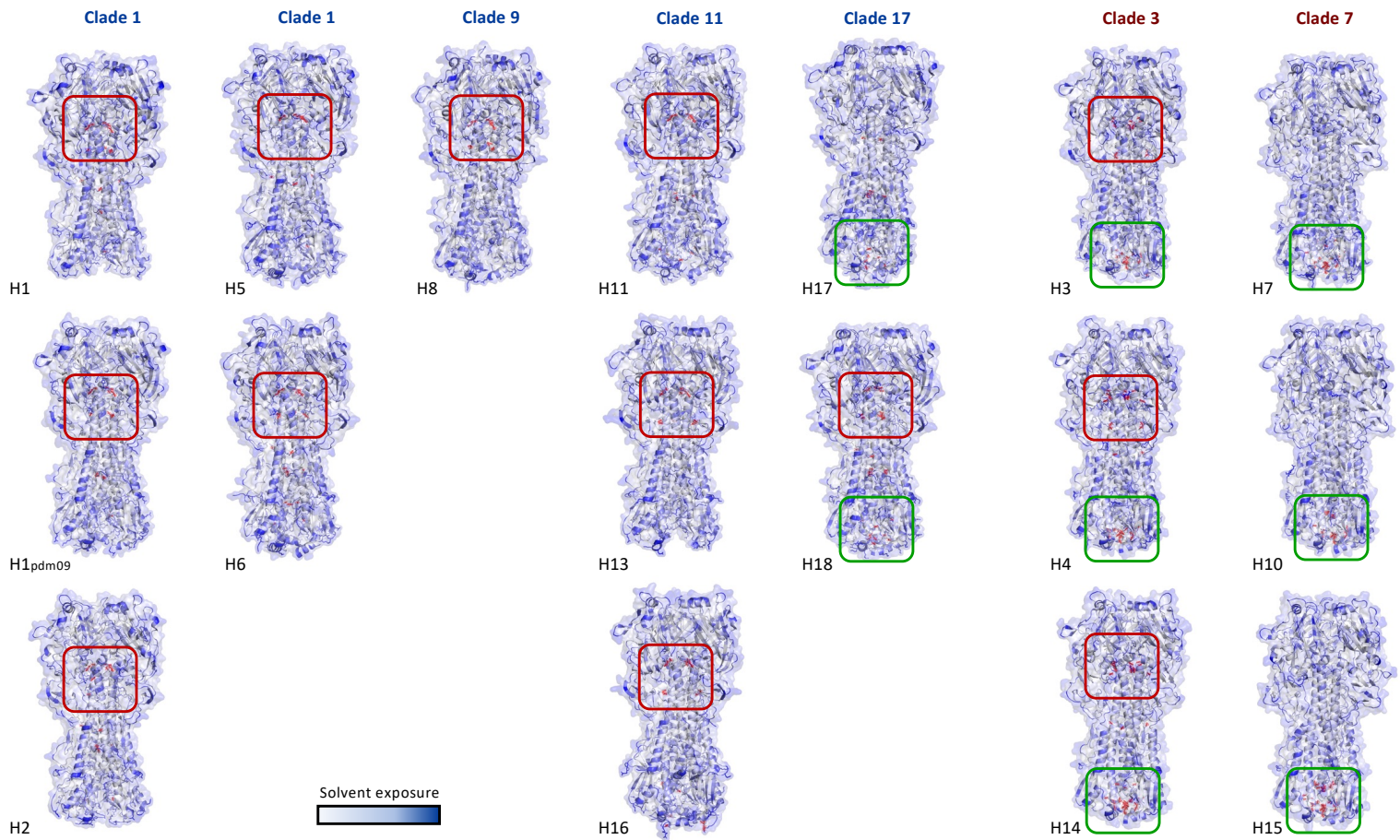
The relative per-residue solvent accessible surface area of the positions involved in the ITR-p salt-bridges complex in the H1, H5, H3 and H7 context were calculated with PyMOL (v2.4.1). Ten trimeric structures were used per subtypes. The calculated values represent the exposure of a residue expressed as a percentage of its full exposure with 0% = fully buried and 100% = full exposed to solvent, calculated by removing all residues except its neighbors.

## Annex 1 - Extended data figure 6. Expression of mutant HAs.



**Annex 1 - Extended Data Figure 6. Expression of mutant HAs.** Total expression of the mutant precursors HA<sub>0</sub> in the H1 context (a) and in the H3 context (c). Proteolytic cleavage of mutants expressed at the cell surface by TPCK-trypsin in the H1 context (b) and in the H3 context (d). Cleavage is attested by the presence of HA<sub>1</sub> band. Expression of HAs mutants at the cell surface quantified by flow-cytometry in the H1 context (e) and in the H3 context (f). Expression of mutant proteins expressed as a percentage of WT surface expression. WT expression level is represented as a black line. Data were compared to WT expression level using one-way ANOVA test followed by Dunnett's multiple comparisons test (\*\*\*\* adjusted P < 0.0001, \*\*\* adjusted P < 0.001, \*\* adjusted P < 0.01, \* adjusted P < 0.05, no symbol means not significant).

Annex 1 - Extended data figure 7. ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges distribution throughout the stem.

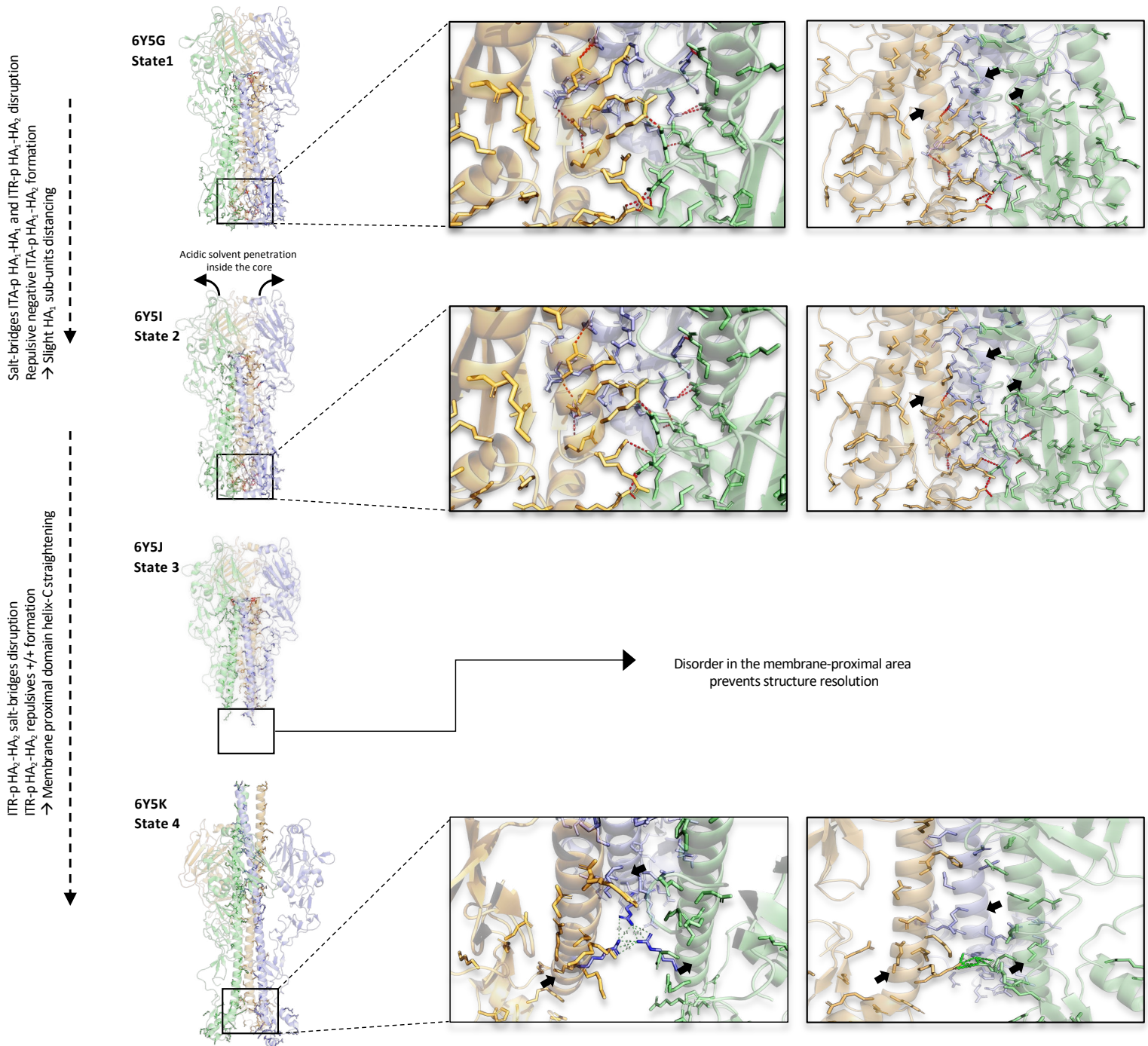


**Annex 1 - Extended Data Figure 7. ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges distribution throughout the stem.**

Representatives of subtypes for which at least one structure has been resolved. On each structure, ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges were represented with red dashes to visualize their distribution. Subtypes were vertically grouped according to their clade appartenance (blue = clades from group 1, red = clades from group 2). ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges, either located at the top of a-helices or at the membrane-proximal domain, were boxed in red and green, respectively. PDB accessing codes of represented structures are the followings: H1:1RU7, H1pdm09:3LZG, H2:2WR3, H5:5KDO, H6:5T08, H8:6V46, H11:6V47, H13:4KPS, H16:4F23, H17:4H32, H18:4K3X, H3:6MXU, H4:5XL8, H14:6V48, H7:4NL6, H10:4CYV, H15:6V49.



Annex 1 - Extended data figure 8. ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges disruption and ITR-p HA<sub>2</sub>-HA<sub>2</sub> +/- repulsions formation during the conformational change.



**Annex 1 - Extended Data Figure 8. ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges disruption and ITR-p HA<sub>2</sub>-HA<sub>2</sub> +/- repulsions formation during the conformational change.** Cartoon representations of HA trimer structures of state 1, 2, 3 and 4 intermediates during the CC (DOI: [10.1038/s41586-020-2333-6](https://doi.org/10.1038/s41586-020-2333-6)). Protomers are represented in different colors. Side chains of charged amino acids are represented as sticks. ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges are shown as red dots and ITR-p

HA<sub>2</sub>-HA<sub>2</sub> +/- repulsions, only present in state 4, as green dots with distance labels in Å. The insets show views of the membrane proximal domain where most of these interactions are found. Black arrows in insets indicate the C-helix curved (6Y5G, 6Y5I) to straightened (6Y5K) steps.

## Annex 1 - Extended data table 1. Dataset of influenza A virus hemagglutinin trimeric structures

PDB access-ion code	Subtype	Strain	Host of origin	Mutation	Complex	Method	Resolution (Å)	Reference (DOI)
<b>H1N1</b>								
4GXX	H1N1	A/Brevig Mission/1/1918	Human	D190E/D225G (HA1)	–	X-RAY DIFFRACTION	1.799	<a href="https://doi.org/10.1371/journal.ppat.1003067">10.1371/journal.ppat.1003067</a>
6ONA	H1N1	A/Hickox/JY2/1940	Human	–	–	X-RAY DIFFRACTION	1.95	<a href="https://doi.org/10.2210/pdb6ONA/pdb">10.2210/pdb6ONA/pdb</a>
3UBQ	H1N1	A/California/04/2009	Human	G205C/R220C (HA1)	3SLN	X-RAY DIFFRACTION	2	<a href="https://doi.org/10.1128/JVI.06322-11">10.1128/JVI.06322-11</a>
5VMC	H1N1	A/Brevig Mission/1/1918	Human	Δ133 (HA1)	6SLN	X-RAY DIFFRACTION	2.15	<a href="https://doi.org/10.1016/j.virol.2017.10.010">10.1016/j.virol.2017.10.010</a>
3UBE	H1N1	A/California/04/2009	Human	G205C/R220C (HA1)	LSTc	X-RAY DIFFRACTION	2.15	<a href="https://doi.org/10.1128/JVI.06322-11">10.1128/JVI.06322-11</a>
4M4Y	H1N1	A/California/04/2009	Human	E47G (HA2)	–	X-RAY DIFFRACTION	2.2	<a href="https://doi.org/10.1128/JVI.01388-13">10.1128/JVI.01388-13</a>
1RVX	H1N1	A/Puerto Rico/8/1934	Human	–	LSTa	X-RAY DIFFRACTION	2.2	<a href="https://doi.org/10.1126/science.1093155">10.1126/science.1093155</a>
3UBJ	H1N1	A/California/04/2009	Human	–	LSTa	X-RAY DIFFRACTION	2.25	<a href="https://doi.org/10.1128/JVI.06322-11">10.1128/JVI.06322-11</a>
1RVZ	H1N1	A/Puerto Rico/8/1934	Human	–	LSTc	X-RAY DIFFRACTION	2.25	<a href="https://doi.org/10.1126/science.1093155">10.1126/science.1093155</a>
1RU7	H1N1	A/Puerto Rico/8/1934	Human	–	–	X-RAY DIFFRACTION	2.3	<a href="https://doi.org/10.1126/science.1093155">10.1126/science.1093155</a>
6D8W	H1N1	A/Jiangsu/ALSI/2011	Human	–	–	X-RAY DIFFRACTION	2.35	<a href="https://doi.org/10.2210/pdb6D8W/pdb">10.2210/pdb6D8W/pdb</a>
5VMF	H1N1	A/Brevig Mission/1/1918	Human	Q226L/G228S (HA1)	6SLN	X-RAY DIFFRACTION	2.35	<a href="https://doi.org/10.1016/j.virol.2017.10.010">10.1016/j.virol.2017.10.010</a>
5VMG	H1N1	A/Brevig Mission/1/1918	Human	Δ133/D190E/D225G/Q226L/G228S (HA1)	6SLN	X-RAY DIFFRACTION	2.45	<a href="https://doi.org/10.1016/j.virol.2017.10.010">10.1016/j.virol.2017.10.010</a>
4EDB	H1N1	A/Thailand/CU44/2006	Human	–	Fab fragment	X-RAY DIFFRACTION	2.5	<a href="https://doi.org/10.1099/vir.0.051136-0">10.1099/vir.0.051136-0</a>
6N41	H1N1	A/Netherlands/002P1/1951	Human	–	–	X-RAY DIFFRACTION	2.5	<a href="https://doi.org/10.2210/pdb6N41/pdb">10.2210/pdb6N41/pdb</a>
3UBN	H1N1	A/California/04/2009	Human	G205C/R220C (HA1)	6SLN	X-RAY DIFFRACTION	2.5079	<a href="https://doi.org/10.1128/JVI.06322-11">10.1128/JVI.06322-11</a>
6OSR	H1N1	A/Melbourne/1/1946	Human	–	–	X-RAY DIFFRACTION	2.55	<a href="https://doi.org/10.2210/pdb6OSR/pdb">10.2210/pdb6OSR/pdb</a>
3LZG	H1N1	A/California/04/2009	Human	–	–	X-RAY DIFFRACTION	2.6	<a href="https://doi.org/10.1126/science.1186430">10.1126/science.1186430</a>
4JUG	H1N1	A/South Carolina/1/1918	Human	D225G (HA1)	–	X-RAY DIFFRACTION	2.7	<a href="https://doi.org/10.1128/JVI.00545-13">10.1128/JVI.00545-13</a>
4EEF	H1N1	A/Brevig Mission/1/1918	Human	–	F-HB80.4 (inhibitor)	X-RAY DIFFRACTION	2.704	<a href="https://doi.org/10.1038/nbt.2214">10.1038/nbt.2214</a>
5W6R	H1N1	A/Puerto Rico/8/1934	Human	–	CP141099 (peptide)	X-RAY DIFFRACTION	2.731	<a href="https://doi.org/10.1126/science.aan0516">10.1126/science.aan0516</a>
3M6S	H1N1	A/Darwin/2001/2009	Human	S203T/R205K (HA1) V91I (HA2)	–	X-RAY DIFFRACTION	2.8	<a href="https://doi.org/10.1371/currents.RRN1152">10.1371/currents.RRN1152</a>
6LKS	H1N1	A/Thailand/CU44/2006	Human	–	–	X-RAY DIFFRACTION	2.8	<a href="https://doi.org/10.1038/s41598-020-72368-x">10.1038/s41598-020-72368-x</a>
5UGY	H1N1	A/Solomon Islands/3/2006	Human	–	Neutralizing antibody	X-RAY DIFFRACTION	2.801	<a href="https://doi.org/10.1073/pnas.1111497108">10.1073/pnas.1111497108</a>
4JUH	H1N1	A/South Carolina/1/1918	Human	D225G (HA1)	LSTa	X-RAY DIFFRACTION	2.805	<a href="https://doi.org/10.1128/JVI.00545-13">10.1128/JVI.00545-13</a>
3AL4	H1N1	A/California/04/2009	Human	–	–	X-RAY DIFFRACTION	2.872	<a href="https://doi.org/10.1007/s13238-010-0059-1">10.1007/s13238-010-0059-1</a>
1RUZ	H1N1	A/South Carolina/1/1918	Human	–	–	X-RAY DIFFRACTION	2.9	<a href="https://doi.org/10.1126/science.1093155">10.1126/science.1093155</a>
4JUO	H1N1	A/California/04/2009	Human	D225E (HA1)	LSTc	X-RAY DIFFRACTION	2.908	<a href="https://doi.org/10.1128/JVI.00545-13">10.1128/JVI.00545-13</a>
5VMJ	H1N1	A/Brevig Mission/1/1918	Human	Δ133/D190E/D225G/Q226L/G228S (HA1)	3SLN	X-RAY DIFFRACTION	2.95	<a href="https://doi.org/10.1016/j.virol.2017.10.010">10.1016/j.virol.2017.10.010</a>
4JTV	H1N1	A/California/04/2009	Human	–	LSTc	X-RAY DIFFRACTION	2.997	<a href="https://doi.org/10.1128/JVI.00545-13">10.1128/JVI.00545-13</a>
4JTX	H1N1	A/California/04/2009	Human	D225E (HA1)	–	X-RAY DIFFRACTION	2.997	<a href="https://doi.org/10.1128/JVI.00545-13">10.1128/JVI.00545-13</a>
2WRG	H1N1	A/Brevig Mission/1/1918	Human	–	6SLN	X-RAY DIFFRACTION	3	<a href="https://doi.org/10.1073/pnas.0906849106">10.1073/pnas.0906849106</a>
4LXV	H1N1	A/Washington/5/2011	Human	–	–	X-RAY DIFFRACTION	3	<a href="https://doi.org/10.1128/JVI.02278-13">10.1128/JVI.02278-13</a>
4JUJ	H1N1	A/South Carolina/1/1918	Human	D225G (HA1)	LSTc	X-RAY DIFFRACTION	3	<a href="https://doi.org/10.1128/JVI.00545-13">10.1128/JVI.00545-13</a>

1RD8	H1N1	A/South Carolina/1/1918	Human	-	-	X-RAY DIFFRACTION	3	<a href="https://doi.org/10.1126/science.1093373">10.1126/science.1093373</a>
4GXU	H1N1	A/Brevig Mission/1/1918	Human	-	1F1 antibody	X-RAY DIFFRACTION	3	<a href="https://doi.org/10.1371/journal.ppat.1003067">10.1371/journal.ppat.1003067</a>
5WKO	H1N1	A/California/04/2009	Human	E47G (HA2)	27F3 antibody	X-RAY DIFFRACTION	3.013	<a href="https://doi.org/10.1016/j.celrep.2017.08.084">10.1016/j.celrep.2017.08.084</a>
5UJZ	H1N1	A/Solomon Islands/03/2006	Human	-	Antibody (RBS)	ELECTRON MICROSCOPY	3.294	<a href="https://doi.org/10.1016/j.jmb.2017.05.011">10.1016/j.jmb.2017.05.011</a>
5UK0	H1N1	A/Solomon Islands/03/2006	Human	-	Antibody (RBS)	ELECTRON MICROSCOPY	3.492	<a href="https://doi.org/10.1016/j.jmb.2017.05.011">10.1016/j.jmb.2017.05.011</a>
6WJ1	H1N1	A/California/04/2009	Human	-	54-4H03 Fab fragment	X-RAY DIFFRACTION	3.5	<a href="https://doi.org/10.1016/j.chom.2020.06.003">10.1016/j.chom.2020.06.003</a>
6XGC	H1N1	A/Michigan/45/2015	Human	-	1C4 antibody	ELECTRON MICROSCOPY	4.1	<a href="https://doi.org/10.1126/scitranslmed.abe5449">10.1126/scitranslmed.abe5449</a>
5UK1	H1N1	A/Solomon Islands/03/2006	Human	-	Antibody (RBS)	ELECTRON MICROSCOPY	4.8	<a href="https://doi.org/10.1016/j.jmb.2017.05.011">10.1016/j.jmb.2017.05.011</a>
5UK2	H1N1	A/Solomon Islands/03/2006	Human	-	Antibody (RBS)	ELECTRON MICROSCOPY	4.8	<a href="https://doi.org/10.1016/j.jmb.2017.05.011">10.1016/j.jmb.2017.05.011</a>
<b>H1N1</b>								
2WRH	H1N1	A/mallard/Alberta/35/1976	Avian	-	6SLN	X-RAY DIFFRACTION	3	<a href="https://doi.org/10.1073/pnas.0906849106">10.1073/pnas.0906849106</a>
6HJN	H1N1	A/mallard/Alberta/35/1976	Avian	-	-	ELECTRON MICROSCOPY	3.3	<a href="https://doi.org/10.1073/pnas.1810927115">10.1073/pnas.1810927115</a>
6HJP	H1N1	A/mallard/Alberta/35/1976	Avian	-	FISW84 Fab Fragment	ELECTRON MICROSCOPY	3.3	<a href="https://doi.org/10.1073/pnas.1810927115">10.1073/pnas.1810927115</a>
6HJQ	H1N1	A/mallard/Alberta/35/1976	Avian	-	FISW84 Fab Fragment	ELECTRON MICROSCOPY	4.1	<a href="https://doi.org/10.1073/pnas.1810927115">10.1073/pnas.1810927115</a>
6HJR	H1N1	A/mallard/Alberta/35/1976	Avian	-	-	ELECTRON MICROSCOPY	4.2	<a href="https://doi.org/10.1073/pnas.1810927115">10.1073/pnas.1810927115</a>
<b>H1N1</b>								
1RV0	H1N1	A/Swine/Iowa/15/1930	Swine	-	LSTa	X-RAY DIFFRACTION	2.5	<a href="https://doi.org/10.1126/science.1093155">10.1126/science.1093155</a>
1RVT	H1N1	A/Swine/Iowa/15/1930	Swine	-	LSTc	X-RAY DIFFRACTION	2.5	<a href="https://doi.org/10.1126/science.1093155">10.1126/science.1093155</a>
1RUY	H1N1	A/Swine/Iowa/15/1930	Swine	-	-	X-RAY DIFFRACTION	2.7	<a href="https://doi.org/10.1126/science.1093155">10.1126/science.1093155</a>
<b>H1N2</b>								
4F3Z	H1N2	A/sw/Indiana/P12439/00	Swine	G205C/R220C (HA1)	-	X-RAY DIFFRACTION	3.2	<a href="https://doi.org/10.1128/JVI.00697-12">10.1128/JVI.00697-12</a>
<b>H2N1</b>								
2WR3	H2N1	A/duck/Ontario/1977	Avian	-	3SLN	X-RAY DIFFRACTION	2.5	<a href="https://doi.org/10.1073/pnas.0906849106">10.1073/pnas.0906849106</a>
2WR4	H2N1	A/duck/Ontario/1977	Avian	-	6SLN	X-RAY DIFFRACTION	2.5	<a href="https://doi.org/10.1073/pnas.0906849106">10.1073/pnas.0906849106</a>
2WR5	H2N1	A/duck/Ontario/1977	Avian	-	-	X-RAY DIFFRACTION	2.6	<a href="https://doi.org/10.1073/pnas.0906849106">10.1073/pnas.0906849106</a>
<b>H2N2</b>								
2WR7	H2N2	A/Singapore/1/1957	Human	-	6SLN	X-RAY DIFFRACTION	2.5	<a href="https://doi.org/10.1073/pnas.0906849106">10.1073/pnas.0906849106</a>
2WRC	H2N2	A/Singapore/1/1957	Human	-	-	X-RAY DIFFRACTION	2.7	<a href="https://doi.org/10.1073/pnas.0906849106">10.1073/pnas.0906849106</a>
4HLZ	H2N2	A/Japan/305/1957	Human	-	C179 Fab Fragment	X-RAY DIFFRACTION	2.9	<a href="https://doi.org/10.1128/JVI.02975-12">10.1128/JVI.02975-12</a>
3QOQ	H2N2	A/Japan/305/1957	Human	R106H (HA2)	-	X-RAY DIFFRACTION	2.9	<a href="https://doi.org/10.1128/JVI.02430-10">10.1128/JVI.02430-10</a>
2WRD	H2N2	A/Japan/305/1957	Human	-	-	X-RAY DIFFRACTION	3	<a href="https://doi.org/10.1073/pnas.0906849106">10.1073/pnas.0906849106</a>
2WRE	H2N2	A/Japan/305/1957	Human	-	6SLN	X-RAY DIFFRACTION	3.001	<a href="https://doi.org/10.1073/pnas.0906849106">10.1073/pnas.0906849106</a>
2WRB	H2N2	A/Singapore/1/1957	Human	-	3SLN	X-RAY DIFFRACTION	3.1	<a href="https://doi.org/10.1073/pnas.0906849106">10.1073/pnas.0906849106</a>
4HG4	H2N2	A/Japan/305/1957	Human	-	2G1 Fab Fragment	X-RAY DIFFRACTION	3.2	<a href="https://doi.org/10.1038/nsmb.2500">10.1038/nsmb.2500</a>

2WR1	H2N2	A/chicken/New York/29878/1991	Avian	-	6SLN	X-RAY DIFFRACTION	2.1	<a href="https://doi.org/10.1073/pnas.0906849106">10.1073/pnas.0906849106</a>
2WR0	H2N2	A/chicken/New York/29878/1991	Avian	-	-	X-RAY DIFFRACTION	2.4	<a href="https://doi.org/10.1073/pnas.0906849106">10.1073/pnas.0906849106</a>
2WR2	H2N2	A/chicken/New York/29878/1991	Avian	-	3SLN	X-RAY DIFFRACTION	2.4	<a href="https://doi.org/10.1073/pnas.0906849106">10.1073/pnas.0906849106</a>
2WRF	H2N2	A/chicken/Potsdam/4705/1984	Avian	-	6SLN	X-RAY DIFFRACTION	3.1	<a href="https://doi.org/10.1073/pnas.0906849106">10.1073/pnas.0906849106</a>
<b>H2N3</b>								
4W8N	H2N3	A/swine/Missouri/2124514/2006	Swine	-	-	X-RAY DIFFRACTION	2.9	<a href="https://doi.org/10.1016/j.virol.2015.01.002">10.1016/j.virol.2015.01.002</a>
<b>H3N2</b>								
6MXU	H3N2	A/Texas/1/1977	Human	-	-	X-RAY DIFFRACTION	1.85	<a href="https://doi.org/10.2210/pdb6MXU/pdb">10.2210/pdb6MXU/pdb</a>
4FNK	H3N2	A/Hong Kong/1/1968	Human	-	-	X-RAY DIFFRACTION	1.901	<a href="https://doi.org/10.1038/nature11414">10.1038/nature11414</a>
6N08	H3N2	A/Netherlands/209/1980	Human	-	-	X-RAY DIFFRACTION	1.916	<a href="https://doi.org/10.2210/pdb6N08/pdb">10.2210/pdb6N08/pdb</a>
6NHR	H3N2	A/Hong Kong/1/1968	Human	I45F (HA2)	-	X-RAY DIFFRACTION	2.1	<a href="https://doi.org/10.1126/science.aaz5143">10.1126/science.aaz5143</a>
5VTZ	H3N2	A/Hong Kong/1/1968	Human	G225Q/L226A (HA1)	-	X-RAY DIFFRACTION	2.15	<a href="https://doi.org/10.1016/j.chom.2017.05.011">10.1016/j.chom.2017.05.011</a>
6BKM	H3N2	A/Hong Kong/1/1968	Human	E190D (HA1)	-	X-RAY DIFFRACTION	2.2	<a href="https://doi.org/10.1038/s41467-018-03663-5">10.1038/s41467-018-03663-5</a>
6TZB	H3N2	A/Hong Kong/1/1968	Human	-	6SLN	X-RAY DIFFRACTION	2.243	<a href="https://doi.org/10.1038/s41467-020-15102-5">10.1038/s41467-020-15102-5</a>
5VTV	H3N2	A/Hong Kong/1/1968	Human	G225M/L226T/S228A (HA1)	3SLN	X-RAY DIFFRACTION	2.25	<a href="https://doi.org/10.1016/j.chom.2017.05.011">10.1016/j.chom.2017.05.011</a>
5VU4	H3N2	A/Hong Kong/1/1968	Human	G225Q/L226A (HA1)	6SLN	X-RAY DIFFRACTION	2.25	<a href="https://doi.org/10.1016/j.chom.2017.05.011">10.1016/j.chom.2017.05.011</a>
6BKQ	H3N2	A/Hong Kong/1/1968	Human	E190D (HA1)	6SLN	X-RAY DIFFRACTION	2.25	<a href="https://doi.org/10.1038/s41467-018-03663-5">10.1038/s41467-018-03663-5</a>
6NHP	H3N2	A/Hong Kong/1/1968	Human	I45T (HA2)	-	X-RAY DIFFRACTION	2.25	<a href="https://doi.org/10.1126/science.aaz5143">10.1126/science.aaz5143</a>
6P6P	H3N2	A/Sichuan/2/1987	Human	-	-	X-RAY DIFFRACTION	2.31	<a href="https://doi.org/10.2210/pdb6P6P/pdb">10.2210/pdb6P6P/pdb</a>
5VTW	H3N2	A/Hong Kong/1/1968	Human	G225M/L226T/S228A (HA1)	6SLN	X-RAY DIFFRACTION	2.35	<a href="https://doi.org/10.1016/j.chom.2017.05.011">10.1016/j.chom.2017.05.011</a>
5VTY	H3N2	A/Hong Kong/1/1968	Human	G225Q/L226A (HA1)	3SLN	X-RAY DIFFRACTION	2.36	<a href="https://doi.org/10.1016/j.chom.2017.05.011">10.1016/j.chom.2017.05.011</a>
6MYM	H3N2	A/Philippines/2/1982	Human	-	-	X-RAY DIFFRACTION	2.45	<a href="https://doi.org/10.2210/pdb6MXU/pdb">10.2210/pdb6MXU/pdb</a>
5VTU	H3N2	A/Hong Kong/1/1968	Human	G225L/L226S (HA1)	-	X-RAY DIFFRACTION	2.45	<a href="https://doi.org/10.1016/j.chom.2017.05.011">10.1016/j.chom.2017.05.011</a>
6MZK	H3N2	A/Pennsylvania/14/2010	Human	-	-	X-RAY DIFFRACTION	2.5	<a href="https://doi.org/10.2210/pdb6MZK/pdb">10.2210/pdb6MZK/pdb</a>
5VTR	H3N2	A/Hong Kong/1/1968	Human	G225L/L226S	6SLN	X-RAY DIFFRACTION	2.5	<a href="https://doi.org/10.1016/j.chom.2017.05.011">10.1016/j.chom.2017.05.011</a>
6NHQ	H3N2	A/Hong Kong/1/1968	Human	I45M (HA2)	-	X-RAY DIFFRACTION	2.5	<a href="https://doi.org/10.1126/science.aaz5143">10.1126/science.aaz5143</a>
5K9Q	H3N2	A/Hong Kong/1/1968	Human	-	16.a.26 neutralizing antibody	X-RAY DIFFRACTION	2.503	<a href="https://doi.org/10.1016/j.cell.2016.06.043">10.1016/j.cell.2016.06.043</a>
5T6N	H3N2	A/Hong Kong/1/1968	Human	-	Arbidol	X-RAY DIFFRACTION	2.541	<a href="https://doi.org/10.1073/pnas.1617020114">10.1073/pnas.1617020114</a>
6CEX	H3N2	A/Hong Kong/1/1968	Human	-	N-Cyclohexylta urine	X-RAY DIFFRACTION	2.57	<a href="https://doi.org/10.1073/pnas.1801999115">10.1073/pnas.1801999115</a>
5THF	H3N2	A/Hong Kong/1/1968	Human	K-S insertion between Ser and Lys 157-158 (HA1)	-	X-RAY DIFFRACTION	2.59	<a href="https://doi.org/10.1016/j.celrep.2017.03.054">10.1016/j.celrep.2017.03.054</a>
5KUY	H3N2	A/Hong Kong/1/1968	Human	R122G (HA2)	HSB.2A inhibitor	X-RAY DIFFRACTION	2.598	<a href="https://doi.org/10.1038/nbt.3907">10.1038/nbt.3907</a>
1HGE	H3N2	A/Aichi/2-1/1968	Human	-	Receptor analog	X-RAY DIFFRACTION	2.6	<a href="https://doi.org/10.1021/bi00155a013">10.1021/bi00155a013</a>
5VTX	H3N2	A/Hong Kong/1/1968	Human	G225M/L226T/S228A (HA1)	-	X-RAY DIFFRACTION	2.65	<a href="https://doi.org/10.1016/j.chom.2017.05.011">10.1016/j.chom.2017.05.011</a>
1HGD	H3N2	A/Aichi/2-1/1968	Human	-	Receptor analog	X-RAY DIFFRACTION	2.7	<a href="https://doi.org/10.1021/bi00155a013">10.1021/bi00155a013</a>
1HGH	H3N2	A/Aichi/2-1/1968	Human	-	Receptor analog	X-RAY DIFFRACTION	2.7	<a href="https://doi.org/10.1021/bi00155a013">10.1021/bi00155a013</a>

1HGI	H3N2	A/Aichi/2-1/1968	Human	-	Receptor analog	X-RAY DIFFRACTION	2.7	<a href="https://doi.org/10.1021/bi00155a013">10.1021/bi00155a013</a>
1HGJ	H3N2	A/Aichi/2-1/1968	Human	-	Receptor analog	X-RAY DIFFRACTION	2.7	<a href="https://doi.org/10.1021/bi00155a013">10.1021/bi00155a013</a>
5KAN	H3N2	A/Hong Kong/1/1968	Human	-	16.g.07 neutralizing antibody	X-RAY DIFFRACTION	2.785	<a href="https://doi.org/10.1016/j.cell.2016.06.043">10.1016/j.cell.2016.06.043</a>
3EYM	H3N2	A/Aichi/2-1/1968	Human	-	Inhibitor	X-RAY DIFFRACTION	2.8	<a href="https://doi.org/10.1073/pnas.0807142105">10.1073/pnas.0807142105</a>
2YPG	H3N2	A/Aichi/2/1968	Human	-	LSTc	X-RAY DIFFRACTION	2.85	<a href="https://doi.org/10.1073/pnas.1218841110">10.1073/pnas.1218841110</a>
3HMG	H3N2	A/Aichi/2/1968	Human	-	-	X-RAY DIFFRACTION	2.9	<a href="https://doi.org/10.1016/0022-2836(90)90234-D">10.1016/0022-2836(90)90234-D</a>
1HGG	H3N2	A/Aichi/2-1/1968	Human	-	Receptor analog	X-RAY DIFFRACTION	2.9	<a href="https://doi.org/10.1021/bi00155a013">10.1021/bi00155a013</a>
6WXB	H3N2	A/Hong Kong/1/1968	Human	-	-	ELECTRON MICROSCOPY	2.9	<a href="https://doi.org/10.1107/S2059798320012474">10.1107/S2059798320012474</a>
5VTQ	H3N2	A/Hong Kong/1/1968	Human	G225L/L226S (HA1)	3SLN	X-RAY DIFFRACTION	2.95	<a href="https://doi.org/10.1016/j.chom.2017.05.011">10.1016/j.chom.2017.05.011</a>
4GMS	H3N2	A/Victoria/3/1975	Human	-	S139/1 Fab fragment	X-RAY DIFFRACTION	2.95	<a href="https://doi.org/10.1073/pnas.1212371109">10.1073/pnas.1212371109</a>
2HMG	H3N2	A/Hong Kong/19/1968	Human	-	-	X-RAY DIFFRACTION	3	<a href="https://doi.org/10.1016/0022-2836(90)90234-D">10.1016/0022-2836(90)90234-D</a>
4HMG	H3N2	A/Aichi/2/1968	Human	-	-	X-RAY DIFFRACTION	3	<a href="https://doi.org/10.1016/0022-2836(90)90234-D">10.1016/0022-2836(90)90234-D</a>
3VUN	H3N2	A/Aichi/2/1968	Human	-	Receptor	X-RAY DIFFRACTION	3	<a href="https://doi.org/10.2210/pdb3VUN/pdb">10.2210/pdb3VUN/pdb</a>
4ZCJ	H3N2	A/Hong Kong/1/1968	Human	C30 (HA1) / C47 (HA2)	-	X-RAY DIFFRACTION	3	<a href="https://doi.org/10.1128/JVI.00808-15">10.1128/JVI.00808-15</a>
1HGF	H3N2	A/Aichi/2-1/1968	Human	-	Receptor analog	X-RAY DIFFRACTION	3	<a href="https://doi.org/10.1021/bi00155a013">10.1021/bi00155a013</a>
5HMG	H3N2	A/Aichi/2/1968	Human	-	-	X-RAY DIFFRACTION	3.2	<a href="https://doi.org/10.1016/0022-2836(90)90234-D">10.1016/0022-2836(90)90234-D</a>
3ZTJ	H3N2	A/Aichi/2/1968	Human	-	Neutralizing antibody	X-RAY DIFFRACTION	3.41	<a href="https://doi.org/10.1126/science.1205669">10.1126/science.1205669</a>
1KEN	H3N2	A/Aichi/2/1968	Human	-	Neutralizing antibody	X-RAY DIFFRACTION	3.5	<a href="https://doi.org/10.1006/viro.2001.1320">10.1006/viro.2001.1320</a>
4UBD	H3N2	A/Hong Kong/1/1968	Human	-	Neutralizing antibody	X-RAY DIFFRACTION	3.5	<a href="https://doi.org/10.1038/ncomms8708">10.1038/ncomms8708</a>
5KAQ	H3N2	A/Hong Kong/1-4-MA21-1/1968	Human	-	31.a.83 Neutralizing antibody	X-RAY DIFFRACTION	3.514	<a href="https://doi.org/10.1016/j.cell.2016.06.043">10.1016/j.cell.2016.06.043</a>
6PDX	H3N2	A/Switzerland/9715293/2013	Human	-	C585 Fab fragment	X-RAY DIFFRACTION	3.99	<a href="https://doi.org/10.1128/JVI.01035-19">10.1128/JVI.01035-19</a>
3WHE	H3N2	A/Aichi/2-1/1968	Human	-	Neutralizing antibody	X-RAY DIFFRACTION	4	<a href="https://doi.org/10.1128/JVI.00420-14">10.1128/JVI.00420-14</a>
4NM8	H3N2	A/Hong Kong/1/1968	Human	-	CR8043 Neutralizing antibody	X-RAY DIFFRACTION	4.0041	<a href="https://doi.org/10.1073/pnas.1319058110">10.1073/pnas.1319058110</a>
4FQR	H3N2	A/Hong Kong/1/1968	Human	-	C05 Neutralizing antibody	X-RAY DIFFRACTION	4.1	<a href="https://doi.org/10.1038/nature11414">10.1038/nature11414</a>
4OSI	H3N2	A/Victoria/361/2011	Human	-	F045-092 Neutralizing antibody	X-RAY DIFFRACTION	6.501	<a href="https://doi.org/10.1038/ncomms4614">10.1038/ncomms4614</a>
<b>H3N8</b>								
5XRS	H3N2	A/swine/Minnesota/11/2010	Swine	-	LSTc	X-RAY DIFFRACTION	2.907	<a href="https://doi.org/10.1038/s41467-018-04704-9">10.1038/s41467-018-04704-9</a>
5XRT	H3N2	A/swine/Minnesota/11/2010	Swine	-	-	X-RAY DIFFRACTION	3.15	<a href="https://doi.org/10.1038/s41467-018-04704-9">10.1038/s41467-018-04704-9</a>
6N4F	H3N2	A/canine/IL/11613/2015	Canine	-	-	X-RAY DIFFRACTION	3.01	<a href="https://doi.org/10.1093/infdis/jiw620">10.1093/infdis/jiw620</a>
<b>H3N8</b>								
4WA1	H3N8	A/harbor seal/Massachusetts/1/2011	Seal	-	-	X-RAY DIFFRACTION	1.898	<a href="https://doi.org/10.1128/JVI.02723-14">10.1128/JVI.02723-14</a>
4WA2	H3N8	A/harbor seal/Massachusetts/1/2011	Seal	-	3SLN	X-RAY DIFFRACTION	2.5	<a href="https://doi.org/10.1128/JVI.02723-14">10.1128/JVI.02723-14</a>
1MQM	H3N8	A/duck/Ukraine/1/1963	Avian	-	LSTa	X-RAY DIFFRACTION	2.6	<a href="https://doi.org/10.1016/s0042-6822(03)00668-0">10.1016/s0042-6822(03)00668-0</a>
1MQL	H3N8	A/duck/Ukraine/1/1963	Avian	-	-	X-RAY DIFFRACTION	2.9	<a href="https://doi.org/10.1016/s0042-6822(03)00668-0">10.1016/s0042-6822(03)00668-0</a>
1MQN	H3N8	A/duck/Ukraine/1/1963	Avian	-	LSTc	X-RAY DIFFRACTION	3.2	<a href="https://doi.org/10.1016/s0042-6822(03)00668-0">10.1016/s0042-6822(03)00668-0</a>

4U00	H3N8	A/equine/Richmond/1/2007	Equine	–	–	X-RAY DIFFRACTION	1.9	<a href="https://doi.org/10.1073/pnas.1406606111">10.1073/pnas.1406606111</a>
4UNW	H3N8	A/equine/Newmarket/1993	Equine	–	–	X-RAY DIFFRACTION	2.6	<a href="https://doi.org/10.1073/pnas.1406606111">10.1073/pnas.1406606111</a>
4U03	H3N8	A/equine/Richmond/1/2007	Equine	S30T (HA1)	–	X-RAY DIFFRACTION	2.87	<a href="https://doi.org/10.1073/pnas.1406606111">10.1073/pnas.1406606111</a>
4UNZ	H3N8	A/equine/Newmarket/1993	Equine	–	6S04-Sialyl Lewis X	X-RAY DIFFRACTION	2.9	<a href="https://doi.org/10.1073/pnas.1406606111">10.1073/pnas.1406606111</a>
4U01	H3N8	A/equine/Richmond/1/2007	Equine	–	3SLN	X-RAY DIFFRACTION	3	<a href="https://doi.org/10.1073/pnas.1406606111">10.1073/pnas.1406606111</a>
4UNX	H3N8	A/eq/Newmarket/93	Equine	–	3SLN	X-RAY DIFFRACTION	3.2	<a href="https://doi.org/10.1073/pnas.1406606111">10.1073/pnas.1406606111</a>
<b>H4N6</b>								
4U05	H3N8	A/canine/Colorado/17864/2006	Canine	–	3SLN	X-RAY DIFFRACTION	2.7	<a href="https://doi.org/10.1073/pnas.1406606111">10.1073/pnas.1406606111</a>
4UNY	H3N8	A/canine/Colorado/17864/2006	Canine	–	6S04-3SLN	X-RAY DIFFRACTION	2.9	<a href="https://doi.org/10.1073/pnas.1406606111">10.1073/pnas.1406606111</a>
4U07	H3N8	A/canine/Colorado/17864/2006	Canine	–	6S04 Sialyl Lewis X	X-RAY DIFFRACTION	3	<a href="https://doi.org/10.1073/pnas.1406606111">10.1073/pnas.1406606111</a>
<b>H5N1</b>								
5XL8	H4N6	A/Duck/Czechoslovakia/1956	Avian	G228S (HA1)	–	X-RAY DIFFRACTION	2.001	<a href="https://doi.org/10.1016/j.celrep.2017.07.028">10.1016/j.celrep.2017.07.028</a>
5XL9	H4N6	A/Duck/Czechoslovakia/1956	Avian	G228S (HA1)	LSTa	X-RAY DIFFRACTION	2.39	<a href="https://doi.org/10.1016/j.celrep.2017.07.028">10.1016/j.celrep.2017.07.028</a>
5Y2M	H4N6	A/Duck/Czechoslovakia/1956	Avian	–	AF4H1K1 antibody	X-RAY DIFFRACTION	3.8	<a href="https://doi.org/10.2210/pdb5Y2L/pdb">10.2210/pdb5Y2L/pdb</a>
6V44	H4N6	A/swine/Missouri/A01727926/2015	Swine	–	–	X-RAY DIFFRACTION	2.2	<a href="https://doi.org/10.1016/j.heliyon.2020.e04068">10.1016/j.heliyon.2020.e04068</a>
5XL2	H4N6	A/swine/Ontario/01911-1/99	Swine	–	–	X-RAY DIFFRACTION	2.3	<a href="https://doi.org/10.1016/j.celrep.2017.07.028">10.1016/j.celrep.2017.07.028</a>
<b>H5N1</b>								
6CF5	H5N1	A/Viet Nam/1203/2004	Human	–	N-Cyclohexylta urine	X-RAY DIFFRACTION	2.04	<a href="https://doi.org/10.1073/pnas.1801999115">10.1073/pnas.1801999115</a>
4KDO	H5N1	A/Vietnam/1203/2004 (high pathogenic)	Human	–	LSTc	X-RAY DIFFRACTION	2.396	<a href="https://doi.org/10.1007/s13238-013-3906-z">10.1007/s13238-013-3906-z</a>
3ZNM	H5N1	A/Vietnam/1194/2004	Human	–	Sialyl-Lewis X	X-RAY DIFFRACTION	2.4	<a href="https://doi.org/10.1016/j.virusres.2013.08.007">10.1016/j.virusres.2013.08.007</a>
4KDN	H5N1	A/Vietnam/1203/2004 (high pathogenic)	Human	–	LSTa	X-RAY DIFFRACTION	2.483	<a href="https://doi.org/10.1007/s13238-013-3906-z">10.1007/s13238-013-3906-z</a>
4KDM	H5N1	A/Vietnam/1203/2004 (high pathogenic)	Human	N158D/N224K/Q226L/T318I (HA1)	–	X-RAY DIFFRACTION	2.496	<a href="https://doi.org/10.1007/s13238-013-3906-z">10.1007/s13238-013-3906-z</a>
3ZNL	H5N1	A/Vietnam/1194/2004	Human	–	Sulfated Lewis X	X-RAY DIFFRACTION	2.5	<a href="https://doi.org/10.1016/j.virusres.2013.08.007">10.1016/j.virusres.2013.08.007</a>
4KTH	H5N1	A/Hubei/1/2010	Human	–	–	X-RAY DIFFRACTION	2.6	<a href="https://doi.org/10.1371/journal.pone.0075209">10.1371/journal.pone.0075209</a>
4KDQ	H5N1	A/Xinjiang/1/2006	Human	–	–	X-RAY DIFFRACTION	2.604	<a href="https://doi.org/10.1007/s13238-013-3906-z">10.1007/s13238-013-3906-z</a>
3ZNK	H5N1	A/Vietnam/1194/2004	Human	–	Su-3SLN	X-RAY DIFFRACTION	2.71	<a href="https://doi.org/10.1016/j.virusres.2013.08.007">10.1016/j.virusres.2013.08.007</a>
2IBX	H5N1	A/Vietnam/1194/2004	Human	–	–	X-RAY DIFFRACTION	2.8	<a href="https://doi.org/10.1038/nature05264">10.1038/nature05264</a>
2FKO	H5N1	A/Vietnam/1203/2004 (high pathogenic)	Human	–	–	X-RAY DIFFRACTION	2.95	<a href="https://doi.org/10.1126/science.1124513">10.1126/science.1124513</a>
4N5Z	H5N1	A/Viet Nam/1203/2004	Human	N158D/N224K/Q226L/T318I (HA1)	–	X-RAY DIFFRACTION	2.9537	<a href="https://doi.org/10.1128/JVI.02690-13">10.1128/JVI.02690-13</a>
6E7G	H5N1	A/Viet Nam/1203/2004	Human	–	–	X-RAY DIFFRACTION	3.094	<a href="https://doi.org/10.1016/j.celrep.2019.05.048">10.1016/j.celrep.2019.05.048</a>
4N5Y	H5N1	A/Viet Nam/1203/2004	Human	Y161A (HA1)	–	X-RAY DIFFRACTION	3.1615	<a href="https://doi.org/10.1128/JVI.02690-13">10.1128/JVI.02690-13</a>
3FKU	H5N1	A/Viet Nam/1203/2004	Human	–	F10 neutralizing antibody	X-RAY DIFFRACTION	3.2	<a href="https://doi.org/10.1038/nsmb.1566">10.1038/nsmb.1566</a>
6E7H	H5N1	A/Viet Nam/1203/2004	Human	Y161A (HA1)	3'-GcLN	X-RAY DIFFRACTION	3.3	<a href="https://doi.org/10.1016/j.celrep.2019.05.048">10.1016/j.celrep.2019.05.048</a>
4MHH	H5N1	A/Viet Nam/1203/2004	Human	–	H5M9 Fab fragment	X-RAY DIFFRACTION	3.564	<a href="https://doi.org/10.1128/JVI.01577-13">10.1128/JVI.01577-13</a>
5JW4	H5N1	A/Vietnam/1194/2004	Human	N182K (HA1)	MEDI8852 Fab Fragment	X-RAY DIFFRACTION	3.7	<a href="https://doi.org/10.1016/j.cell.2016.05.073">10.1016/j.cell.2016.05.073</a>

6B3M	H5N1	A/Vietnam/1203/2004 (high pathogenic)	Human	-	70-1F02 antibody	X-RAY DIFFRACTION	3.92	<a href="https://doi.org/10.1128/JVI.00949-18">10.1128/JVI.00949-18</a>
<b>H5N1</b>								
4CQY	H5N1	A/turkey/Turkey/1/2005	Avian	$\Delta$ 133/155T (HA1)	LSTa	X-RAY DIFFRACTION	2.05	<a href="https://doi.org/10.1016/j.virol.2014.03.008">10.1016/j.virol.2014.03.008</a>
4BH1	H5N1	A/turkey/Turkey/1/2005	Avian	-	3SLN	X-RAY DIFFRACTION	2.15	<a href="https://doi.org/10.1038/nature12144">10.1038/nature12144</a>
6VMZ	H5N1	A/chicken/Vietnam/4/2003	Avian	-	CBS1117	X-RAY DIFFRACTION	2.2	<a href="https://doi.org/10.26508/lsa.202000724">10.26508/lsa.202000724</a>
4CQW	H5N1	A/turkey/Turkey/1/2005	Avian	$\Delta$ 133/155T (HA1)	3SLN	X-RAY DIFFRACTION	2.3	<a href="https://doi.org/10.1016/j.virol.2014.03.008">10.1016/j.virol.2014.03.008</a>
4CQX	H5N1	A/turkey/Turkey/1/2005	Avian	$\Delta$ 133/155T (HA1)	6SLN	X-RAY DIFFRACTION	2.3	<a href="https://doi.org/10.1016/j.virol.2014.03.008">10.1016/j.virol.2014.03.008</a>
4JUN	H5N1	A/duck/Guangxi/2396/2004	Avian	-	-	X-RAY DIFFRACTION	2.3398	<a href="https://doi.org/10.2210/pdb4JUN/pdb">10.2210/pdb4JUN/pdb</a>
4BH0	H5N1	A/turkey/Turkey/1/2005	Avian	-	6SLN	X-RAY DIFFRACTION	2.36	<a href="https://doi.org/10.1038/nature12144">10.1038/nature12144</a>
3S11	H5N1	A/Goose/Guangdong/1/1996	Avian	-	-	X-RAY DIFFRACTION	2.5	<a href="https://doi.org/10.1371/journal.ppat.1002398">10.1371/journal.ppat.1002398</a>
4MHI	H5N1	A/Goose/Guangdong/1/96	Avian	-	-	X-RAY DIFFRACTION	2.595	<a href="https://doi.org/10.1128/JVI.01577-13">10.1128/JVI.01577-13</a>
5E2Y	H5N1	A/duck/Egypt/101855S/2010	Avian	Q226L (HA1)	-	X-RAY DIFFRACTION	2.6	<a href="https://doi.org/10.1016/j.celrep.2015.10.027">10.1016/j.celrep.2015.10.027</a>
5E2Z	H5N1	A/duck/Egypt/101855S/2010	Avian	Q226L (HA1)	LSTa	X-RAY DIFFRACTION	2.624	<a href="https://doi.org/10.1016/j.celrep.2015.10.027">10.1016/j.celrep.2015.10.027</a>
4BGZ	H5N1	A/turkey/Turkey/1/2005	Avian	-	-	X-RAY DIFFRACTION	2.68	<a href="https://doi.org/10.1038/nature12144">10.1038/nature12144</a>
5E30	H5N1	A/duck/Egypt/101855S/2010	Avian	Q226L (HA1)	LSTc	X-RAY DIFFRACTION	2.7	<a href="https://doi.org/10.1016/j.celrep.2015.10.027">10.1016/j.celrep.2015.10.027</a>
4JUL	H5N1	A/duck/Laos/3295/2006	Avian	-	-	X-RAY DIFFRACTION	2.7927	<a href="https://doi.org/10.2210/pdb4JUL/pdb">10.2210/pdb4JUL/pdb</a>
4CQV	H5N1	A/turkey/Turkey/1/2005	Avian	$\Delta$ 133/155T (HA1)	-	X-RAY DIFFRACTION	2.86	<a href="https://doi.org/10.1016/j.virol.2014.03.008">10.1016/j.virol.2014.03.008</a>
4MHJ	H5N1	A/Goose/Guangdong/1/96	Avian	-	H5M9 Fab fragment	X-RAY DIFFRACTION	6.98	<a href="https://doi.org/10.1128/JVI.01577-13">10.1128/JVI.01577-13</a>
<b>H5N2</b>								
5YKC	H5N2	A/chicken/Taiwan/0502/2012	Avian	-	-	X-RAY DIFFRACTION	2.823	<a href="https://doi.org/10.2210/pdb5YT9/pdb">10.2210/pdb5YT9/pdb</a>
<b>H5N6</b>								
5HU8	H5N6	A/Sichuan/26221/2014	Human	-	-	X-RAY DIFFRACTION	2.45	<a href="https://doi.org/10.1128/JVI.00180-16">10.1128/JVI.00180-16</a>
<b>H5N8</b>								
5HUF	H5N8	A/gyrfalcon/Washington/41088-6/2014	Avian	-	-	X-RAY DIFFRACTION	2.81	<a href="https://doi.org/10.1128/JVI.00180-16">10.1128/JVI.00180-16</a>
<b>H6N1</b>								
5TOB	H6N1	A/Taiwan/2/2013	Human	G225D (HA1)	6SLN	X-RAY DIFFRACTION	2.001	<a href="https://doi.org/10.15252/emmm.201707726">10.15252/emmm.201707726</a>
5TOE	H6N1	A/Taiwan/2/2013	Human	G225D (HA1)	LSTa	X-RAY DIFFRACTION	2.089	<a href="https://doi.org/10.15252/emmm.201707726">10.15252/emmm.201707726</a>
5TO8	H6N1	A/Taiwan/2/2013	Human	G225D (HA1)	-	X-RAY DIFFRACTION	2.1895	<a href="https://doi.org/10.15252/emmm.201707726">10.15252/emmm.201707726</a>
4XKG	H6N1	A/Taiwan/2/2013	Human	-	6SLN	X-RAY DIFFRACTION	2.25	<a href="https://doi.org/10.1016/j.chom.2015.02.005">10.1016/j.chom.2015.02.005</a>
4XKE	H6N1	A/Taiwan/2/2013	Human	-	3SLN	X-RAY DIFFRACTION	2.356	<a href="https://doi.org/10.1016/j.chom.2015.02.005">10.1016/j.chom.2015.02.005</a>
4WST	H6N1	A/Taiwan/1/2013	Human	-	-	X-RAY DIFFRACTION	2.4	<a href="https://doi.org/10.1128/JVI.03456-14">10.1128/JVI.03456-14</a>
4XKF	H6N1	A/Taiwan/2/2013	Human	-	LSTa	X-RAY DIFFRACTION	2.446	<a href="https://doi.org/10.1016/j.chom.2015.02.005">10.1016/j.chom.2015.02.005</a>



4XKD	H6N1	A/Taiwan/2/2013	Human	-	-	X-RAY DIFFRACTION	2.482	<a href="https://doi.org/10.1016/j.chom.2015.02.005">10.1016/j.chom.2015.02.005</a>
4WSU	H6N1	A/Taiwan/1/2013	Human	-	3SLN	X-RAY DIFFRACTION	2.7	<a href="https://doi.org/10.1128/JVI.03456-14">10.1128/JVI.03456-14</a>
5TOD	H6N1	A/Taiwan/2/2013	Human	G225D (HA1)	3SLN	X-RAY DIFFRACTION	2.864	<a href="https://doi.org/10.15252/emmm.201707726">10.15252/emmm.201707726</a>
4WSV	H6N1	A/Taiwan/1/2013	Human	-	6SLN	X-RAY DIFFRACTION	3.1	<a href="https://doi.org/10.1128/JVI.03456-14">10.1128/JVI.03456-14</a>
<b>H6N2</b>								
4YYO	H6N1	A/chicken/Taiwan/A2837/2013	Avian	-	-		2.59	<a href="https://doi.org/10.2210/pdb4YY0/pdb">10.2210/pdb4YY0/pdb</a>
<b>H6N6</b>								
4WSR	H6N2	A/chicken/New York/14677-13/1998	Avian	-	-	X-RAY DIFFRACTION	2.5	<a href="https://doi.org/10.1128/JVI.03456-14">10.1128/JVI.03456-14</a>
4WSS	H6N2	A/chicken/New York/14677-13/1998	Avian	-	LSTa	X-RAY DIFFRACTION	2.8	<a href="https://doi.org/10.1128/JVI.03456-14">10.1128/JVI.03456-14</a>
<b>H6N6</b>								
5BNY	H6N6	A/Chicken/Guangdong/S1311/2010	Avian	-	-	X-RAY DIFFRACTION	2.66	<a href="https://doi.org/10.1371/journal.pone.0134576">10.1371/journal.pone.0134576</a>
5BQY	H6N6	A/Chicken/Guangdong/S1311/2010	Avian	-	LSTa	X-RAY DIFFRACTION	2.78	<a href="https://doi.org/10.1371/journal.pone.0134576">10.1371/journal.pone.0134576</a>
5BQZ	H6N6	A/Chicken/Guangdong/S1311/2010	Avian	-	LSTc	X-RAY DIFFRACTION	2.89	<a href="https://doi.org/10.1371/journal.pone.0134576">10.1371/journal.pone.0134576</a>
<b>H7N2</b>								
3M5G	H7N2	A/New York/107/2003	Human	-	-	X-RAY DIFFRACTION	2.6	<a href="https://doi.org/10.1371/journal.ppat.1001081">10.1371/journal.ppat.1001081</a>
3M5J	H7N2	A/New York/107/2003	Human	-	LSTb	X-RAY DIFFRACTION	2.6	<a href="https://doi.org/10.1371/journal.ppat.1001081">10.1371/journal.ppat.1001081</a>
3M5H	H7N2	A/New York/107/2003	Human	-	3SLN	X-RAY DIFFRACTION	2.7	<a href="https://doi.org/10.1371/journal.ppat.1001081">10.1371/journal.ppat.1001081</a>
3M5I	H7N2	A/New York/107/2003	Human	-	6SLN	X-RAY DIFFRACTION	3	<a href="https://doi.org/10.1371/journal.ppat.1001081">10.1371/journal.ppat.1001081</a>
6MLM	H7N2	A/New York/107/2003	Human	-	H7.5 antibody	ELECTRON MICROSCOPY	3.5	<a href="https://doi.org/10.1371/journal.pbio.3000139">10.1371/journal.pbio.3000139</a>
<b>H7N7</b>								
4DJ6	H7N7	A/Netherlands/219/2003	Human	-	-	X-RAY DIFFRACTION	2.61	<a href="https://doi.org/10.1128/JVI.00281-12">10.1128/JVI.00281-12</a>
4DJ8	H7N7	A/Netherlands/219/2003	Human	-	6SLN	X-RAY DIFFRACTION	2.8	<a href="https://doi.org/10.1128/JVI.00281-12">10.1128/JVI.00281-12</a>
4DJ7	H7N7	A/Netherlands/219/2003	Human	-	3SLN	X-RAY DIFFRACTION	2.81	<a href="https://doi.org/10.1128/JVI.00281-12">10.1128/JVI.00281-12</a>
4FQV	H7N7	A/Netherlands/219/2003	Human	-	CR9114 antibody	X-RAY DIFFRACTION	5.75	<a href="https://doi.org/10.1126/science.1222908">10.1126/science.1222908</a>
<b>H7N9</b>								
4LN6	H7N9	A/Shanghai/02/2013	Human	-	-	X-RAY DIFFRACTION	2.12	<a href="https://doi.org/10.1128/JVI.01854-13">10.1128/JVI.01854-13</a>
5T6S	H7N9	A/Shanghai/2/2013	Human	-	Arbidol	X-RAY DIFFRACTION	2.36	<a href="https://doi.org/10.1073/pnas.1617020114">10.1073/pnas.1617020114</a>
6IDD	H7N9	A/Shanghai/1/2013	Human	S138A/G186V/T221P/Q226L (HA1)	-	X-RAY DIFFRACTION	2.38	<a href="https://doi.org/10.1016/j.celrep.2019.10.047">10.1016/j.celrep.2019.10.047</a>
4LN8	H7N9	A/shanghai/2/2013	Human	-	LSTb	X-RAY DIFFRACTION	2.5	<a href="https://doi.org/10.1128/JVI.01854-13">10.1128/JVI.01854-13</a>

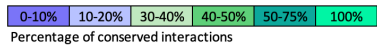
6FYU	H7N9	A/shanghai/2/2013	Human	-	SD36	X-RAY DIFFRACTION	2.643	<a href="https://doi.org/10.1126/science.aag0620">10.1126/science.aag0620</a>
4LN3	H7N9	A/Shanghai/1/2013	Human	-	-	X-RAY DIFFRACTION	2.65	<a href="https://doi.org/10.1128/JVI.01854-13">10.1128/JVI.01854-13</a>
6D7U	H7N9	A/Guangdong/175F003/2016	Human	-	-	X-RAY DIFFRACTION	2.7	<a href="https://doi.org/10.1128/JVI.00375-18">10.1128/JVI.00375-18</a>
6ID2	H7N9	A/Anhui/1/2013	Human	P221T (HA1)	-	X-RAY DIFFRACTION	2.705	<a href="https://doi.org/10.1016/j.celrep.2019.10.047">10.1016/j.celrep.2019.10.047</a>
6D7C	H7N9	A/Hong Kong/61/2016	Human	-	-	X-RAY DIFFRACTION	2.95	<a href="https://doi.org/10.1128/JVI.00375-18">10.1128/JVI.00375-18</a>
6D8B	H7N9	A/Hong Kong/61/2016	Human	-	-	X-RAY DIFFRACTION	2.95	<a href="https://doi.org/10.1128/JVI.00375-18">10.1128/JVI.00375-18</a>
4LCX	H7N9	A/Shanghai/1/2013	Human	-	-	X-RAY DIFFRACTION	3.094	<a href="https://doi.org/10.1126/science.1242917">10.1126/science.1242917</a>
4LN4	H7N9	A/Shanghai/02/2013	Human	-	LSTb	X-RAY DIFFRACTION	3.1	<a href="https://doi.org/10.1128/JVI.01854-13">10.1128/JVI.01854-13</a>
6I18	H7N9	A/Anhui/1/2013	Human	-	L4B-18 neutralizing antibody	X-RAY DIFFRACTION	3.32	<a href="https://doi.org/10.1038/s41564-018-0303-Z">10.1038/s41564-018-0303-Z</a>
6I19	H7N9	A/Anhui/1/2013	Human	-	L3A-44 neutralizing antibody	X-RAY DIFFRACTION	3.5	<a href="https://doi.org/10.1038/s41564-018-0303-Z">10.1038/s41564-018-0303-Z</a>
6D8D	H7N9	A/Hong Kong/125/2017	Human	-	LSTb	X-RAY DIFFRACTION	3.55	<a href="https://doi.org/10.1128/JVI.00375-18">10.1128/JVI.00375-18</a>
<b>H8N4</b>								
6V46	H8N4	A/turkey/Ontario/6118/1968	Avian	-	-	X-RAY DIFFRACTION	2.3	<a href="https://doi.org/10.1016/j.heliyon.2020.e04068">10.1016/j.heliyon.2020.e04068</a>
<b>H10N2</b>								
4CYV	H10N2	A/mallard/Sweden/51/2002	Avian	-	-	X-RAY DIFFRACTION	2.3	<a href="https://doi.org/10.1038/nature13443">10.1038/nature13443</a>
4CYZ	H10N2	A/mallard/Sweden/51/2002	Avian	-	LSTa	X-RAY DIFFRACTION	2.4	<a href="https://doi.org/10.1038/nature13443">10.1038/nature13443</a>
4CYW	H10N2	A/mallard/Sweden/51/2002	Avian	-	6SLN	X-RAY DIFFRACTION	2.6	<a href="https://doi.org/10.1038/nature13443">10.1038/nature13443</a>
4CZ0	H10N2	A/mallard/Sweden/51/2002	Avian	-	Su-3SLN	X-RAY DIFFRACTION	3.2	<a href="https://doi.org/10.1038/nature13443">10.1038/nature13443</a>
<b>H10N7</b>								
6TVA	H10N7	A/harbour seal/Netherlands/PV14- 221_TS/2015	Seal	-	3SLN	X-RAY DIFFRACTION	1.65	<a href="https://doi.org/10.1016/j.chom.2020.08.011">10.1016/j.chom.2020.08.011</a>
6TVB	H10N7	A/harbour seal/Netherlands/PV14- 221_TS/2015	Seal	-	6SLN	X-RAY DIFFRACTION	1.65	<a href="https://doi.org/10.1016/j.chom.2020.08.011">10.1016/j.chom.2020.08.011</a>
6TVC	H10N7	A/harbour seal/Netherlands/PV14- 221_TS/2015	Seal	-	-	X-RAY DIFFRACTION	1.84	<a href="https://doi.org/10.1016/j.chom.2020.08.011">10.1016/j.chom.2020.08.011</a>
6TVS	H10N7	A/harbor seal/S1047_14_L/Germany/201 4	Seal	G226L (HA1)	3SLN	X-RAY DIFFRACTION	1.9	<a href="https://doi.org/10.1016/j.chom.2020.08.011">10.1016/j.chom.2020.08.011</a>
6TWS	H10N7	A/harbor seal/S1047_14_L/Germany/201 4	Seal	G226L/G228S	-	X-RAY DIFFRACTION	2	<a href="https://doi.org/10.1016/j.chom.2020.08.011">10.1016/j.chom.2020.08.011</a>
6TVT	H10N7	A/harbor seal/S1047_14_L/Germany/201 4	Seal	G226L/Del228 (HA1)	6SLN	X-RAY DIFFRACTION	2.2	<a href="https://doi.org/10.1016/j.chom.2020.08.011">10.1016/j.chom.2020.08.011</a>
6TWI	H10N7	A/harbor seal/S1047_14_L/Germany/201 4	Seal	G226L/G228S	3SLN	X-RAY DIFFRACTION	2.27	<a href="https://doi.org/10.1016/j.chom.2020.08.011">10.1016/j.chom.2020.08.011</a>
6TJW	H10N7	A/harbor seal/S1047_14_L/Germany/201 4	Seal	G226L/Del228 (HA1)	-	X-RAY DIFFRACTION	2.31	<a href="https://doi.org/10.1016/j.chom.2020.08.011">10.1016/j.chom.2020.08.011</a>
6TXO	H10N7	A/harbor seal/S1047_14_L/Germany/201 4	Seal	G226L/Del228 (HA1)	3SLN	X-RAY DIFFRACTION	2.4	<a href="https://doi.org/10.1016/j.chom.2020.08.011">10.1016/j.chom.2020.08.011</a>
6TWW	H10N7	A/harbor seal/S1047_14_L/Germany/201 4	Seal	G226L (HA1)	6SLN	X-RAY DIFFRACTION	2.55	<a href="https://doi.org/10.1016/j.chom.2020.08.011">10.1016/j.chom.2020.08.011</a>
6TVF	H10N7	A/harbor seal/S1047_14_L/Germany/201 4	Seal	-	6SLN	X-RAY DIFFRACTION	2.6	<a href="https://doi.org/10.1016/j.chom.2020.08.011">10.1016/j.chom.2020.08.011</a>

6TVR	H10N7	A/harbor seal/S1047_14_L/Germany/2014	Seal	G226L (HA1)	-	X-RAY DIFFRACTION	2.63	<a href="https://doi.org/10.1016/j.chom.2020.08.011">10.1016/j.chom.2020.08.011</a>
6TWH	H10N7	A/harbor seal/S1047_14_L/Germany/2014	Seal	G226L/G228S	-	X-RAY DIFFRACTION	2.68	<a href="https://doi.org/10.1016/j.chom.2020.08.011">10.1016/j.chom.2020.08.011</a>
6TVD	H10N7	A/harbor seal/S1047_14_L/Germany/2014	Seal	-	3SLN	X-RAY DIFFRACTION	2.7	<a href="https://doi.org/10.1016/j.chom.2020.08.011">10.1016/j.chom.2020.08.011</a>
6TJY	H10N7	A/harbor seal/S1047_14_L/Germany/2014	Seal	-	-	X-RAY DIFFRACTION	2.82	<a href="https://doi.org/10.1016/j.chom.2020.08.011">10.1016/j.chom.2020.08.011</a>
6TY1	H10N7	A/harbor seal/S1047_14_L/Germany/2014	Seal	G226L/G228S	6SLN	X-RAY DIFFRACTION	3.2	<a href="https://doi.org/10.1016/j.chom.2020.08.011">10.1016/j.chom.2020.08.011</a>
<b>H10N8</b>								
4D00	H10N7	A/Chicken/Germany/n/1949	Avian	-	6SLN	X-RAY DIFFRACTION	2.5	<a href="https://doi.org/10.1038/nature13443">10.1038/nature13443</a>
4WSW	H10N7	A/green-winged teal/Texas/Y171/2006	Avian	-	-	X-RAY DIFFRACTION	2.8	<a href="https://doi.org/10.1128/JVI.03456-14">10.1128/JVI.03456-14</a>
<b>H10N8</b>								
5TH1	H10N8	A/Jiangxi-Donghu/346/2013	Human	K158aA/Q226L/G228S (HA1)	6SLNLN	X-RAY DIFFRACTION	2.191	<a href="https://doi.org/10.1016/j.celrep.2017.03.054">10.1016/j.celrep.2017.03.054</a>
5TH0	H10N8	A/Jiangxi-Donghu/346/2013	Human	K158aA/D193T/Q226L/G228S (HA1)	3SLN	X-RAY DIFFRACTION	2.25	<a href="https://doi.org/10.1016/j.celrep.2017.03.054">10.1016/j.celrep.2017.03.054</a>
5TGO	H10N8	A/Jiangxi-Donghu/346/2013	Human	K158aA/D193T/Q226L/G228S (HA1)	-	X-RAY DIFFRACTION	2.35	<a href="https://doi.org/10.1016/j.celrep.2017.03.054">10.1016/j.celrep.2017.03.054</a>
5TGU	H10N8	A/Jiangxi-Donghu/346/2013	Human	K158aA/D193T/Q226L/G228S (HA1)	6SLNLN	X-RAY DIFFRACTION	2.35	<a href="https://doi.org/10.1016/j.celrep.2017.03.054">10.1016/j.celrep.2017.03.054</a>
4QY2	H10N8	A/Jiangxi-Donghu/IPB13/2013	Human	-	LSTc	X-RAY DIFFRACTION	2.399	<a href="https://doi.org/10.1038/ncomms6600">10.1038/ncomms6600</a>
5THB	H10N8	A/Jiangxi-Donghu/346/2013	Human	D193T/Q226L/G228S (HA1)	-	X-RAY DIFFRACTION	2.41	<a href="https://doi.org/10.1016/j.celrep.2017.03.054">10.1016/j.celrep.2017.03.054</a>
4QY0	H10N8	A/Jiangxi-Donghu/IPB13/2013	Human	-	-	X-RAY DIFFRACTION	2.47	<a href="https://doi.org/10.1038/ncomms6600">10.1038/ncomms6600</a>
4XQ5	H10N8	A/Jiangxi-Donghu/IPB13/2013	Human	-	-	X-RAY DIFFRACTION	2.592	<a href="https://doi.org/10.1016/j.chom.2015.02.006">10.1016/j.chom.2015.02.006</a>
4QY1	H10N8	A/Jiangxi-Donghu/IPB13/2013	Human	-	3SLN	X-RAY DIFFRACTION	2.594	<a href="https://doi.org/10.1038/ncomms6600">10.1038/ncomms6600</a>
4WSX	H10N8	Jiangxi-Donghu/346/2013	Human	-	-	X-RAY DIFFRACTION	2.7	<a href="https://doi.org/10.1128/JVI.03456-14">10.1128/JVI.03456-14</a>
5THC	H10N8	A/Jiangxi-Donghu/346/2013	Human	D193T/Q226L/G228S (HA1)	6SLNLN	X-RAY DIFFRACTION	2.792	<a href="https://doi.org/10.1016/j.celrep.2017.03.054">10.1016/j.celrep.2017.03.054</a>
4XQ0	H10N8	A/Jiangxi-Donghu/346/2013	Human	-	6SLNLN	X-RAY DIFFRACTION	2.85	<a href="https://doi.org/10.1016/j.chom.2015.02.006">10.1016/j.chom.2015.02.006</a>
5TGV	H10N8	A/Jiangxi-Donghu/346/2013	Human	K158aA/D193T/Q226L/G228S (HA1)	3SLN	X-RAY DIFFRACTION	2.974	<a href="https://doi.org/10.1016/j.celrep.2017.03.054">10.1016/j.celrep.2017.03.054</a>
4XQU	H10N8	A/Jiangxi-Donghu/IPB13/2013	Human	-	3SLN	X-RAY DIFFRACTION	3.25	<a href="https://doi.org/10.1016/j.chom.2015.02.006">10.1016/j.chom.2015.02.006</a>
<b>H11N9</b>								
6V47	H11N9	A/duck/Memphis/546/1974	Avian	-	-	X-RAY DIFFRACTION	2.8	<a href="https://doi.org/10.1016/j.heliyon.2020.e04068">10.1016/j.heliyon.2020.e04068</a>
<b>H13N6</b>								
4KPQ	H13N6	A/Gull/Maryland/704/1977	Avian	-	-	X-RAY DIFFRACTION	2.502	<a href="https://doi.org/10.1128/JVI.00235-13">10.1128/JVI.00235-13</a>
4KPS	H13N6	A/Gull/Maryland/704/1977	Avian	-	-	X-RAY DIFFRACTION	2.587	<a href="https://doi.org/10.1128/JVI.00235-13">10.1128/JVI.00235-13</a>
<b>H14N5</b>								
6V48	H14N5	A/mallard/Gurjev/263/1982	Avian	-	-	X-RAY DIFFRACTION	3	<a href="https://doi.org/10.1016/j.heliyon.2020.e04068">10.1016/j.heliyon.2020.e04068</a>

<b>H15N9</b>								
6V49	H15N9	A/wedge-tailed shearwater/Western Australia/2576/1979	Avian	-	-	X-RAY DIFFRACTION	2.5	<a href="https://doi.org/10.1016/j.heliyon.2020.e04068">10.1016/j.heliyon.2020.e04068</a>
<b>H16N3</b>								
4F23	H16N3	A/black-headed gull/Sweden/2/99	Avian	-	-	X-RAY DIFFRACTION	1.7	<a href="https://doi.org/10.1128/JVI.01606-12">10.1128/JVI.01606-12</a>
4FIU	H16N3	A/black-headed gull/Sweden/2/99	Avian	V327G (HA1)	-	X-RAY DIFFRACTION	1.999	<a href="https://doi.org/10.1128/JVI.01606-12">10.1128/JVI.01606-12</a>
<b>H17N10</b>								
4H32	H17N10	A/little yellow-shouldered bat/Guatemala/060/2010	Bat	-	-	X-RAY DIFFRACTION	2.7	<a href="https://doi.org/10.1016/j.celrep.2013.01.025">10.1016/j.celrep.2013.01.025</a>
<b>H18N11</b>								
4K3X	H18N11	A/flat-faced bat/Peru/033/2010	Bat	-	-	X-RAY DIFFRACTION	2.149	<a href="https://doi.org/10.1371/journal.ppat.1003657">10.1371/journal.ppat.1003657</a>
4MC5	H18N11	A/flat-faced bat/Peru/033/2010	Bat	-	-	X-RAY DIFFRACTION	2.238	<a href="https://doi.org/10.1371/journal.ppat.1003657">10.1371/journal.ppat.1003657</a>
<b>Intermediary and Post-Fusion</b>								
7K39	H3N2	A/Hong Kong/1/1968 (pH 5.2 - A)	Human	-	Antibody	ELECTRON MICROSCOPY	3	<a href="https://doi.org/10.1371/journal.ppat.1009062">10.1371/journal.ppat.1009062</a>
7K37	H3N2	A/Hong Kong/1/1968 (pH 7.8)	Human	-	Antibody	ELECTRON MICROSCOPY	2.8	<a href="https://doi.org/10.1371/journal.ppat.1009062">10.1371/journal.ppat.1009062</a>
7K3A	H3N2	A/Hong Kong/1/1968 (pH 5.2 - B)	Human	-	Antibody	ELECTRON MICROSCOPY	4.2	<a href="https://doi.org/10.1371/journal.ppat.1009062">10.1371/journal.ppat.1009062</a>
7K3B	H3N2	A/Hong Kong/1/1968 (pH 5.2 - C)	Human	-	Antibody	ELECTRON MICROSCOPY	3.4	<a href="https://doi.org/10.1371/journal.ppat.1009062">10.1371/journal.ppat.1009062</a>
6Y5G	H3N2	A/Aichi/2-1/1968 (pH 8)	Human	-	-	ELECTRON MICROSCOPY	3	<a href="https://doi.org/10.1038/s41586-020-2333-6">10.1038/s41586-020-2333-6</a>
6Y5H	H3N2	A/Aichi/2-1/1968 (pH 5 - State I)	Human	-	-	ELECTRON MICROSCOPY	3	<a href="https://doi.org/10.1038/s41586-020-2333-6">10.1038/s41586-020-2333-6</a>
6Y5I	H3N2	A/Aichi/2-1/1968 (pH 5 - State II)	Human	-	-	ELECTRON MICROSCOPY	5.5	<a href="https://doi.org/10.1038/s41586-020-2333-6">10.1038/s41586-020-2333-6</a>
6Y5J	H3N2	A/Aichi/2-1/1968 (pH 5 - State III)	Human	-	-	ELECTRON MICROSCOPY	5.6	<a href="https://doi.org/10.1038/s41586-020-2333-6">10.1038/s41586-020-2333-6</a>
6Y5K	H3N2	A/Aichi/2-1/1968 (pH 5 - State IV)	Human	-	-	ELECTRON MICROSCOPY	4.2	<a href="https://doi.org/10.1038/s41586-020-2333-6">10.1038/s41586-020-2333-6</a>
6Y5L	H3N2	A/Aichi/2-1/1968 (pH 5 - State V)	Human	-	-	ELECTRON MICROSCOPY	3.6	<a href="https://doi.org/10.1038/s41586-020-2333-6">10.1038/s41586-020-2333-6</a>
1QU1	H3N2	A/Aichi/2/1968 (Post fusion)	Human	-	-	X-RAY DIFFRACTION	1.9	<a href="https://doi.org/10.1073/pnas.96.16.8967">10.1073/pnas.96.16.8967</a>
1HTM	H3N2	A/Aichi/2/1968 (Post fusion)	Human	-	-	X-RAY DIFFRACTION	2.5	<a href="https://doi.org/10.1038/371037a0">10.1038/371037a0</a>



Repulsive +/-															
HA1-HA1	0,0%	0,0%	0,0%	0,0%	33,3%	0,0%	0,0%	0,0%	0,0%	0,0%	27,3%	33,3%	37,5%	0,0%	0,0%
HA1-HA2	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	100,0%	0,0%	
HA2-HA2	0,0%	0,0%	31,6%	0,0%	30,0%	33,3%	33,3%	0,0%	0,0%	0,0%	30,0%	60,0%	20,0%	33,3%	
Inter	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	33,3%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	
InterHA1-HA1	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	
InterHA1-HA2	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	
InterHA2-HA2	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	50,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	
Intra	0,0%	0,0%	22,2%	0,0%	31,0%	23,1%	0,0%	0,0%	0,0%	16,7%	32,1%	51,7%	26,1%	22,2%	
IntraHA1-HA1	0,0%	0,0%	0,0%	0,0%	33,3%	0,0%	0,0%	0,0%	0,0%	27,3%	33,3%	37,5%	0,0%	0,0%	
IntraHA1-HA2	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	100,0%	0,0%	
IntraHA2-HA2	0,0%	0,0%	40,0%	0,0%	35,3%	40,0%	0,0%	0,0%	0,0%	0,0%	35,3%	75,0%	25,0%	40,0%	
Total	0,0%	0,0%	17,6%	0,0%	25,0%	18,8%	13,0%	0,0%	0,0%	13,6%	25,0%	42,9%	22,2%	17,1%	





Repulsive HA1-HA1	39	39	39	33	42
Repulsive HA1-HA2	5	9	7	2	0
Repulsive HA2-HA2	49	45	45	22	55
Repulsive ITR-p	3	9	9	4	16
Repulsive ITR-p HA1-HA1	0	0	0	0	0
Repulsive ITR-p HA1-HA2	2	0	0	0	0
Repulsive ITR-p HA2-HA2	1	9	9	4	16
Repulsive ITA-p	90	84	82	53	81
Repulsive ITA-p HA1-HA1	39	39	39	33	42
Repulsive ITA-p HA1-HA2	3	9	7	2	0
Repulsive ITA-p HA2-HA2	48	36	36	18	39
<b>Repulsive (total)</b>	<b>93</b>	<b>93</b>	<b>91</b>	<b>57</b>	<b>97</b>

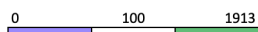
+0%	+0%	-15,38%	+27,27%
+80,00%	-22,22%	-71,43%	-100,00%
-8,16%	+0%	-51,11%	+150,00%
+200,00%	+0%	-55,56%	+300,00%
-100,00%			
+800,00%	+0%	-55,56%	+300,00%
-6,67%	-2,38%	-35,37%	+52,83%
+0%	+0%	-15,38%	+27,27%
+200,00%	-22,22%	-71,43%	-100,00%
-25,00%	+0%	-50,00%	+116,67%
+0%	-2,15%	-37,36%	+70,18%

Repulsive -/- HA1-HA1	6	6	6	0	6
Repulsive -/- HA1-HA2	3	9	7	0	0
Repulsive -/- HA2-HA2	12	15	15	10	29
Repulsive -/- ITR-p	0	9	9	4	5
Repulsive -/- ITR-p HA1-HA1	0	0	0	0	0
Repulsive -/- ITR-p HA1-HA2	0	0	0	0	0
Repulsive -/- ITR-p HA2-HA2	0	9	9	4	5
Repulsive -/- ITA-p	21	21	19	6	30
Repulsive -/- ITA-p HA1-HA1	6	6	6	0	6
Repulsive -/- ITA-p HA1-HA2	3	9	7	0	0
Repulsive -/- ITA-p HA2-HA2	12	6	6	6	24
<b>Repulsive -/- (total)</b>	<b>21</b>	<b>30</b>	<b>28</b>	<b>10</b>	<b>35</b>

+0%	+0%	-100,00%	
+200,00%	-22,22%	-100,00%	
+25,00%	+0%	-33,33%	+190,00%
	+0%	-55,56%	+25,00%
+800%	+0%	-55,56%	+25,00%
+0%	-9,52%	-68,42%	+400,00%
+0%	+0%	-100,00%	
200,00%	-22,22%	-100,00%	
-50,00%	+0%	+0%	+300,00%
+42,86%	-6,67%	-64,29%	+250,00%

Repulsive +/- HA1-HA1	33	33	33	33	36
Repulsive +/- HA1-HA2	2	0	0	2	0
Repulsive +/- HA2-HA2	37	30	30	12	26
Repulsive +/- ITR-p	3	0	0	0	11
Repulsive +/- ITR-p HA1-HA1	0	0	0	0	0
Repulsive +/- ITR-p HA1-HA2	2	0	0	0	0
Repulsive +/- ITR-p HA2-HA2	1	0	0	0	11
Repulsive +/- ITA-p	69	63	63	47	51
Repulsive +/- ITA-p HA1-HA1	33	33	33	33	36
Repulsive +/- ITA-p HA1-HA2	0	0	0	2	0
Repulsive +/- ITA-p HA2-HA2	36	30	30	12	15
<b>Repulsive +/- (total)</b>	<b>72</b>	<b>63</b>	<b>63</b>	<b>47</b>	<b>62</b>

+0%	+0%	+0%	+9,09%
-100,00%			-100,00%
-18,92%	+0%	-60,00%	+116,67%
-100,00%			+1000%
-100,00%			
-100,00%			+1000%
-8,70%	+0%	-25,40%	+8,51%
+0%	+0%	+0%	+9,09%
			-100,00%
-16,67%	+0%	-60,00%	+25,00%
-12,50%	+0%	-25,40%	+31,91%



ITR-p : Inter-protomer

ITA-p : Intra-protomer

All : All interaction types

\* newly formed interactions, variation was calculated as if one interaction pre-existed

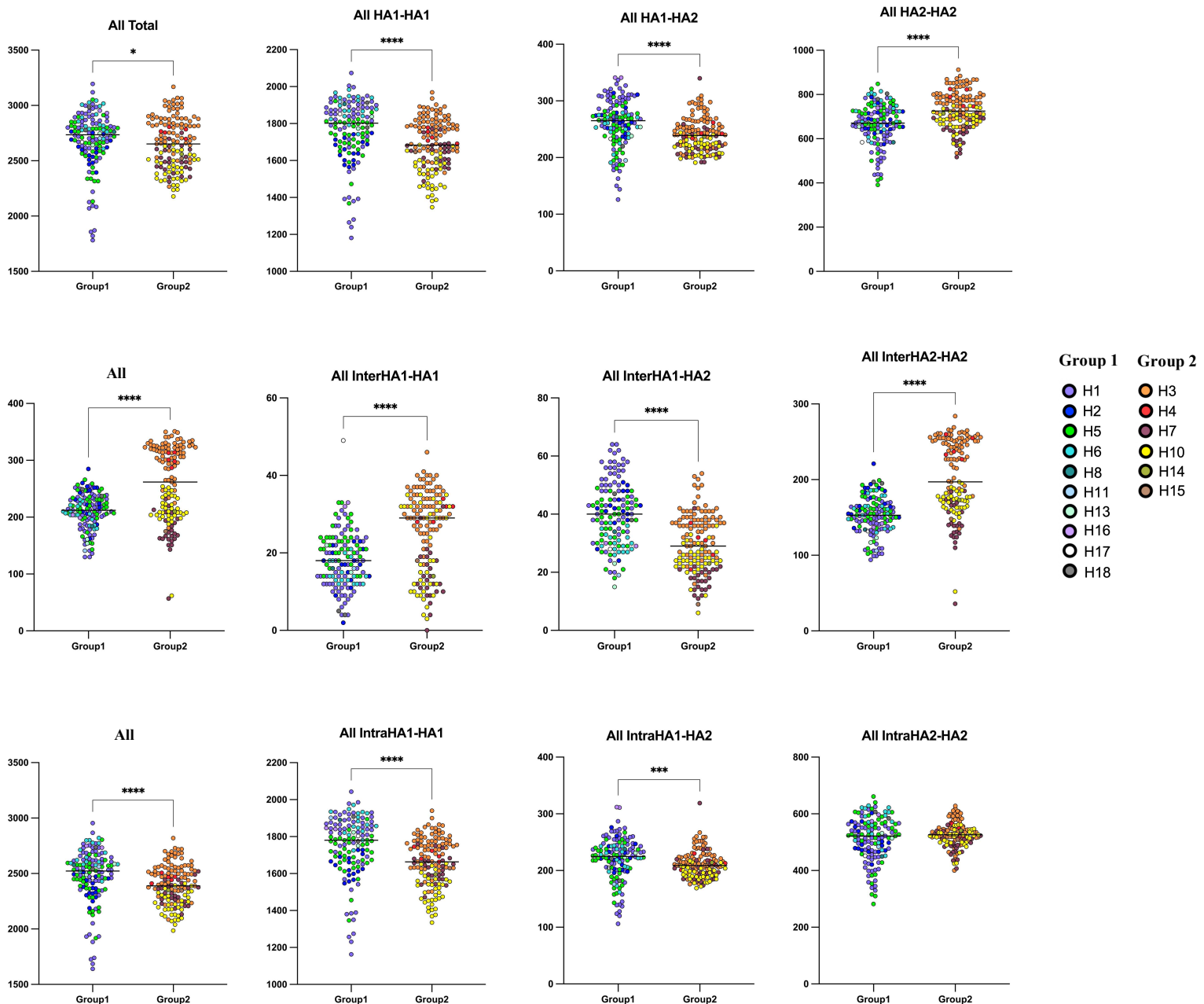


### 3.2.7 Supplementary tables/figures

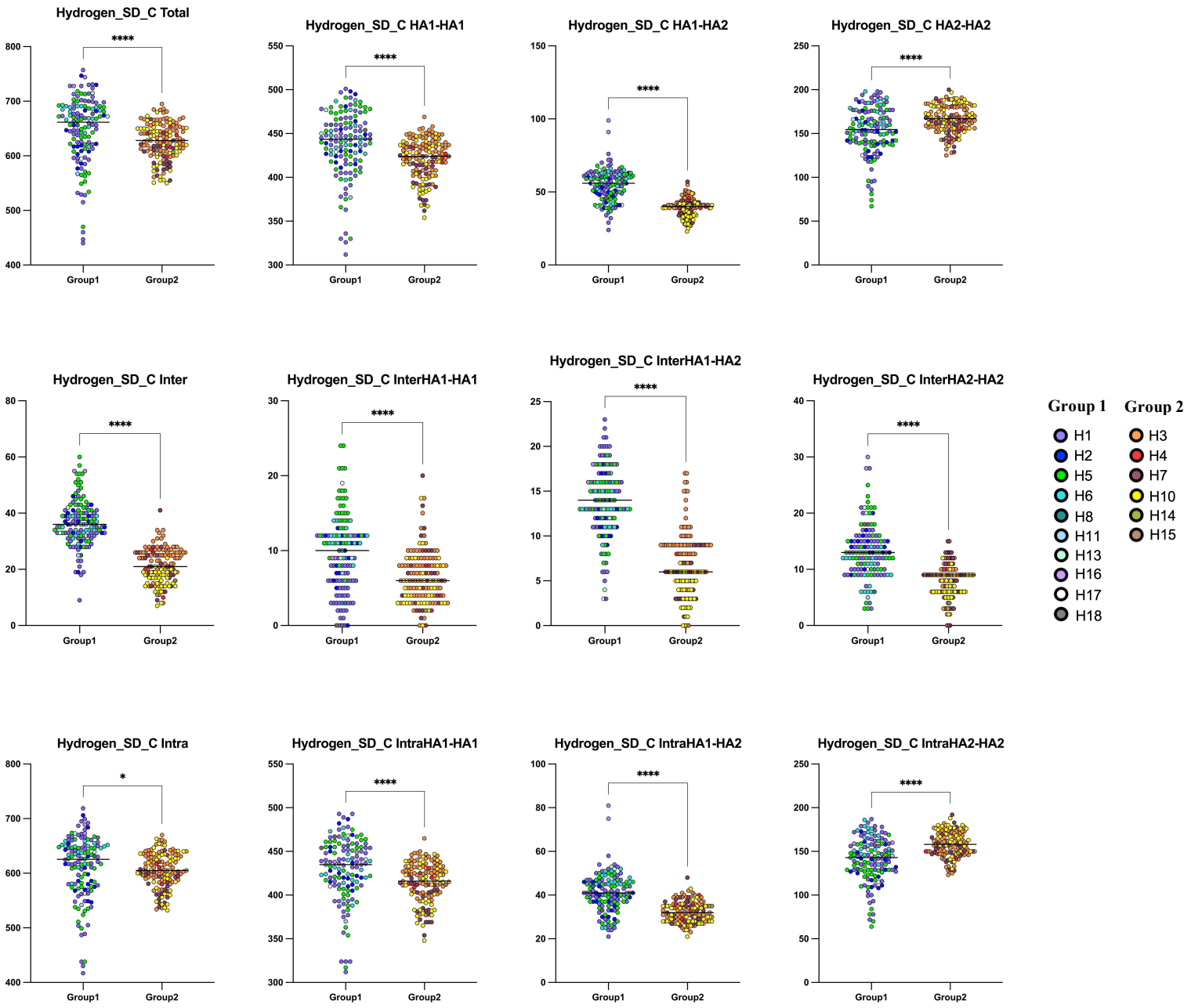
Annex 1 - Supplementary figure 1. Interaction types and distributions values plotted according to HA groups.

Dotplots depicting the number of calculated interactions from each structure. Values were plotted according to HA group classification. Dot colors differentiate subtypes. One dot represents one value of one structure. Median value for each group is represented as a black line. Data were compared between group 1 (n=132) and group 2 (n=137) by two-tailed Mann-Whitney test (\*\*\*\* P < 0.0001. \*\*\* P < 0.001. \*\* P < 0.01. \* P < 0.05. no symbol means not significant).

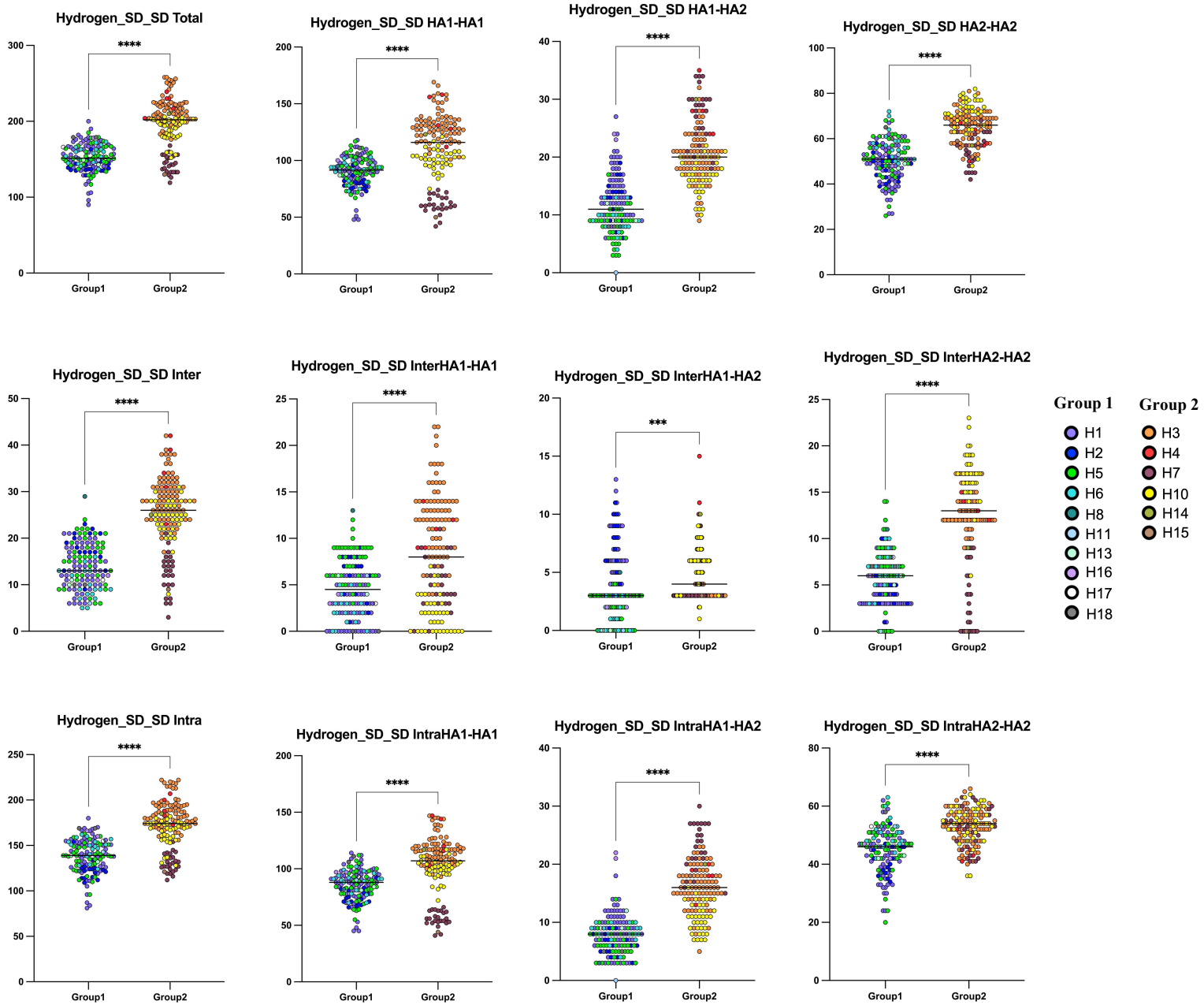
## Total interactions



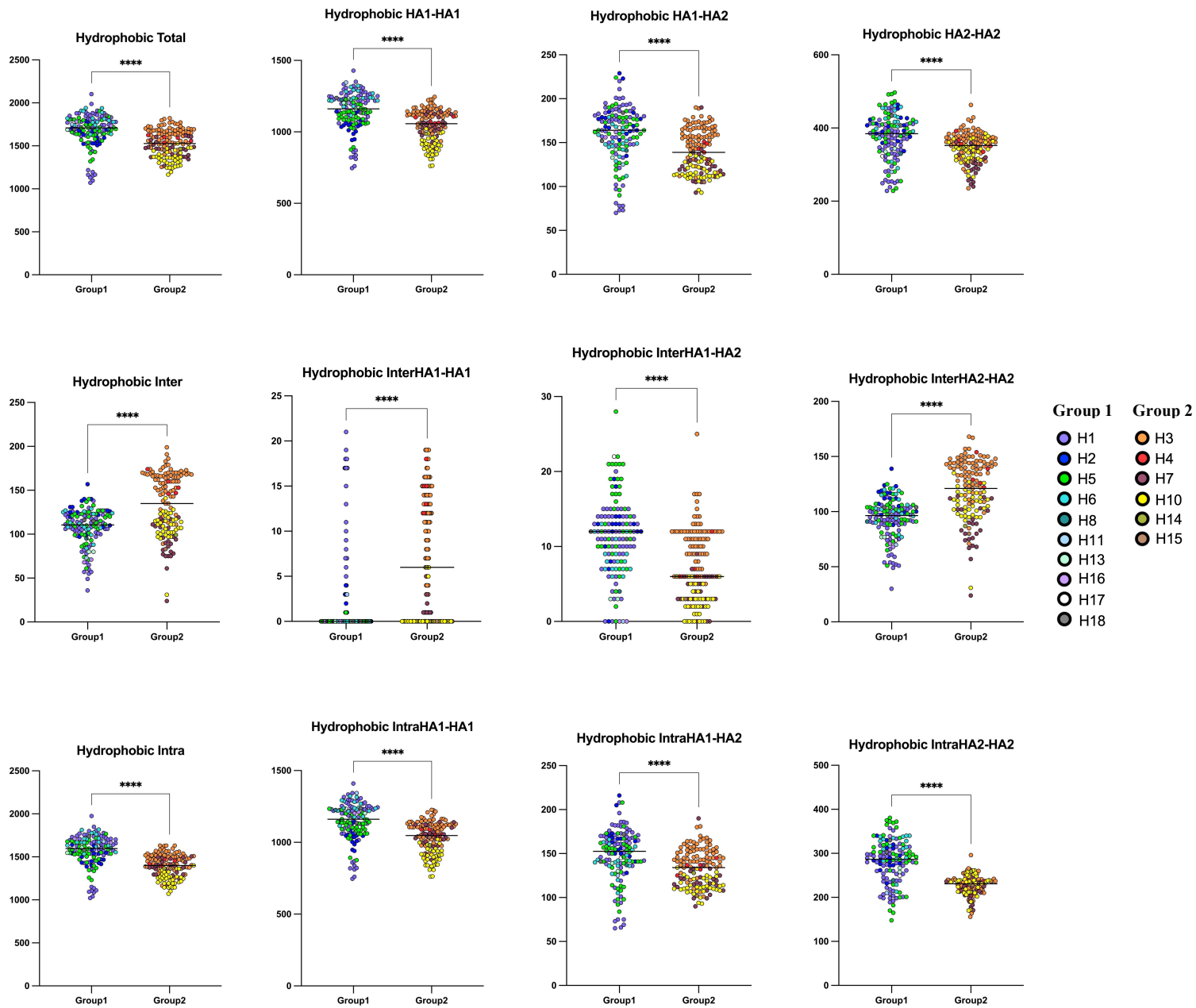
# Hydrogen SD-C



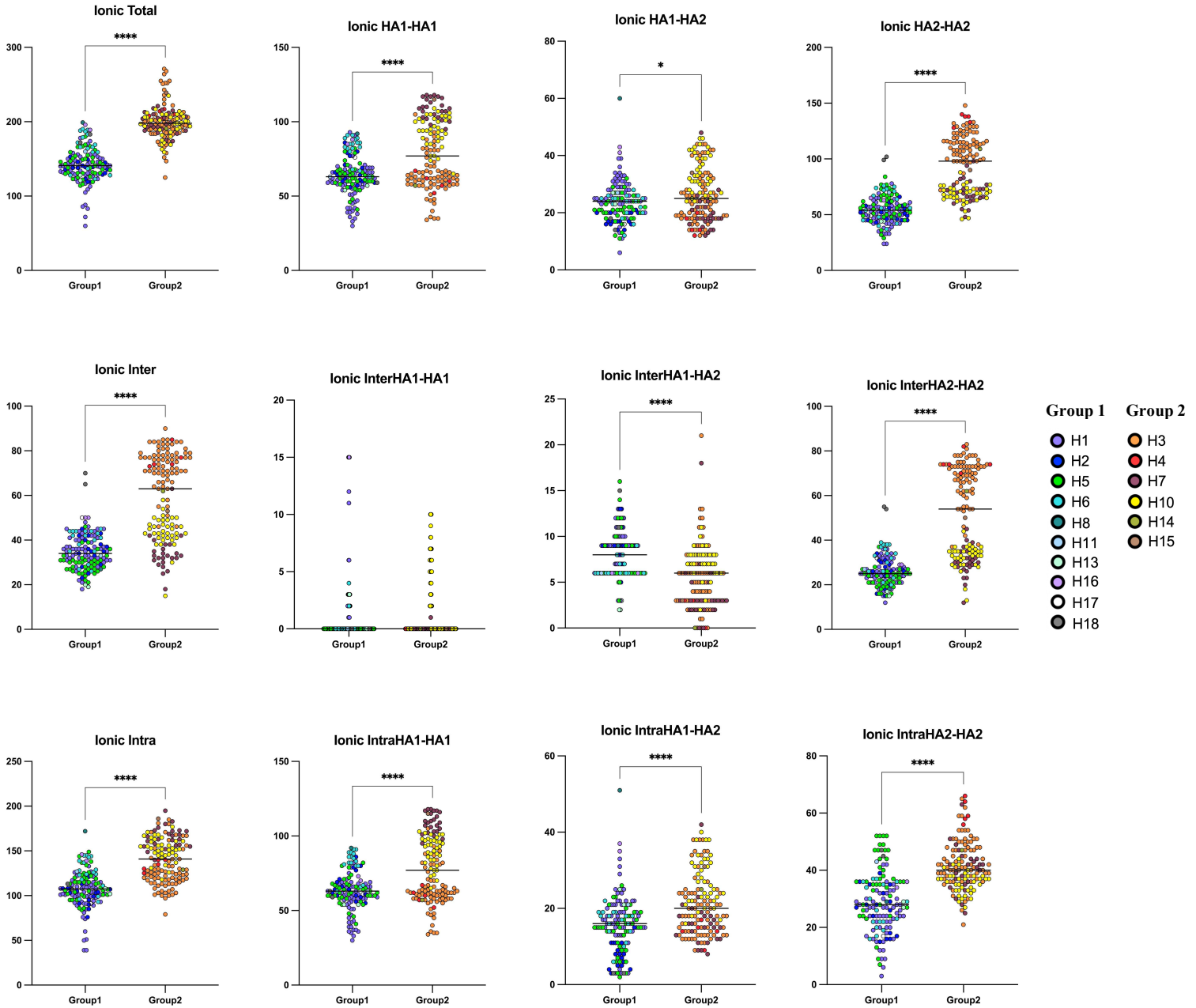
# Hydrogen SD-SD



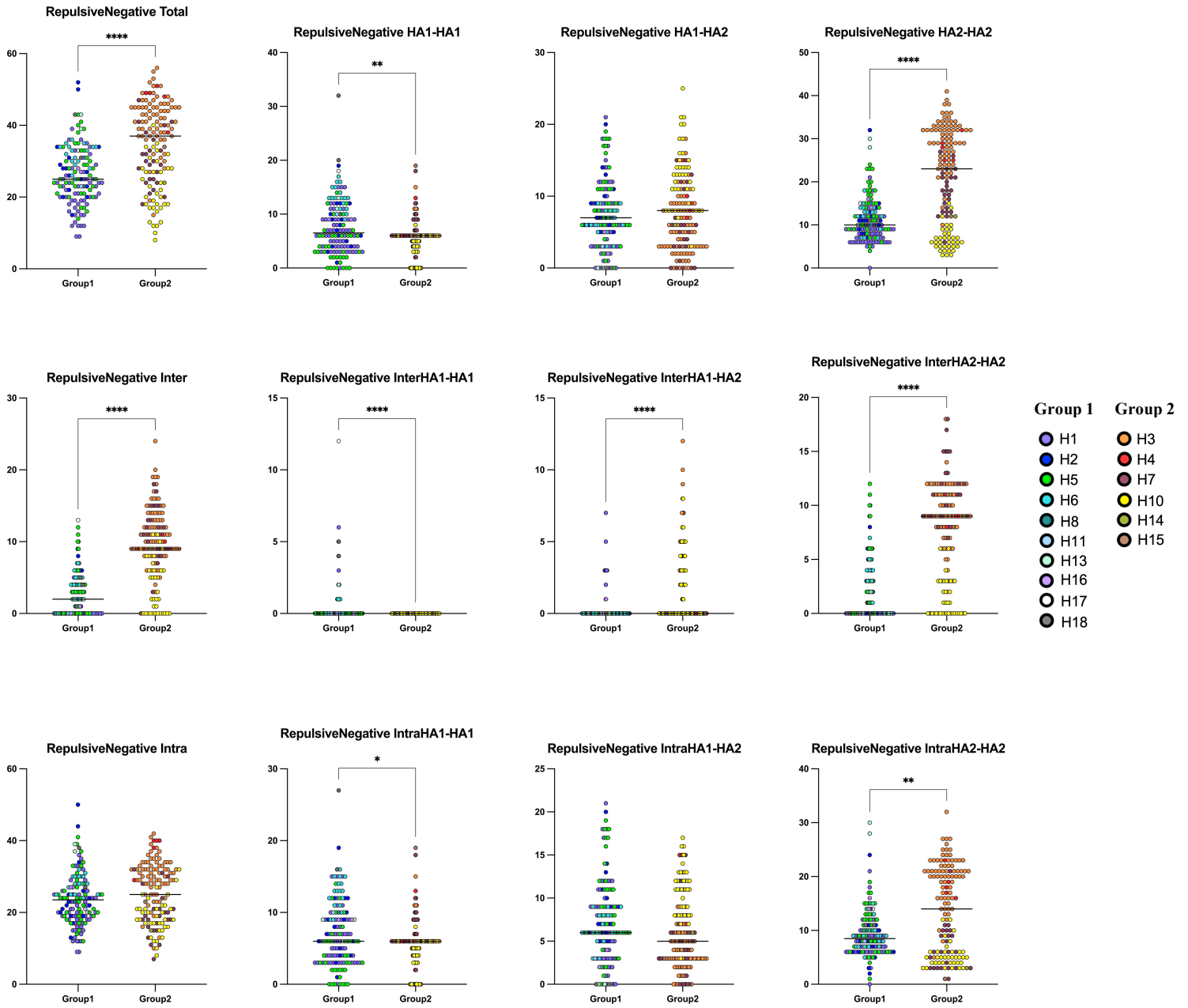
# Hydrophobic



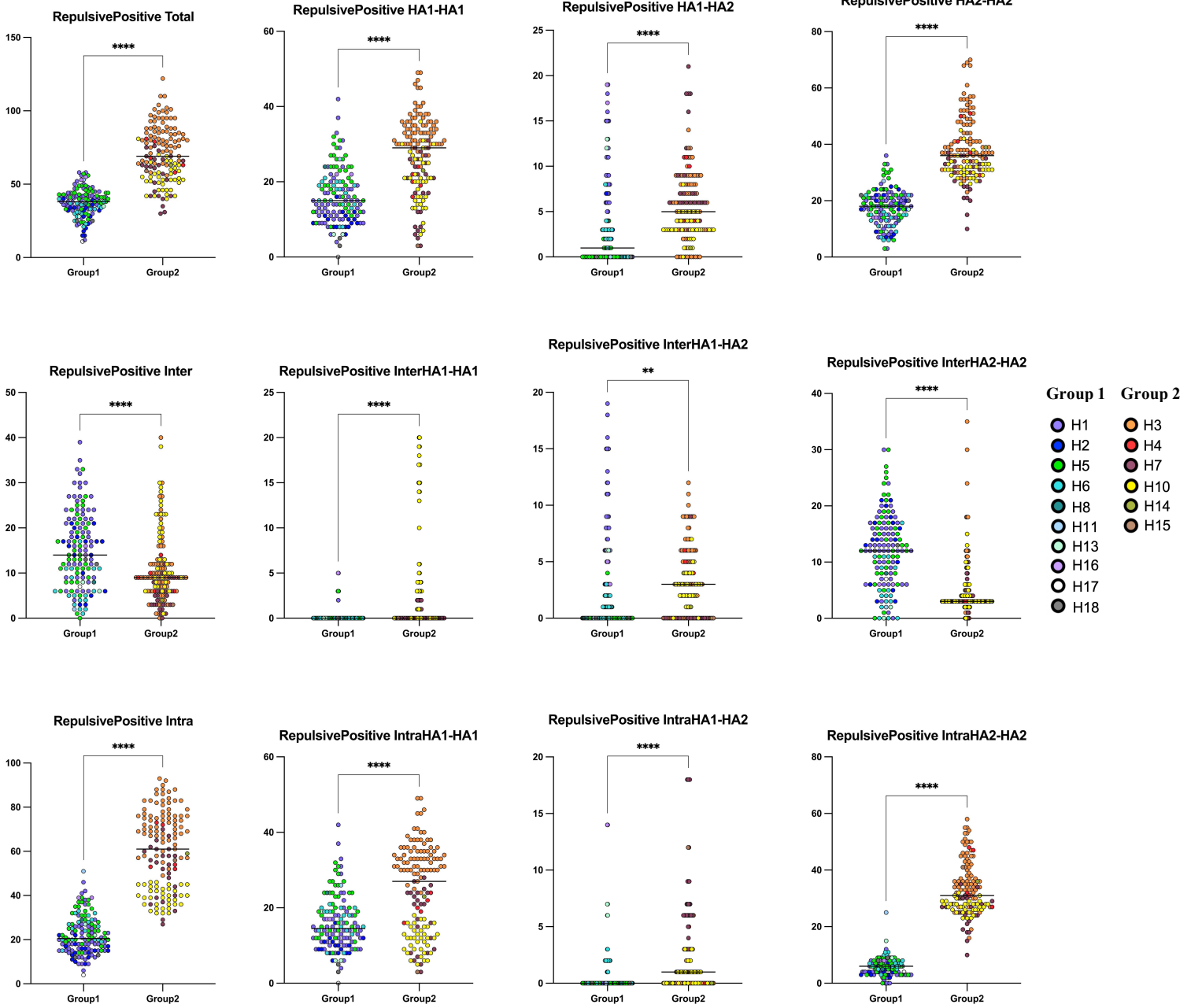
# Salt-bridges (ionic)



# Repulsives -/- (negative)



# Repulsives +/+ (positive)







### 3.2.8 Material and methods

#### **Data preparation**

Available trimeric hemagglutinin (HA) pre-fusion, intermediates and post-fusion structures were gathered from the Protein Data Bank (PDB). Files representing monomeric structures were not considered so that our data set provides meaningful interface information. A total of 281 trimeric HA structures were retrieved, including 269 pre-fusion structures and 12 intermediate/post-fusion structures, with 16 of the 18 HA subtypes represented (H9 and H12 were not represented). The data set is accessible in Extended Data Table 1 with the following information: PDB accession code, subtype, strain, host of origin, amino acid substitution(s) (i.e. mutations) compared to the corresponding wild-type strain and location in HA<sub>1</sub> or HA<sub>2</sub> when present (of note most mutations were located in the receptor binding region of HA<sub>1</sub>); nature of the complexed element (sialic acids, sialic acid analogs, antibody, Fab fragment) when present together with HA in the resolved structure (i.e. complex), structure resolution method, resolution (Å) and references.

Files were browsed and elements of interest were selected according to the standard 80 characters identification nomenclature of PDB files ATOM lines (Code n°1). PDB files contained various chains and numbering nomenclatures. They were thus formatted to harmonize the data by applying a unique nomenclature as follows. First, amino acid numbering between structure files was standardized to easily delimitate HA<sub>1</sub> from HA<sub>2</sub> in each structure (Codes n°2 and 3). This step is only useful if a universal numbering is not applied (see 3D-coffee alignment in next paragraph). Meanwhile, a unique chain nomenclature was applied: A/B C/D E/F with A = HA<sub>1</sub> of protomer 1, B = HA<sub>2</sub> of protomer 1, C = HA<sub>1</sub> of protomer 2, D = HA<sub>2</sub> of protomer 2, E = HA<sub>1</sub> of protomer 3 and F = HA<sub>2</sub> of protomer 3

(Codes n°2 and 3). Each structure was classified into HA groups, clades, HA subtypes (e.g. H1, H3), subtypes (e.g. H1N1, H3N2) and infected host for further investigations (Code n°4).

Cleaned files (ATOM lines and standardized nomenclature) were used as basis to produce fasta files which were then used for sequence alignment (Code n°5). The sequences of each of the three protomers of each structure were used to perform the alignment, to avoid discrepancies caused by potential slight differences in structure resolution between protomers. Sequence alignment was performed with 3D-coffee, a special mode of T-coffee (v13.45.45.71ca9cd) that incorporates structural information of the template with `mustang_pair` and `slow_pair` methods<sup>194,195</sup>. A universal numbering was determined from the final sequence alignment and implemented within the cleaned PDB files before computing interactions (Code n°6). The alignment can be accessed in the GitHub repository. Specifications within cleaned files are the following: atom number, atom identification name, residue three letter code, chain, residue number in universal alignment, X coordinates, Y coordinates, Z coordinates, sub-unit, protomer number, residue number in original file.

### **Interaction computations**

Interactions within the structures were computed for each of the 281 files (Code n°7). Our program first screens the distance separating each atom from all the other atoms of the structure and potential interactions were considered when the distance was equal or below 5 Å. Inter-atomic distances were not computed if both atoms belong to the same position, except if they belong to different protomers. Then, interactions were discriminated using atoms identification and specific distance cut-offs based on standard published

criteria<sup>148,151,152,159,174,196</sup>. The program assessed if both precise atoms and their corresponding residues match with an interaction type and if the distance is comprised within the accepted range for this specific interaction type.

For hydrophobic interactions, nonpolar side chain atoms (CB, CG, CE, CD1, CD2, CE2, CE3, CZ2, CZ3, CH2, CE1, CZ, CG1, CG2, CD, CH2) in hydrophobic residues (Ala, Met, Trp, Phe, Tyr, Val, Leu, Ile, Pro) were analyzed and interactions validated if the inter-atomic distance was less than 4.2 Å. Indeed, hydrophobic effect is caused by spontaneous water exclusion resulting from hydrophobic properties of carbon atoms<sup>173</sup>. Hydrophobic interactions between nonpolar side chain atoms (< 4.2 Å) drive the protein folding initiation and upholding by minimizing the non-polar surface, hence constituting the core backbone of globular proteins.

For hydrogen interactions, sidechain-core chain (SD-C) and sidechain-sidechain (SD-SD) interactions were considered. They occur between the partial positive charge of a hydrogen covalently bonded to an electronegative donor atom and the partial negative charge of a lone pair of electrons belonging to a nearby acceptor electronegative atom, and mostly take part in the solvent-exposed layer of the protein<sup>174,175</sup>. SD-C and SD-SD interactions were validated if inter-atomic distance between proton donor and proton acceptor atoms was less than < 3.5 Å<sup>174,175</sup>. More specifically, distances between proton donor atoms (OG, OG1, NE2, ND2, ND1, NE2, NZ, NE, NH1, NH2, OH, NE1) in proton donor residues (Ser, Thr, Gln, Asn, His, Lys, Arg, Tyr, Trp) and proton acceptor atoms (OG, OG1, OE1, OE2, OD1, OD2, ND1, NE2, OH) in proton acceptor residues (Ser, Thr, Glu, Asp, Gln, Asn, His, Tyr) were computed. In SD-C hydrogen interactions, proton donor (N) of core-chain or proton acceptor (O) of core-chain can belong to any residue.

For electrostatic bonds, which comprise attractive and repulsive interactions between charged atoms, attraction (salt-bridge/ionic interaction) occurs when a cation is located close to an anion ( $< 4 \text{ \AA}$ ) whilst repulsion materializes between two identical charges ( $< 5 \text{ \AA}$ )<sup>148,152</sup>. Accordingly, salt-bridges between positively charged atoms (NE, NH1, NH2, NZ, NE2, ND1) in positively charged residues (Arg, Lys, His) and negatively charged atoms (OD1, OD2, OE1, OE2) in negatively charged residues (Asp, Glu), as well as -/- repulsions (between negatively charged atoms) and +/+ repulsions (between positively charged atoms), were computed. All positively charged atoms and all negatively charged atoms were taken into account as the charges are equally distributed between ionizable groups of atoms due to resonance stabilization of the charge.

Then, for each interaction type, inter-atomic interactions were counted according to their location throughout the protein: HA<sub>1</sub>-HA<sub>1</sub>, HA<sub>1</sub>-HA<sub>2</sub>, HA<sub>2</sub>-HA<sub>2</sub> and total interactions; intra-protomer (ITA-p), ITA-p HA<sub>1</sub>-HA<sub>1</sub>, ITA-p HA<sub>1</sub>-HA<sub>2</sub>, ITA-p HA<sub>2</sub>-HA<sub>2</sub>; inter-protomer (ITR-p), ITR-p HA<sub>1</sub>-HA<sub>1</sub>, ITR-p HA<sub>1</sub>-HA<sub>2</sub>, ITR-p HA<sub>2</sub>-HA<sub>2</sub> (Code n°7). Hence, a total of twelve conditions for each interaction type and for all combined interactions were computed.

Data points were grouped according to HA groups, clades, subtypes, and host classification for further analyses (Code n°8).

The median number of interactions (inter-atomic) within the different classifications were calculated for each interaction type and location with GraphPad Prism 9 (v.3.1). Dot plots, grouped bar plots, PCA plots and heatmaps were produced with GraphPad Prism 9 (v.3.1). Regarding PCAs, parallel analysis was carried out to select the principal components. Monte Carlo simulations on random data (1000 iterations) were computed and principal components with eigenvalues greater than the 95<sup>th</sup> percentile of the eigenvalues from the

simulations were retained. Principle components with the highest variance were selected for plotting.

Inter-amino acid interactions (rather than inter-atomic) were also counted for each structure (Code n°8-1 and data points were grouped according to classifications (Code n°8-2) for later use in the analysis of interactions conservation.

### **Interactions conservation**

The universal numbering that we implemented, based on sequence alignment which considers structural properties, was used to assess the degree of conservation of each computed interaction across classifications (Code n°9). Conservation within and between subtypes as well as across HA groups, clades and host of origin classifications were determined.

To do so, all previously computed interactions were browsed to create a list including a unique representative of each existing interaction (inter-amino acids). Then, each single representative was used as a basis to identify, within each structure, the most related interaction. To determine whether the interaction is conserved within a given structure, the following prerequisites were set: (i) The “most related interaction” should be retrieved within or between each of the three protomers, i.e. interactions with a symmetric arrangement. Hence, an interaction solely found on a single protomer (or two) would not be characterized as conserved in the structure; (ii) variation, even slight shift, in amino acid positions (based on universal numbering) was not tolerated so that only interactions between the two residues at the exact same positions were taken into account; (iii) Mutations were tolerated under the strict condition of preserving the precise interaction type.

Then, to be considered as conserved within a given classification (e.g. within clade 3), a specific interaction had to be identified in at least 90% of the total structures within that classification.

Furthermore, the number of conserved interactions were counted according to the different classifications (Code n°10). The percentage of conserved interactions within classifications was calculated for each interaction type and location. Since we solely considered the conservation of symmetric interactions, the median number of inter-amino acids interactions (which were previously computed, see end of “interaction computations” section) corresponding to the precise assessed type/location condition was divided by three for each classification to reflect symmetric interactions. Hence, according to the different classifications, the percentage of conservation was determined as the number of conserved interactions (inter-amino acids) divided by the median value of the number of assessed type/location interactions divided by three.

Finally, for each conserved interaction, the total number of different amino acid combinations resulting from mutations preserving the interaction type were also computed (Code n°10). The mean number of combinations enabling to preserve the interaction type was calculated for each condition according to interaction types and locations throughout the protein. The distribution of the percentage of interactions with different numbers of retrieved combinations was also calculated. Dot/bar plots and heatmaps were produced with GraphPad Prism 9 (v.3.1).

### **Amino acid conservation**

Influenza A viruses HA protein sequences were gathered from the FluDB database. Subtype-based and total sequences alignments were performed using the MAAFT multiple alignment

tool (v7) for large sequence datasets. Amino acid representativeness for the desired position was computed accordingly to HA classifications with python (v3.10).

### **Data Availability**

Codes were written in Perl v5.30.2 and python v3.10. Codes, inputs and outputs are available in the GitHub repository [https://github.com/ValentinOzeel/InfluenzaA\\_Interactions](https://github.com/ValentinOzeel/InfluenzaA_Interactions).

### **Cells**

HEK-293T cells (human embryonic kidney cell line; ATCC CRL-3216) and HEK293-T7 cells constitutively expressing T7 RNA polymerase in the nucleus under the control of the cytomegalovirus promoter (kindly provided by Yves Jacob, Unité de Génétique Moléculaire des Virus ARN, Institut Pasteur, Paris) were maintained in Dulbecco's modified Eagle's medium (DMEM; Gibco, 41965-039) supplemented with 1% penicillin-streptomycin (Gibco; 15140-122) and 10% heat-inactivated fetal bovine serum (Gibco; 10270106). A549 (Human alveolar basal epithelial adenocarcinoma cell line; provided by M. Schwemmler, Freiburg, Germany) were maintained in the same medium as HEK-293T cells. MDCK (Madin-Darby canine kidney; ATCC CCL-34) and MDCK-SIAT1 cells (kindly provided by M. Matrosovitch, St. Jude Children's Research Hospital, Memphis), expressing two-fold higher amounts of  $\alpha$ 2,6-Linked sialic acids and two folds lower amounts of  $\alpha$ 2,3-Linked sialic acids than MDCK cells, were maintained in minimum essential medium (MEM) supplemented with 1% penicillin-streptomycin and 5% heat-inactivated fetal bovine serum (FSC, Gibco; 10270106). All cell lines were frequently tested for mycoplasma contamination by PCR and were confirmed to be mycoplasma-free.

Cells were grown at 37°C, in the presence of 5% CO<sub>2</sub> in a Forma Series II Water Jacketed CO<sub>2</sub> Incubator (Thermofisher).

### **Plasmids**

Plasmids were amplified and purified with the NucleoBond Xtra Midi Plus kit (Macherey-Nagel; 740422.50). DNA purity was validated by measurement of the A<sub>260</sub>/A<sub>280</sub> ratio (1.83 to 1.87) using a NanoDrop lite spectrophotometer (Fisher Scientific).

A reverse-genetic system composed of eight pRF483 bidirectional plasmids (kindly provided by R. Fouchier, Erasmus MC, Rotterdam, The Netherlands), corresponding to the eight viral segments of the influenza A/PR/8/34 virus, was used for virus production<sup>197</sup>. The same system was also generated for the A/Centre/1003/2012 (H3N2) virus, provided by the National Reference Center for Respiratory Viruses (Institut Pasteur, Paris).

Different mutations (Fig. 1) were introduced into the pRF483-HA plasmid (H1 or H3) by site-directed mutagenesis (Q5 Site-Directed Mutagenesis Kit, New England Biolabs; E0554S). The complete sequence of WT and mutant pRF483-HA plasmids, and of the WT pRF483 plasmids corresponding to each of the other segments was validated by Sanger sequencing (GATC Eurofins, Cologne, Germany).

The pTM1-T7Fvo plasmid expressing the firefly luciferase under control of a T7 promoter was used for quantitative fusion assays. The firefly gene was amplified by PCR with high-fidelity polymerase using specific primers. After purification steps, the PCR product and pTM1 plasmid were digested with EcoRI and Sall restriction enzymes and ligated using the rapid DNA ligation kit (Roche; 11635379001). Complete sequence of the construct was validated by Sanger sequencing (GATC Eurofins, Cologne, Germany).



pE-GFPN1 and pmCherry plasmids were used as transfection controls direct microscope visualization and flow cytometry analysis, respectively.

### **Transfections**

HEK-293T cells were seeded to obtain sub-confluent monolayers (350 000 cells per well in 12 well plates). The next day, monolayers were transfected using FuGENE HD transfection reagent (Promega; E2311) at a DNA ( $\mu\text{g}$ ) to  $\mu\text{l}$  *FugeneHD* ( $\mu\text{L}$ ) ratio of 1:4 in Opti-MEM (Gibco; 31985070), according to manufacturer instructions. *After 30h, transfection efficiency was validated* by GFP visualization with an Olympus CKX41 microscope (between 95 and 100% transfection efficiency) or by flow-cytometry with mCherry.

Otherwise stated, the amount of pRF-483-HA (500ng), pRF-483-NA (200ng) and pE-GFPN1/pmCherry (100ng) plasmids co-transfected remained constant among experiments. pRF-483-HA was replaced by pRF483-M for negative control conditions.

pTM1-T7Fvo (600ng) was added to the mix for quantitative fusion assay experiments otherwise the total amount of transfected DNA was adjusted with DNA salmon sperm (600ng).

### **Western blot analysis, total HA expression and HA cleavage by TPCK-treated trypsin**

12 well plates were coated with poly-D-lysine >300kDa (Sigma; P1024) diluted at 0.01% in sterile milli-Q water. After 5h of incubation at room temperature, plates were delicately washed with sterile milli-Q water before seeding cells.

The next day, monolayers of HEK-293T cells were transfected with pE-GFPN1, wild-type pRF483-NA and wild-type or mutated pRF483-HA. Transfected cells were lysed 30h post-transfection in reducing 1X NuPAGE LDS sample buffer (Invitrogen; NP0007) supplemented

with 3%  $\beta$ -mercaptoethanol (Sigma; M6250). Lysates were collected, loaded in QIAshredder columns (Qiagen; 79656) and centrifuged twice at 3000rpm to homogenize and shred samples. Then, samples were boiled for 10min at 85°C. Proteins were separated by electrophoresis (90 minutes at 130V) in Novex NuPAGE MES SDS running buffer (Invitrogen; NP0002) using a NuPAGE 4-12% Bis-Tris midi gel (Invitrogen; WG1402BOX). Samples were transferred by wet transfer in 1X Novex NuPAGE transfer buffer (Invitrogen; NP0006-1) with 10% absolute ethanol onto an ethanol-activated PVDF blotting membrane (GE healthcare LifeScience; 10600023). Membranes were blocked in 1X Dulbecco's phosphate-buffered saline buffer (DPBS, Gibco; 14190144) with 5% non-fat dry milk and 0.1% Tween-20 (Sigma; P9416) for 1h at room temperature. After washing steps, membranes were incubated overnight at 4°C in DPBS-T buffer with 5% bovine serum albumin (Sigma; A9647), 0.03% sodium azide (Interchim; NJK63A) and anti-HA polyclonal rabbit primary antibodies targeting the whole HA (1/2500 dilution, Sinobiological; H1N1 11684-T52/H3N2 11056-T62). The next day, membranes were washed and incubated 1 hour at RT with horseradish peroxidase-coupled secondary antibodies anti-rabbit IgGs (1/10000 dilution, Thermofisher Scientific) in DPBST- 5% BSA. After washing steps, proteins were detected using chemiluminescent reagent Pierce ECL 2 (Thermofisher; 11517371) and visualized with the ChemiDoc MP imaging system (Bio-rad). Furthermore, membranes were washed and incubated overnight at 4°C with GAPDH mouse monoclonal antibodies (Invitrogen: MA1-16757) diluted (1:2000) in PSBT- 5% BSA and GAPDH proteins were revealed using the same method as described above.

For HA<sub>0</sub> cleavage detection, transfected cells were washed twice with FCS-free DMEM.

Washing media was replaced by fresh pre-warmed FCS-free DMEM containing 1 µg/ml TPCK-treated trypsin (Sigma; 4352157).

The plates were incubated 20min at 37°C, 5% CO<sub>2</sub> and cell lysis was performed as described above.

### **Immunofluorescence labelling/Surface expression**

To assess surface expression, monolayers of HEK-293T cells were transfected as described previously except pe-GFP-N1 was replaced by pmCherry to discriminate transfected cells.

HA expression at the surface of transfected HEK-293T cells was measured 30h after transfection with a Beckman Coulter CytoFLEX S flow cytometer. Cells were first detached with TrypLE express (Thermofisher; 12605028) to preserve surface proteins integrity and neutralized in DMEM 10% FCS. Harvested cells were centrifuged at 1500rpm for 5 minutes and resuspended in ice cold DPBS 5% FCS. Cells were centrifuged, incubated 30 minutes at RT with corresponding primary antibodies diluted at 1:100 in DPBS 5% FCS. Cells were spun-down, washed with ice-cold DPBS and incubated 30 minutes at RT with goat anti-rabbit IgG (H+L) highly cross-adsorbed secondary antibody Alexa Fluor Plus 488 (Invitrogen; A32731) diluted at 1:200 in DPBS 5% FCS. Cells were washed again and finally incubated 30 minutes at RT with LIVE/DEAD™ Fixable Violet Dead Cell Stain (Thermofisher; L34955) according to manufacturer recommendations. Finally, cells were washed, resuspended in ice cold DPBS 5% FCS and placed on ice before measurements.

A total of 20,000 alive singlet events were acquired for each sample and fluorescence intensity was investigated with the following settings: exciting 488 nm laser and measurements through a 530/30 nm bandpass filter for FITCS, exciting 561 nm laser and

measurements through a 610/20 nm bandpass filter for mCherry and exciting 405 nm laser and measurements through a 450/50 nm bandpass filter for violet LIVE/DEAD.

A very slight compensation matrix was created using unstained non-transfected cells as well as single-color controls: WT-HA transfected cells stained with primary and secondary FITC antibodies, non-stained cells transfected with mCherry and WT-HA transfected cells stained with LIVE/DEAD solely.

Living singlets were selected according to SSC-A/FSC-A, SSC-A/SSC-H, FSC-A/FSC-H, LIVE/DEAD negative and mCherry positive parameters.

Median FITC fluorescence intensity (MFI) of sample minus median fluorescence intensity of stained non-transfected cells, in arbitrary units (a.u.), was used as a measure of surface HA expression.

Data were collected with CytExpert acquisition software and analyzed with FlowJo software (v10.8.0).

#### **HA-based Qualitative fusion assay:**

HEK293T cells were transfected as described above. 30 hours after transfection, monolayers were gently washed twice with warm FCS-free DMEM + 10mM HEPES. Medium was replaced by warm FCS-free DMEM containing 10mM HEPES and TPCK-treated trypsin (Sigma; 4352157) at a final concentration of 1 µg/ml and cells were incubated 20min at 37°C 5%CO<sub>2</sub>. Then, the medium was removed, and cells were incubated in DMEM 10% FCS for 1h to allow cells to recover. Cells were then treated with DPBS calcium+, magnesium+ (Gibco; 14040141) adjusted to 4.6 pH unit with citric acid (suitable for cell culture, Sigma; C2404) during 10minutes (time needed for an early endosome to mature into a late endosome). Acidic DPBS++ was replaced with warm DMEM 10% FCS and cells were incubated 2h at 37°C

5%CO<sub>2</sub> to allow the formation of syncytia. Syncytia formation was verified by GFP visualization with an Olympus CKX41 microscope.

#### **HA-based Quantitative fusion assay:**

HEK-293T monolayers were transfected as described above, with the reporter plasmid pTM1-T7Fvo added to the mix. At 30h after transfection, transfected cells were detached with TrypLE express (Thermofisher; 12605028) to preserve surface antigens, neutralized in warm DMEM supplemented with 10% FCS, 10mM HEPES (Cayman chemical; 600212) and counted with the Cellometer Auto 1000 (Nexcelom bioscience). HEK293-T7 were harvested and counted the same way.

The day preceding the experiment, 6-well plates were coated overnight at 4°C with poly-D-lysine >300kDa (Sigma; P1024) diluted at 0.01% in sterile milli-Q water. Plates were delicately washed with sterile milli-Q water before seeding cells. A suspension of effector cells (20%) and target HEK293-T7 cells (80%) was prepared in DMEM 10% FCS, 10mM HEPES and seeded in coated plates (1 100 000 cells /well) in triplicate. Plates were well shaken to ensure cells dispersion and complete coverage of the well.

After 3hours incubation at 37°C 5%CO<sub>2</sub>, cells were gently washed twice with pre-warmed FCS-free DMEM + 10mM HEPES. Medium was replaced by pre-warmed FCS-free DMEM containing 10mM HEPES and TPCK-treated trypsin (Sigma; 4352157) at a concentration of 1 µg/mL, and incubated 20min at 37°C 5%CO<sub>2</sub>. Then, the medium was removed, and cells were incubated in D10 for 1hour to enable contacts between HA<sub>1</sub>/receptor as well as cell recovery. Then, co-cultures were treated with DPBS calcium+, magnesium+ (Gibco; 14040141) adjusted to 4.6 pH unit with citric acid (suitable for cell culture, Sigma; C2404)

during 10minutes. Acidic DPBS++ was replaced by warm DMEM + 10% FCS and cells were incubated 3h at 37°C 5%CO<sub>2</sub> to allow syncytia formations and firefly luciferase expression. Afterwards, co-cultures were overlaid with glo lysis buffer (Promega; E2661) and incubated 15min at room temperature. Luciferase activity was measured by using the ONE-Glo™ EX Luciferase Assay System (Promega; E8130). One volume of lysate was mixed with one volume of substrate in white opaque plate (Greiner Bio-One; 655083) and luminescence was measured with a *Centro XS LB960* microplate luminometer (Berthold Technologies) and expressed as relative luminescence units (RLU). Background noise (mean of negative control values) was subtracted from each condition. For each experiment, values of each condition were normalized by the positive control values.

#### **Virus rescue:**

HEK-293T cells (60%) and MDCK (40%; for PR8 RG) or MDCK-SIAT (40%; for H3N2 RG) cells were seeded in 12-well plates (1 000 000 cells/well) and co-cultured in DMEM 10% FCS before transfection. Sub-confluent monolayers were co-transfected with WT or mutant pRF483-HA and the other WT pRF483 plasmids corresponding to the seven other viral segments (500ng per plasmid) and with 100ng of pGFPN1 plasmid. At 24h after transfection, cells were washed twice with FCS-free DMEM then placed in FCS-free DMEM + 1µg/mL TPCK-treated trypsin (Sigma; 4352157) and incubated at 35°C 5% CO<sub>2</sub>. After 48h, viral supernatants were harvested, cleared by centrifugation for 5 min at 1000 rpm and aliquoted. Viral rescue was directly evaluated by titrating the supernatant on MDCK or MDCK-SIAT1 cells, in a standard plaque assay and aliquots were stored at -80°C.

**Plaque assay:**

12 well plates were seeded with MDCK or MDCK-SIAT1 (600 000 cells/w) to obtain confluent monolayers the next day at the time of infection. Plates were washed twice with FCS-free DMEM. Ten-fold virus dilutions in FCS-free DMEM (500µL) were inoculated and incubated 1h at 37°C 5% CO<sub>2</sub> for viral adsorption onto the cells.

Then, an overlay of MEM 2×-Avicel RC581 (FMC biopolymer) supplemented with TPCK-treated trypsin (Sigma; 4352157) *at a final concentration of 1 µg/mL* was added to cover each well, essentially as described<sup>198</sup>. After incubation for 72 hours at 35°C 5% CO<sub>2</sub>, cells were fixed and stained with crystal violet by incubation with a solution of 10% violet crystal (RAL), 10% formaldehyde, 20% ethanol and 60% milliQ water at RT for 30 minutes. Plaques were counted and titers were expressed as the number of PFU/mL.

**Viral RNA extraction and sequencing:**

RNA extraction from 150 µL of cleared viral stock was performed with the RNeasy mini kit (Qiagen; 74104), according to manufacturer instructions. Samples were treated on the column, prior to the second washing step, with RNase-free DNase set (Qiagen; 79254) during 30min. RNA was purified with 1.8X SRPI beads (Beckman Coulter; A66514), according to manufacturer instructions. Double stranded cDNA was produced with the Maximam H minus Double-Stranded cDNA synthesis kit (Thermofisher; K2561). cDNA was purified with 1.8X SRPI beads (Beckman Coulter; A66514), according to manufacturer instructions. Samples were sequenced on an Illumina MiSeq using MiSeq reagent kit v3 (2x300 cycles).

**Statistics:**

All statistical calculations were performed using GraphPad Prism 9.0. Data in graphs represent the mean or median  $\pm$  s.d. of three or more independent replicates of an experiment. The results were considered significant at  $P \leq 0.05$ .



### 3.3 Annex 2: Overall differences in interaction networks between clades

The composition and disposition of interaction networks seemed to differ between clades. We therefore sought to highlight the largest differences.

Differences in interaction networks composition and distribution between the most represented clades (clades 1, 3 and 7) were quantified by extensively comparing the median values of computed interactions corresponding to each condition (**Annex 2 - table 1**). Moreover, for each condition, i.e. interaction type and distribution, we determined the proportion of computed interactions relative to overall interactions. The mean of the median number of interactions, collected from clades, was calculated for each condition and compared to the mean of the median number of total interactions (MMTI; **Annex 1 - Extended Data Fig. 2, Annex 2 - table 1**).

Compared to clades 3 and 7, clade 1 HAs mainly displayed more hydrophobic (+4%; +23%) interactions, which represent 59% of the MMTI, and exhibited less salt-bridges (-28%; -29%), -/- repulsive (-42%; -7%) and +/+ repulsive (-54%; -31%) interactions which account for 7%, 1% and 2% of the MMTI, respectively (**Annex 1 - Extended Data Fig. 2, Annex 2 - table 1**).

Compared to clades 1 and 7, clade 3 HAs mainly displayed more ITR-p interactions (+47%; +54%), which represent 9% of the MMTI. Indeed, they displayed substantially more ITR-p HA<sub>2</sub>-HA<sub>2</sub> interactions (+62%; +52%) which represent 77% of the mean of median numbers of ITR-p interactions. Clade 3 HAs also harbored many more -/- (+72%; +59%) and +/+ (+116%; +49%) repulsions compared to clades 1 and 7 HAs.

Compared to clades 1 and 3, clade 7 HAs had fewer overall interactions (-10%; -13%) with less hydrophobic (-19%; -16%) interactions. However, clade 7 HAs notably displayed more

ITA-p salt-bridges (+47%; +27%), which accounted for 5% of the MMTI. They harbored significantly more -/- (+8%) and +/+ repulsions (+45%) than clade 1, even though less (-37% and -33%, respectively) than clade 3 (**Annex 1 - Extended Data Fig 2, Annex 2 - table 1**).

The composition and disposition of interaction networks that enable the folding the metastable trimeric structure of the HA are clearly different according to clades, indicating different physicochemical strategies to preserve highly similar structure and identical function.



Repulsive +/- HA1-HA1	15	33	20	-54,55%	-25,00%	+120,00%	+65,00%	+33,33%	-39,39%	22,67	0,85%
Repulsive +/- HA1-HA2	1	6	4	-83,33%	-75,00%	+500,00%	+50,00%	+300,00%	-33,33%	3,67	0,14%
Repulsive +/- HA2-HA2	18	41	31	-56,10%	-41,94%	+127,78%	+32,26%	+72,22%	-24,39%	30,00	1,13%
Repulsive +/- ITR-p	15	9	9	+66,67%	+66,67%	-40,00%	+0%	-40,00%	+0%	11,00	0,41%
Repulsive +/- ITR-p HA1-HA1	0	0	3		-100,00%		-100,00%			1,00	0,04%
Repulsive +/- ITR-p HA1-HA2	0	5	2		-100,00%		+150,00%		-60,00%	2,33	0,09%
Repulsive +/- ITR-p HA2-HA2	12	3	3		+300,00%	+300,00%	-75,00%	+0%	-75,00%	6,00	0,23%
Repulsive +/- ITA-p	20	73	42		-72,60%	-52,38%	+265,00%	+73,81%	+110,00%	45,00	1,69%
Repulsive +/- ITA-p HA1-HA1	15	33	13		-54,55%	+15,38%	+120,00%	+153,85%	-13,33%	20,33	0,76%
Repulsive +/- ITA-p HA1-HA2	0	0	1			-100,00%		-100,00%		0,33	0,01%
Repulsive +/- ITA-p HA2-HA2	5	36	27		-86,11%	-81,48%	+620,00%	+33,33%	+440,00%	22,67	0,85%
<b>Repulsive +/- (total)</b>	<b>38</b>	<b>82</b>	<b>55</b>		<b>-53,66%</b>	<b>-30,91%</b>	<b>+115,79%</b>	<b>+49,09%</b>	<b>+44,74%</b>	<b>58,33</b>	<b>2,19%</b>

\*: mean of median number of total interactions (MMTI) taken from clades

^: median values of interactions according to clades

§: comparison of median values of interaction number between clades

§: mean of median value taken from clades for each condition (interaction type and distribution) and comparison of their proportion relative to the mean of median number (taken from clades) of total interactions.

### 3.4 Annex 3: Correlation between interactions balance and subtype- or host-specific HA stability

The pH values triggering the CC vary according to the host of origin, with higher pH for avian IAVs than for human IAVs<sup>133</sup>. When examining interaction networks according to the hosts of origin, PCAs showed that avian and human viruses were represented in each of the described HA physicochemical profiles (**Annex 1 - Extended Data Figure 3**). Hence, the overall HA physicochemical strategy is not restrictive for species-barrier crossing, indicating that stability modulation depends on precise and subtle interactions balance regulation.

HA stability values of wild-type strains belonging to H1, H5, H3 and H7 subtypes were gathered from the literature (**Annex 3 - Table 1**). The calculated mean pH values were 5.36, 5.76, 5.31 and 5.64, respectively (**Annex 3 - Figure 1d**). H3 and H1 are the most stable proteins. They revealed a comparable number of total and ITA-p interactions even though H1 harbored slightly more ITA-p HA<sub>1</sub>-HA<sub>1</sub> hydrophobic interactions (**Annex 3 - Figure 1a, Figure 2**). However, H3 proteins displayed a significantly higher number of salt-bridges, repulsions, and ITR-p interactions. Since these two subtypes exhibit different physicochemical profiles which result in similar stability, it indicates that the high number of ITR-p interactions and salt-bridges in H3 are compensated by the destabilizing effect of repulsions.

H7 proteins are less stable than H3 and H1 but share similar stability with H5. H7 displayed significantly less ITA-p interactions compared to H1, H5 and H3 (**Annex 3 - Figure 1a**). In addition, they exhibited less ITR-p interactions than H5 and H3 with less ITR-p HA<sub>1</sub>-HA<sub>2</sub> interactions compared to H1, H5 and H3 (**Annex 3 - Figure 1a, Figure 2**). However, compared to H1 and H5 they showed substantially more salt-bridges, known to increase stability<sup>199</sup>. Finally, H7 had more repulsions than H1 and more ITA-p +/- repulsions than H5. With not

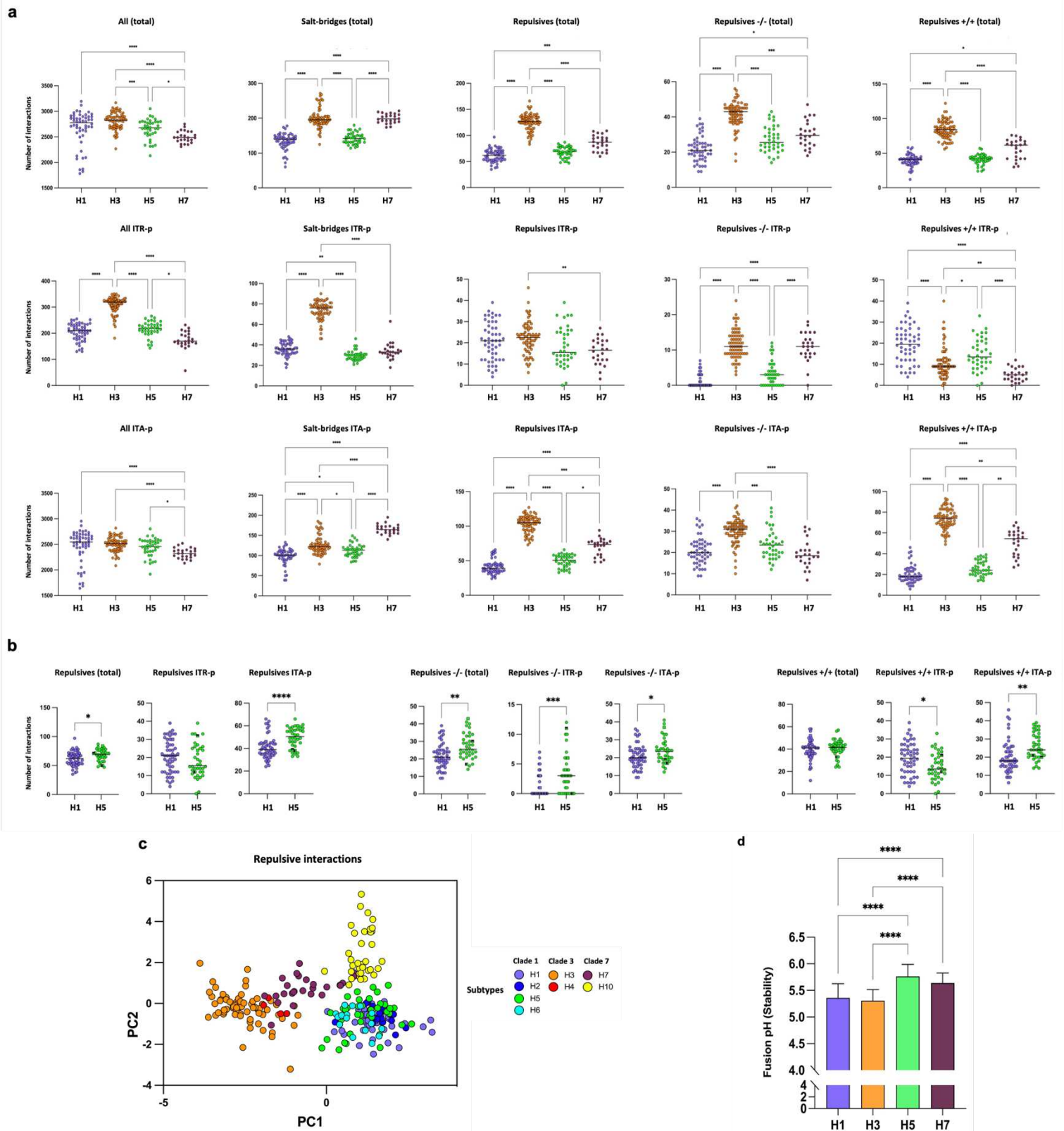
much stability brought by ITA-p and ITR-p interactions, the high number of repulsions harbored by H7 correlates with its low stability.

H5 proteins are less stable than H1 although they share the same overall physicochemical profile. However, H5 proteins harbor less ITR-p HA<sub>1</sub>-HA<sub>2</sub> and more repulsive interactions, especially ITA-p -/- and +/- repulsions which may explain the difference in stability between the two subtypes (**Annex 3 - Figure 1b, Figure 2**).

Moreover, PCA of less represented interactions such as ITR-p and repulsive interactions showed some intra-clade sub-clustering, indicating their role in finely regulating interactions balances (**Annex 1 - Figure 1d, Annex 3 - Figure 1c**).

Collectively, these results illustrate that the balance between destabilization mediated by repulsions and stabilization brought by ITA-p and ITR-p seems to regulate stability.

### Annex 3 - Figure 1. Relation between interactions and HA stability



### Annex 3 - Figure 1. Relation between interactions and HA stability.

a Dot-plots representing the number of total, ITR-p and ITA-p interactions of all interactions, salt-bridges, repulsives, repulsives -/- and repulsives +/- interactions computed from

structures for H1 (n=52), H5 (n=38), H3 (n=70) and H7 (n=24) subtypes. Data were compared using Kruskal-Wallis test followed by Dunn's multiple comparison test (\*\*\*\* adjusted P < 0.0001, \*\*\* adjusted P < 0.001, \*\* adjusted P < 0.01, \* adjusted P < 0.05, no symbol means not significant).

b Dot-plots representing the number of total, ITR-p and ITA-p interactions of repulsives, repulsives -/- and repulsives +/+ interactions for H1 and H5 subtypes. Median value for each group is represented as a black line. Data were compared by two-tailed unpaired t-test (\*\*\*\* P < 0.0001, \*\*\* P < 0.001, \*\* P < 0.01, \* P < 0.05, no symbol means not significant).

c Scatter plot representing PC1 and PC2 coefficients of standardized interaction counts from 258 structures for subtypes H1 (n=52), H2 (n=16), H3 (n=70), H4 (n=5), H5 (n=38), H6 (n=17), H7 (n=24), H10 (n=36). Parallel analysis was carried out to select the principal components. Principle components with the highest variance were selected for plotting. Components represent the standardized interaction number corresponding to repulsives -/- and repulsives +/+ for all location conditions: ITA-p HA<sub>1</sub>-HA<sub>1</sub>, HA<sub>1</sub>-HA<sub>2</sub>, HA<sub>2</sub>-HA<sub>2</sub> and ITR-p HA<sub>1</sub>-HA<sub>1</sub>, HA<sub>1</sub>-HA<sub>2</sub>, HA<sub>2</sub>-HA<sub>2</sub>.

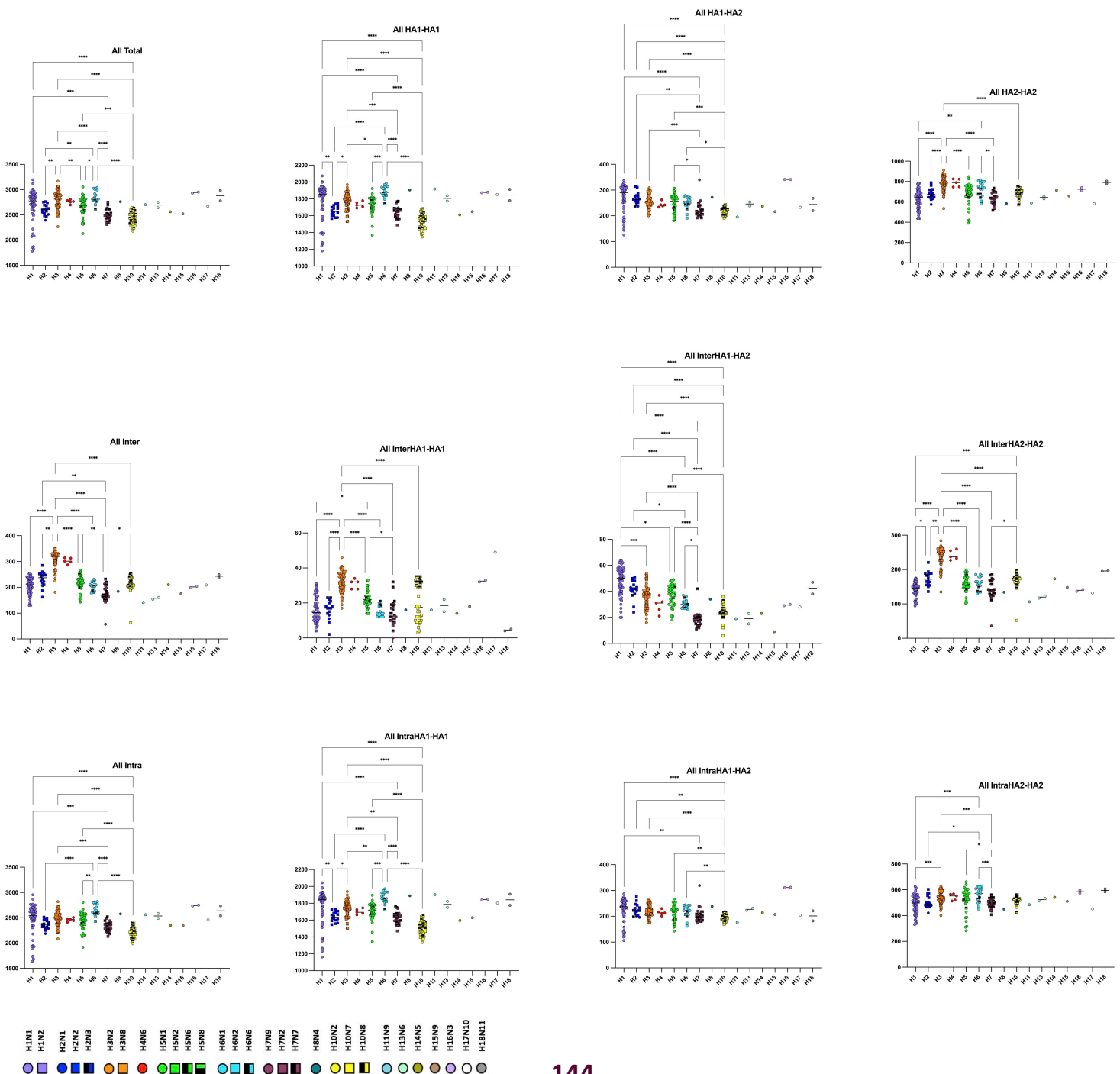
d Bar plot depicting the mean stability values for H1 (n = 55), H5 (n = 38), H3 (n = 31) and H7 (n = 22) subtypes wild-type viruses, collected from the literature. Data were compared using the Kruskal-Wallis test followed by Dunn's multiple comparison test (\*\*\*\* adjusted P < 0.0001, \*\*\* adjusted P < 0.001, no symbol means not significant).



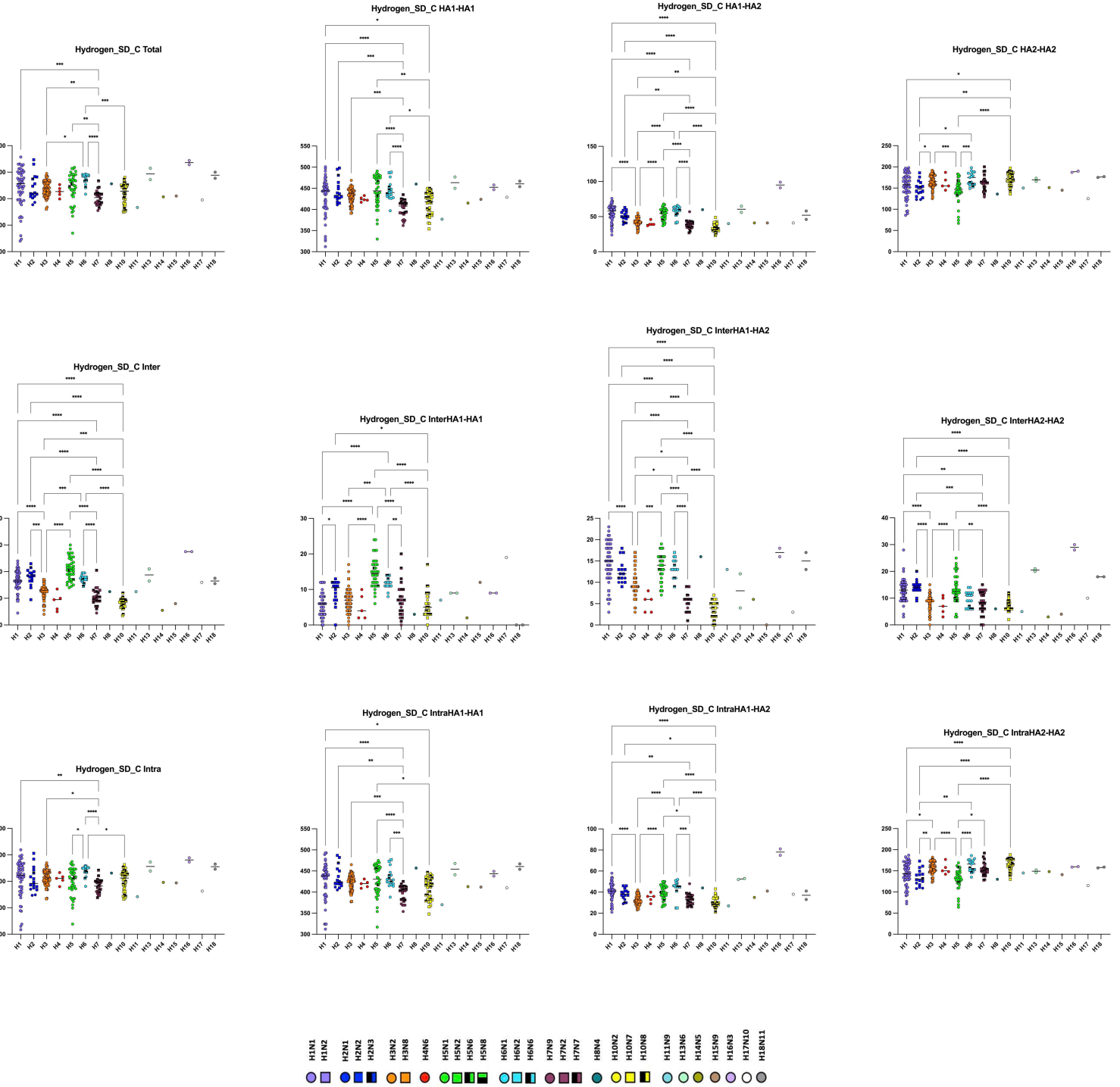
Annex 3 - Figure 2. Interaction types and distributions values plotted according to HA subtypes.

Dotplots depicting the number of calculated interactions from each structure. Values were plotted according to HA subtypes. Colors differentiate HA subtypes; forms differentiate NA subtypes. One dot represents one value of one structure. Median values are represented as a black line. Data were compared between subtypes (H1 (n = 52), H2 (n = 16), H3 (n = 70), H4 (n = 5), H5 (n = 38), H6 (n = 17), H7 (n = 24), H10 (n = 36)) using the Kruskal-Wallis test followed by Dunn's multiple comparison test (\*\*\*\* adjusted P < 0.0001; \*\*\* adjusted P < 0.001; \*\* adjusted P < 0.01; \* adjusted P < 0.05; no symbol means not significant).

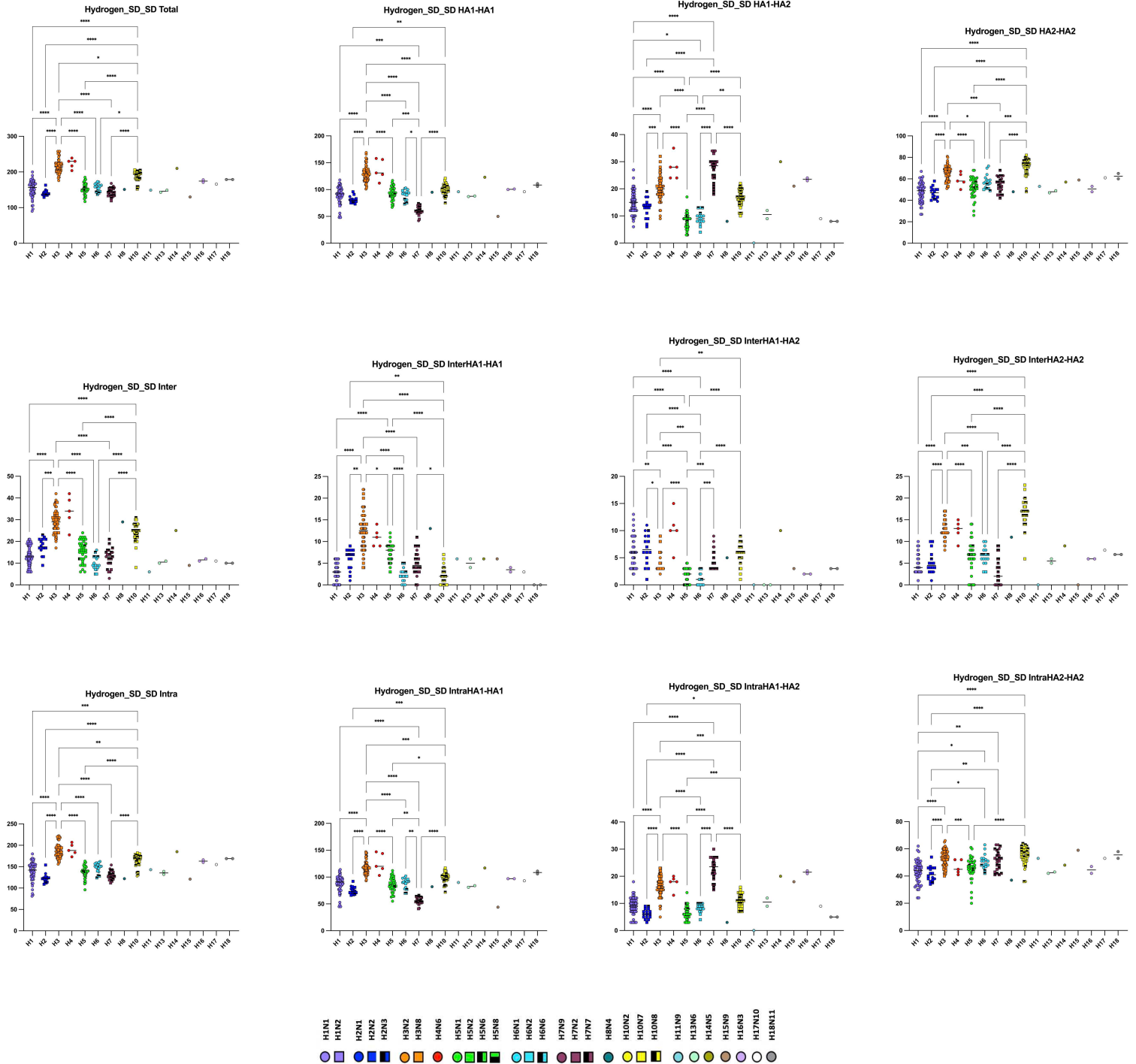
## All interactions



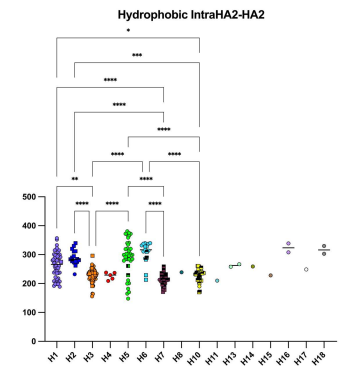
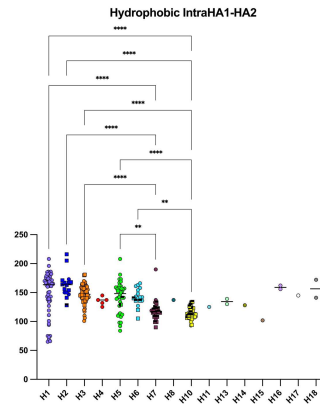
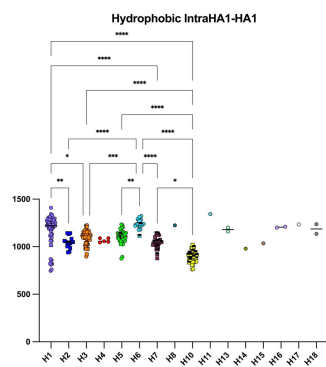
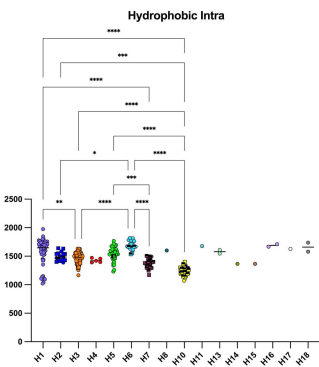
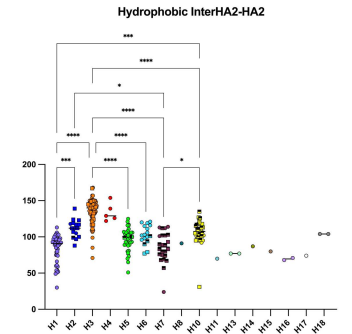
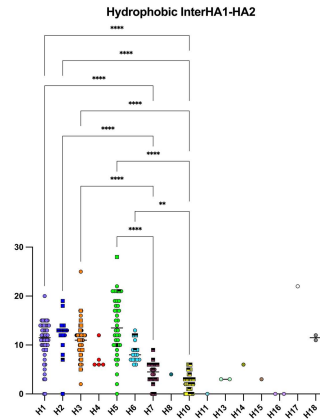
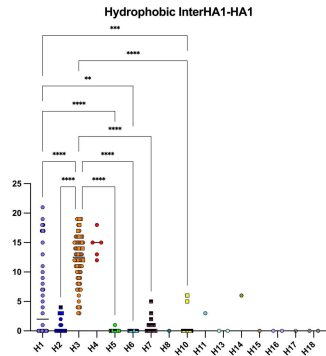
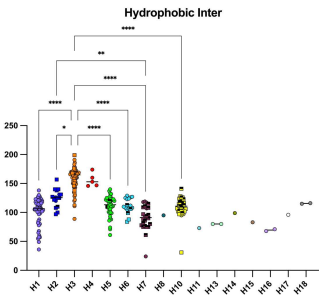
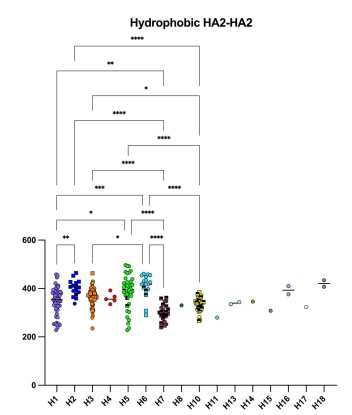
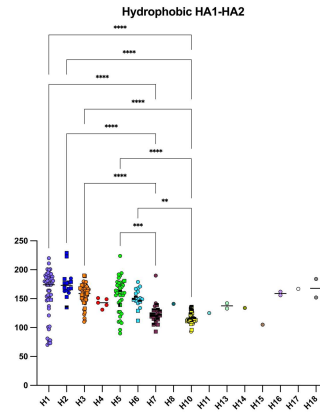
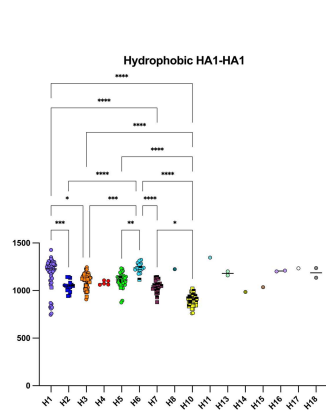
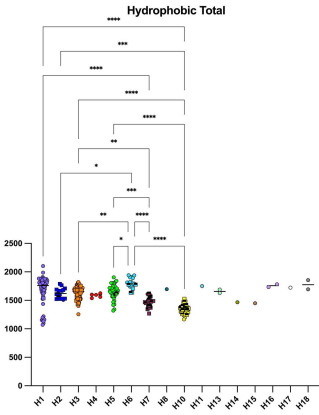
# Hydrogen SD-C



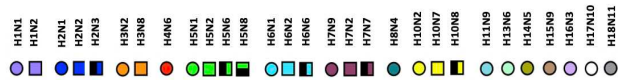
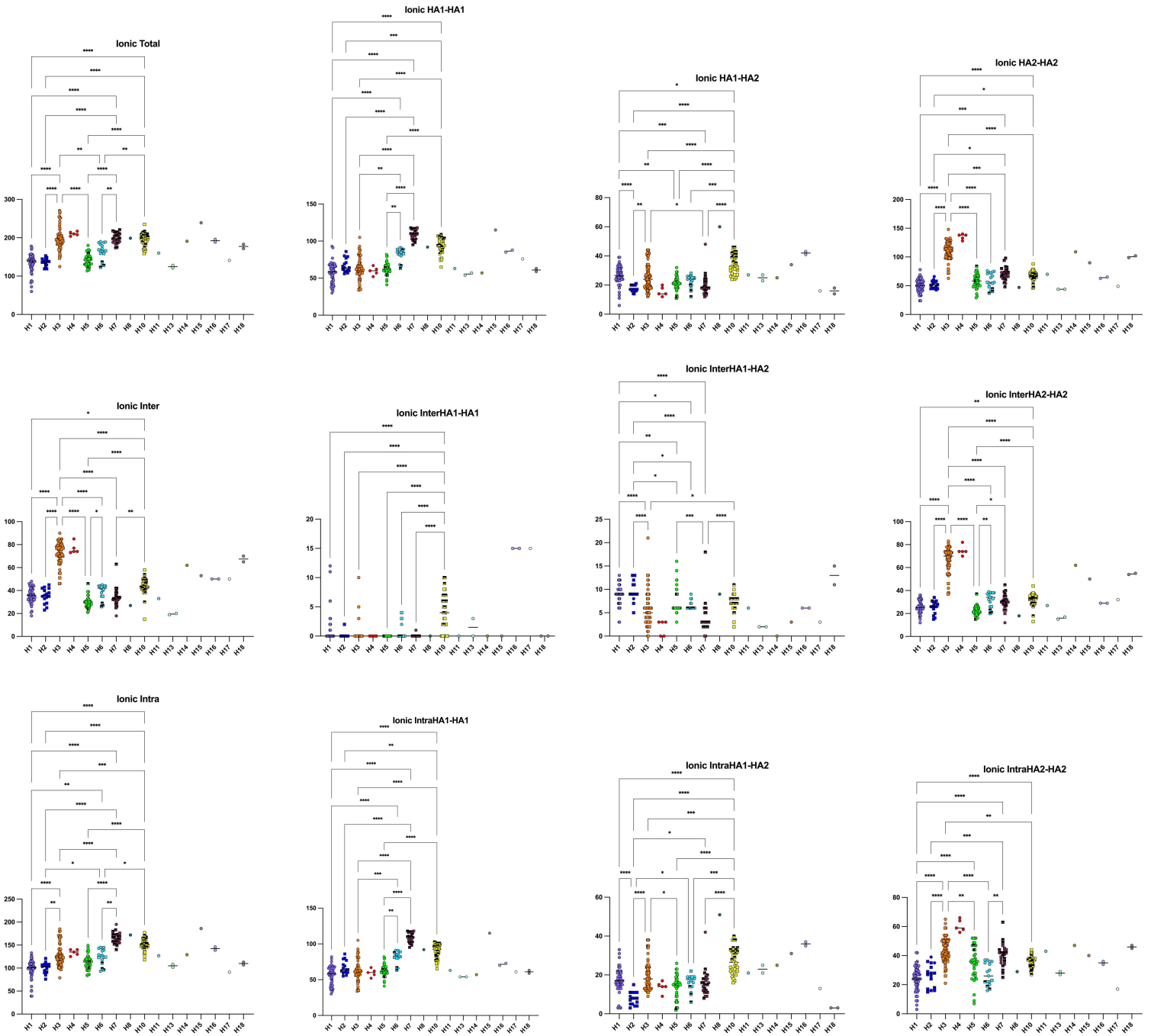
# Hydrogen SD-SD



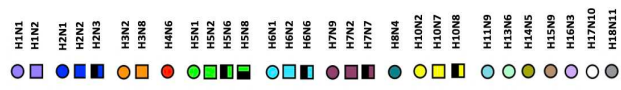
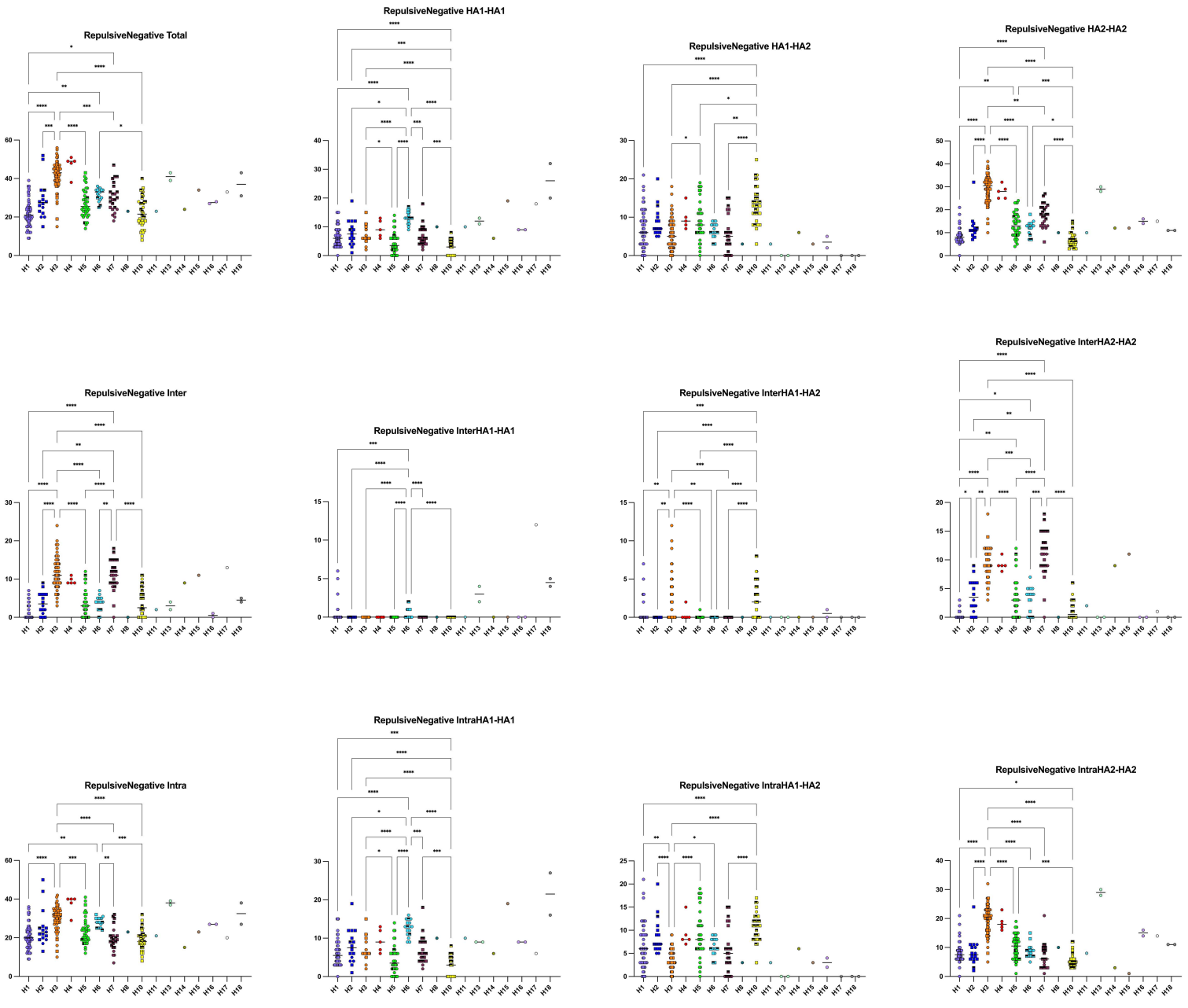
# Hydrophobic



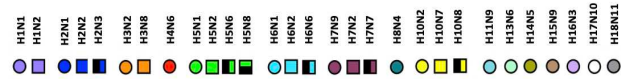
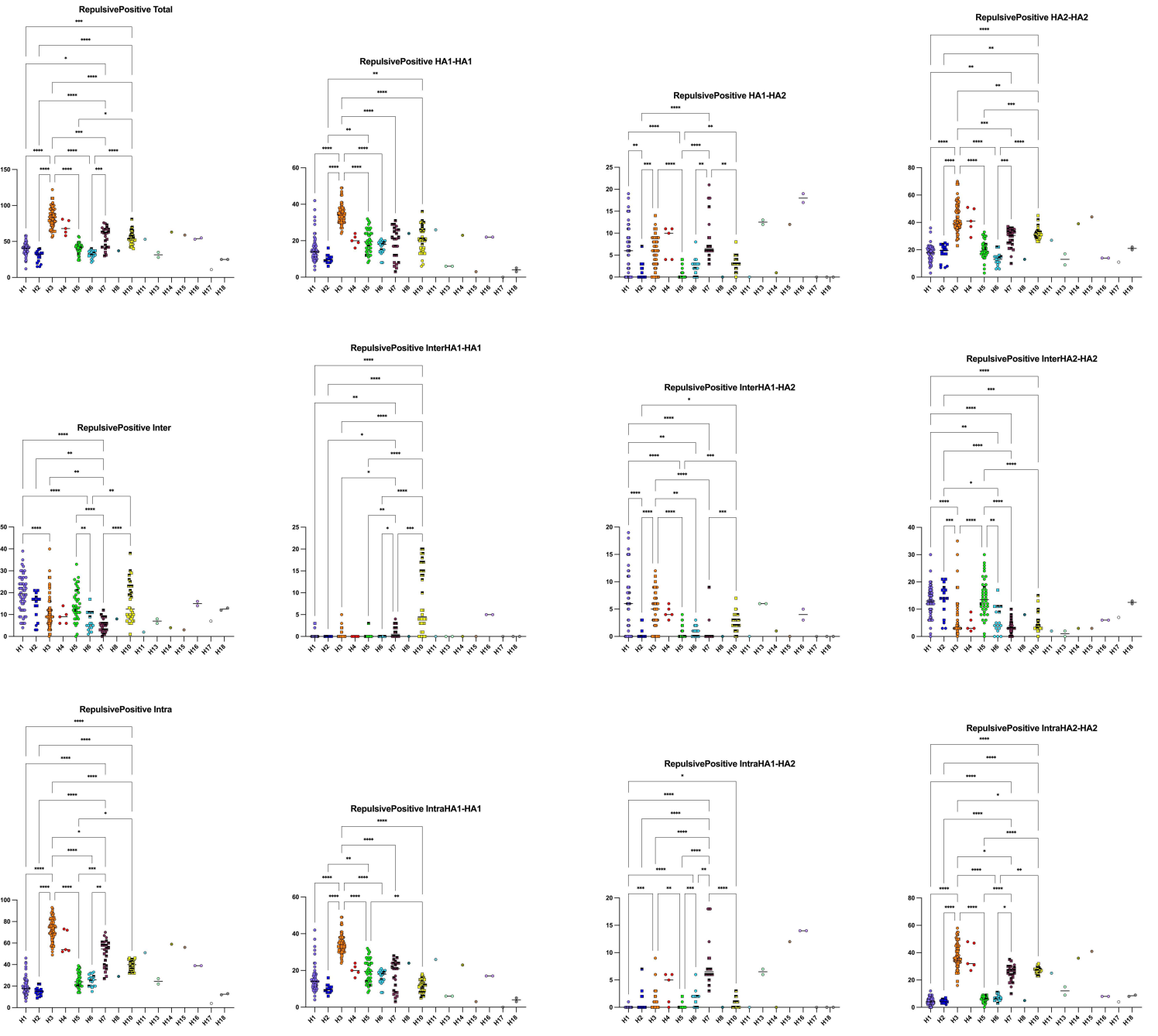
# Salt-bridges (ionic)



# Repulsives -/- (negative)



# Repulsives +/+ (positive)



Annex 3 - Table 1. Influenza A wild-type viruses pH of fusion.

Subtype	Strain	Fusion pH	Reference (DOI)
<b>H1</b>			
H1N1	A/swine/1976/31	5.6	<a href="https://doi.org/10.1016/0264-410x(85)90109-4">10.1016/0264-410x(85)90109-4</a>
H1N1	A/sw/France/OLI/80	6	<a href="https://doi.org/10.1016/0264-410x(85)90109-4">10.1016/0264-410x(85)90109-4</a>
H1N1	A/swine/Germany/2/81	6	<a href="https://doi.org/10.1016/0264-410x(85)90109-4">10.1016/0264-410x(85)90109-4</a>
H1N1	A/duck/Alberta/35/1976	5.6	<a href="https://doi.org/10.1371/journal.ppat.1003151">10.1371/journal.ppat.1003151</a>
H1N1	A/duck/Alberta/35/1976	5.6	<a href="https://doi.org/10.1128/JVI.00057-15">10.1128/JVI.00057-15</a>
H1N1	A/duck/Alberta/35/1976	5.4	<a href="https://doi.org/10.1016/0264-410x(85)90109-4">10.1016/0264-410x(85)90109-4</a>
H1N1	A/duck/Bavaria/2/77	5.4	<a href="https://doi.org/10.1016/0264-410x(85)90109-4">10.1016/0264-410x(85)90109-4</a>
H1N1	A/coot/Schleswig/81	5.4	<a href="https://doi.org/10.1016/0264-410x(85)90109-4">10.1016/0264-410x(85)90109-4</a>
H1N1	A/turkey/Bethlehem-Glilit/1492-B/1982	5.8	<a href="https://doi.org/10.1016/0264-410x(85)90109-4">10.1016/0264-410x(85)90109-4</a>
H1N1	A/mallard/Alberta/119/1998	4.9	<a href="https://doi.org/10.1128/JVI.01223-19">10.1128/JVI.01223-19</a>
H1N1	A/WSN/1933	5	<a href="https://doi.org/10.1371/journal.ppat.1003151">10.1371/journal.ppat.1003151</a>
H1N1	A/WSN/1933	5.5	<a href="https://doi.org/10.1099/jgv.0.001305">10.1099/jgv.0.001305</a>
H1N1	A/WSN/1933	5.2	<a href="https://doi.org/10.1128/JVI.01393-08">10.1128/JVI.01393-08</a>
H1N1	A/WSN/1933	5.2	<a href="https://doi.org/10.1006/viro.1998.9273">10.1006/viro.1998.9273</a>
H1N1	A/Puerto Rico/8/1934	5	<a href="https://doi.org/10.1371/journal.ppat.1003151">10.1371/journal.ppat.1003151</a>
H1N1	A/Puerto Rico/8/1934	5.2	<a href="https://doi.org/10.1128/JVI.00057-15">10.1128/JVI.00057-15</a>
H1N1	A/PR/8/34 Cambridge strain (high virulence)	5.8	<a href="https://doi.org/10.1099/vir.0.035782-0">10.1099/vir.0.035782-0</a>
H1N1	A/PR/8/34 Mount Sinai strain (low virulence)	5.3	<a href="https://doi.org/10.1099/vir.0.035782-0">10.1099/vir.0.035782-0</a>
H1N1	A/Puerto Rico/8/1934	5.4	<a href="https://doi.org/10.1016/0264-410x(85)90109-4">10.1016/0264-410x(85)90109-4</a>
H1N1	A/Puerto Rico/8/1934	5.2	<a href="https://doi.org/10.1128/JVI.06009-11">10.1128/JVI.06009-11</a>
H1N1	A/Brazil/11/78	5.8	<a href="https://doi.org/10.1016/0264-410x(85)90109-4">10.1016/0264-410x(85)90109-4</a>
H1N1	A/Kawasaki/86	4.8	<a href="https://doi.org/10.1006/viro.1998.9273">10.1006/viro.1998.9273</a>
H1N1	A/Memphis/14/1996	4.7	<a href="https://doi.org/10.1128/JVI.01223-19">10.1128/JVI.01223-19</a>
H1N1	A/Brisbane/59/2007	5.2	<a href="https://doi.org/10.1128/JVI.00939-15">10.1128/JVI.00939-15</a>
H1N1	A/Pennsylvania/08/2008	5.8	<a href="https://doi.org/10.1371/journal.ppat.1003151">10.1371/journal.ppat.1003151</a>
H1N1	A/CA/04/2009	5.4	<a href="https://doi.org/10.1128/mSphere.00003-16">10.1128/mSphere.00003-16</a>
H1N1	A/California/04/2009	5.6	<a href="https://doi.org/10.1371/journal.ppat.1003151">10.1371/journal.ppat.1003151</a>
H1N1	A/California/04/2009	5.6	<a href="https://doi.org/10.1073/pnas.1524384113">10.1073/pnas.1524384113</a>
H1N1	A/California/07/2009	5.46	<a href="https://doi.org/10.1038/emi.2016.139">0.1038/emi.2016.139</a>
H1N1	A/California/07/2009	5.4	<a href="https://doi.org/10.1371/journal.ppat.1003831">10.1371/journal.ppat.1003831</a>
H1N1	A/TN/1-560/2009	5.5	<a href="https://doi.org/10.1073/pnas.1524384113">10.1073/pnas.1524384113</a>
H1N1	A/Tennessee/1-560/2009	5.5	<a href="https://doi.org/10.1371/journal.ppat.1006276">10.1371/journal.ppat.1006276</a>
H1N1	A/England/195/2009	5.5	<a href="https://doi.org/10.1128/JVI.00058-19">10.1128/JVI.00058-19</a>
H1N1	A/Tennessee/F1080/2010	5.4	<a href="https://doi.org/10.7554/eLife.56236">10.7554/eLife.56236</a>
H1N1	A/Tennessee/F2090/2011	5.2	<a href="https://doi.org/10.7554/eLife.56236">10.7554/eLife.56236</a>
H1N1	A/Memphis/5/2012	5.2	<a href="https://doi.org/10.7554/eLife.56236">10.7554/eLife.56236</a>
H1N1	A/Georgia/F32551/2012	5.4	<a href="https://doi.org/10.1371/journal.ppat.1003151">10.1371/journal.ppat.1003151</a>



H1N1	A/Georgia/F32551/2012	5.2	<a href="https://doi.org/10.1128/JVI.00057-15">10.1128/JVI.00057-15</a>
H1N1	A/Memphis/43/2013	5.2	<a href="https://doi.org/10.7554/eLife.56236">10.7554/eLife.56236</a>
H1N1	A/Memphis/46/2013	5.2	<a href="https://doi.org/10.7554/eLife.56236">10.7554/eLife.56236</a>
H1N1	A/Memphis/14/2013	5.1	<a href="https://doi.org/10.7554/eLife.56236">10.7554/eLife.56236</a>
H1N1	A/Memphis/40/2013	5.2	<a href="https://doi.org/10.7554/eLife.56236">10.7554/eLife.56236</a>
H1N1	A/Tennessee/F5001/2013	5.3	<a href="https://doi.org/10.7554/eLife.56236">10.7554/eLife.56236</a>
H1N1	A/Tennessee/F5029/2014	5.2	<a href="https://doi.org/10.7554/eLife.56236">10.7554/eLife.56236</a>
H1N1	A/Tennessee/F5031/2014	5.3	<a href="https://doi.org/10.7554/eLife.56236">10.7554/eLife.56236</a>
H1N1	A/Tennessee/F5020C17/2014	5.4	<a href="https://doi.org/10.7554/eLife.56236">10.7554/eLife.56236</a>
H1N1	A/Tennessee/F5008/2014	5.2	<a href="https://doi.org/10.7554/eLife.56236">10.7554/eLife.56236</a>
H1N1	A/Memphis/3/2014	5.2	<a href="https://doi.org/10.7554/eLife.56236">10.7554/eLife.56236</a>
H1N1	A/Memphis/10/2014	5.5	<a href="https://doi.org/10.7554/eLife.56236">10.7554/eLife.56236</a>
H1N1	A/Memphis/5/2014	5.3	<a href="https://doi.org/10.7554/eLife.56236">10.7554/eLife.56236</a>
H1N1	A/Tennessee/F5034A/2014	5.5	<a href="https://doi.org/10.7554/eLife.56236">10.7554/eLife.56236</a>
H1N1	A/Michigan/45/2015	5	<a href="https://doi.org/10.7554/eLife.56236">10.7554/eLife.56236</a>
H1N1	A/Memphis/31/2016	5.4	<a href="https://doi.org/10.7554/eLife.56236">10.7554/eLife.56236</a>
H1N1	A/Memphis/20/2016	5.4	<a href="https://doi.org/10.7554/eLife.56236">10.7554/eLife.56236</a>
H1N1	A/Memphis/25/2016	5.3	<a href="https://doi.org/10.7554/eLife.56236">10.7554/eLife.56236</a>

### H3

H3N2	A/swine/Hong Kong/3/1976	5.4	<a href="https://doi.org/10.1016/0264-410x(85)90109-4">10.1016/0264-410x(85)90109-4</a>
H3N2	A/goose/HongKong/10/76	5.2	<a href="https://doi.org/10.1016/0264-410x(85)90109-4">10.1016/0264-410x(85)90109-4</a>
H3N2	A/chicken/HongKong/8/76	5.2	<a href="https://doi.org/10.1016/0264-410x(85)90109-4">10.1016/0264-410x(85)90109-4</a>
H3N2	A/duck/HongKong/115/77	5.2	<a href="https://doi.org/10.1016/0264-410x(85)90109-4">10.1016/0264-410x(85)90109-4</a>
H3N2	A/Aichi/2/1968	5.2	<a href="https://doi.org/10.1371/journal.ppat.1003151">10.1371/journal.ppat.1003151</a>
H3N2	A/Aichi/2/1968	5.25	<a href="https://doi.org/10.1038/emi.2016.139">0.1038/emi.2016.139</a>
H3N2	A/Aichi/2/1968	5.75	<a href="https://doi.org/10.1016/0092-8674(85)90157-6">10.1016/0092-8674(85)90157-6</a>
H3N2	A/Aichi/2/1968	5	<a href="https://doi.org/10.1128/JVI.00057-15">10.1128/JVI.00057-15</a>
H3N2	A/Aichi/2/1968	5.6	<a href="https://doi.org/10.1002/j.1460-2075.1987.tb02387.x">10.1002/j.1460-2075.1987.tb02387.x</a>
H3N2	A/Aichi/2/1968	5.2	<a href="https://doi.org/10.1073/pnas.93.23.12873">10.1073/pnas.93.23.12873</a>
H3N2	A/Aichi/2/1968	5.3	<a href="https://doi.org/10.1128/JVI.01927-14">10.1128/JVI.01927-14</a>
H3N2	A/Aichi/2/1968	5.2	<a href="https://doi.org/10.1006/viro.1997.8626">10.1006/viro.1997.8626</a>
H3N2	A/Aichi/2/1968	5.2	<a href="https://doi.org/10.1016/j.virol.2007.08.035">10.1016/j.virol.2007.08.035</a>
H3N2	A/Aichi/2/1968	5.5	<a href="https://doi.org/10.1128/JVI.69.11.6643-6651.1995">10.1128/JVI.69.11.6643-6651.1995</a>
H3N2	A/Aichi/2/1968	5.7	<a href="https://doi.org/10.1128/JVI.57.2.603-613.1986">10.1128/JVI.57.2.603-613.1986</a>
H3N2	A/Aichi/2/1968	5.2	<a href="https://doi.org/10.1128/JVI.02325-09">10.1128/JVI.02325-09</a>
H3N2	A/Aichi/2/1968	5.4	<a href="https://doi.org/10.1096/fj.06-7052hyp">10.1096/fj.06-7052hyp</a>
H3N2	A/Aichi/2/1968	5.2	<a href="https://doi.org/10.1006/viro.1998.9273">10.1006/viro.1998.9273</a>
H3N2	A/Hong Kong/1/1968	5.1	<a href="https://doi.org/10.1016/0264-410x(85)90109-4">10.1016/0264-410x(85)90109-4</a>
H3N2	A/Hong Kong/1/1968	4.9	<a href="https://doi.org/10.1128/JVI.01223-19">10.1128/JVI.01223-19</a>
H3N2	A/Hong Kong/1/68	5.2	<a href="https://doi.org/10.1128/JVI.01393-08">10.1128/JVI.01393-08</a>
H3N2	A/Udorn/72	5.6	<a href="https://doi.org/10.1128/JVI.01927-14">10.1128/JVI.01927-14</a>
H3N2	A/Victoria/3/1975	5	<a href="https://doi.org/10.1371/journal.ppat.1003151">10.1371/journal.ppat.1003151</a>
H3N2	A/Netherlands/213/2003	5.4	<a href="https://doi.org/10.1016/j.chom.2020.08.011">10.1016/j.chom.2020.08.011</a>
H3N2	A/Netherlands/213/2003	5.2	<a href="https://doi.org/10.1016/j.cell.2014.02.040">10.1016/j.cell.2014.02.040</a>

H3N2	A/Vienna/28/06	5.4	<a href="https://doi.org/10.1016/j.vaccine.2011.02.100">10.1016/j.vaccine.2011.02.100</a>
H3N2	A/Brisbane/07	5.5	<a href="https://doi.org/10.1128/JVI.01927-14">10.1128/JVI.01927-14</a>
H3N8	A/duck/Ukraine/1/1963	5.6	<a href="https://doi.org/10.1371/journal.ppat.1003151">10.1371/journal.ppat.1003151</a>
H3N8	A/duck/Ukraine/1/1963	5.4	<a href="https://doi.org/10.1128/JVI.00057-15">10.1128/JVI.00057-15</a>
H3N8	A/duck/Ukraine/1/1963	5.4	<a href="https://doi.org/10.1016/0264-410x(85)90109-4">10.1016/0264-410x(85)90109-4</a>
H3N8	A/duck/Chabarovsk/72	5.1	<a href="https://doi.org/10.1016/0264-410x(85)90109-4">10.1016/0264-410x(85)90109-4</a>

## H5

H5N1	A/chicken/Scotland/59	5.8	<a href="https://doi.org/10.1016/0264-410x(85)90109-4">10.1016/0264-410x(85)90109-4</a>
H5N1	A/goose/Hong Kong/437-10/1999 (moderate pathogenicity)	5.35	<a href="https://doi.org/10.1371/journal.ppat.1002398">10.1371/journal.ppat.1002398</a>
H5N1	A/chicken/Hong Kong/YU562/01 (High pathogenicity)	5.7	<a href="https://doi.org/10.1371/journal.ppat.1002398">10.1371/journal.ppat.1002398</a>
H5N1	A/chicken/Vietnam/1194/2004	5.85	<a href="https://doi.org/10.1038/emi.2016.139">0.1038/emi.2016.139</a>
H5N1	A/chicken/Vietnam/C58/04	5.9	<a href="https://doi.org/10.1128/JVI.02069-09">10.1128/JVI.02069-09</a>
H5N1	A/chicken/Vietnam/C58/04	5.9	<a href="https://doi.org/10.1128/JVI.02238-08">10.1128/JVI.02238-08</a>
H5N1	A/chicken/Vietnam/C58/04	5.9	<a href="https://doi.org/10.1128/JVI.03110-12">10.1128/JVI.03110-12</a>
H5N1	A/crow/Kyoto/53/04	5.6	<a href="https://doi.org/10.1074/jbc.M114.611327">10.1074/jbc.M114.611327</a>
H5N1	A/turkey/Turkey/1/2005	5.72	<a href="https://doi.org/10.1038/emi.2016.139">0.1038/emi.2016.139</a>
H5N1	A/chicken/Egypt/CL6/07	5.6	<a href="https://doi.org/10.1074/jbc.M114.611327">10.1074/jbc.M114.611327</a>
H5N1	A/duck/Egypt/D1Br/2007	5.6	<a href="https://doi.org/10.1128/mBio.00081-15">10.1128/mBio.00081-15</a>
H5N1	A/chicken/Vietnam/NCVD-1192/2012	5.99	<a href="https://doi.org/10.1038/emi.2016.139">0.1038/emi.2016.139</a>
H5N1	A/chicken/Vietnam/OIE-2202/2012	5.7	<a href="https://doi.org/10.1038/emi.2016.139">0.1038/emi.2016.139</a>
H5N1	A/Thailand/Kan353/04	5.75	<a href="https://doi.org/10.1074/jbc.M114.611327">10.1074/jbc.M114.611327</a>
H5N1	A/Vietnam/1203/2004	6	<a href="https://doi.org/10.1128/mSphere.00003-16">10.1128/mSphere.00003-16</a>
H5N1	A/Vietnam/1203/2004	5.7	<a href="https://doi.org/10.1038/nature10831">10.1038/nature10831</a>
H5N1	A/Vietnam/1203/2004	5.7	<a href="https://doi.org/10.1128/JVI.01223-19">10.1128/JVI.01223-19</a>
H5N1	A/Indonesia/5/05	5.6	<a href="https://doi.org/10.1016/j.cell.2014.02.040">10.1016/j.cell.2014.02.040</a>
H5N1	A/Indonesia/5/2005	5.7	<a href="https://doi.org/10.1016/j.chom.2020.08.011">10.1016/j.chom.2020.08.011</a>
H5N1	A/Indonesia/5/05	5.6	<a href="https://doi.org/10.1074/jbc.M114.611327">10.1074/jbc.M114.611327</a>
H5N1	A/Shanghai/1/06	5.6	<a href="https://doi.org/10.1074/jbc.M114.611327">10.1074/jbc.M114.611327</a>
H5N2	A/turkey/Kfar Vitkin/71	5.9	<a href="https://doi.org/10.1016/0264-410x(85)90109-4">10.1016/0264-410x(85)90109-4</a>
H5N2	A/Northern pintail/Washington/40964/2014	6	<a href="https://doi.org/10.1128/mSphere.00003-16">10.1128/mSphere.00003-16</a>
H5N2	A/turkey/Minnesota/10915-1/2015	6	<a href="https://doi.org/10.1128/mSphere.00003-16">10.1128/mSphere.00003-16</a>
H5N2	A/turkey/Minnesota/10777/2015	6	<a href="https://doi.org/10.1128/mSphere.00003-16">10.1128/mSphere.00003-16</a>
H5N2	A/turkey/Minnesota/11688-1/2015	6	<a href="https://doi.org/10.1128/mSphere.00003-16">10.1128/mSphere.00003-16</a>
H5N2	A/turkey/South Dakota/11089-3/2015	6	<a href="https://doi.org/10.1128/mSphere.00003-16">10.1128/mSphere.00003-16</a>
H5N2	A/turkey/North Dakota/11419-1/2015	6	<a href="https://doi.org/10.1128/mSphere.00003-16">10.1128/mSphere.00003-16</a>
H5N2	A/snow goose/Missouri/CC15-84A/2015	5.95	<a href="https://doi.org/10.1128/mSphere.00003-16">10.1128/mSphere.00003-16</a>
H5N3	A/tern/South Africa/1961	5.4	<a href="https://doi.org/10.1371/journal.ppat.1003151">10.1371/journal.ppat.1003151</a>
H5N3	A/tern/South Africa/1961	5.4	<a href="https://doi.org/10.1128/JVI.00057-15">10.1128/JVI.00057-15</a>
H5N3	A/Duck/Hong Kong/820/80	5.37	<a href="https://doi.org/10.1074/jbc.M114.611327">10.1074/jbc.M114.611327</a>
H5N8	A/gyrfalcon/Washington/41088-6/2014	6	<a href="https://doi.org/10.1128/mSphere.00003-16">10.1128/mSphere.00003-16</a>
H5N8	A/duck/England/36254/2014	5.9	<a href="https://doi.org/10.1128/mSphere.00003-16">10.1128/mSphere.00003-16</a>
H5N8	A/mallard/Korea/W452/2014	5.9	<a href="https://doi.org/10.1128/mSphere.00003-16">10.1128/mSphere.00003-16</a>

H5N8	A/chicken/Kumamoto/1/7/2014	6	<a href="https://doi.org/10.1128/mSphere.00003-16">10.1128/mSphere.00003-16</a>
H5N9	A/Turkey/Ontario/7732/66	5.12	<a href="https://doi.org/10.1074/jbc.M114.611327">10.1074/jbc.M114.611327</a>
H5N9	A/Turkey/Ontario/7732/66	5.8	<a href="https://doi.org/10.1016/0264-410x(85)90109-4">10.1016/0264-410x(85)90109-4</a>

## H7

H7N1	A/FPV/Rostock/34	5.8	<a href="https://doi.org/10.1016/0264-410x(85)90109-4">10.1016/0264-410x(85)90109-4</a>
H7N1	A/chicken/Germany/34	5.7	<a href="https://doi.org/10.1016/0092-8674(85)90157-6">10.1016/0092-8674(85)90157-6</a>
H7N1	A/parrot/Ulster/73	5.6	<a href="https://doi.org/10.1016/0264-410x(85)90109-4">10.1016/0264-410x(85)90109-4</a>
H7N1	A/chicken/Italy/1279/1999	5.78	<a href="https://doi.org/10.1038/emi.2016.139">0.1038/emi.2016.139</a>
H7N3	A/turkey/England/1963	5.6	<a href="https://doi.org/10.1016/0264-410x(85)90109-4">10.1016/0264-410x(85)90109-4</a>
H7N3	A/turkey/England/1963	5.4	<a href="https://doi.org/10.1371/journal.ppat.1003151">10.1371/journal.ppat.1003151</a>
H7N3	A/turkey/England/1963	5.2	<a href="https://doi.org/10.1128/JVI.00057-15">10.1128/JVI.00057-15</a>
H7N3	A/turkey/Oregon/71	5.8	<a href="https://doi.org/10.1016/0264-410x(85)90109-4">10.1016/0264-410x(85)90109-4</a>
H7N7	A/FPV/Dutch/27	5.6	<a href="https://doi.org/10.1016/0264-410x(85)90109-4">10.1016/0264-410x(85)90109-4</a>
H7N7	A/turkey/England/77	6	<a href="https://doi.org/10.1016/0264-410x(85)90109-4">10.1016/0264-410x(85)90109-4</a>
H7N7	A/Seal/Massachussets/1/80	5.8	<a href="https://doi.org/10.1016/0264-410x(85)90109-4">10.1016/0264-410x(85)90109-4</a>
H7N7	A/wigeon/Osaka/1/01	5.4	<a href="https://doi.org/10.1074/jbc.M114.611327">10.1074/jbc.M114.611327</a>
H7N7	A/Netherlands/219/03	5.4	<a href="https://doi.org/10.1099/vir.0.82256-0">10.1099/vir.0.82256-0</a>
H7N9	A/Anhui/1/13	5.8	<a href="https://doi.org/10.1038/ncomms7553">10.1038/ncomms7553</a>
H7N9	A/Anhui/1/13	5.6	<a href="https://doi.org/10.1128/JVI.03052-15">10.1128/JVI.03052-15</a>
H7N9	A/Anhui/1/13	5.6	<a href="https://doi.org/10.1128/JVI.00216-20">10.1128/JVI.00216-20</a>
H7N9	A/Shanghai/01/2013	5.6	<a href="https://doi.org/10.1038/ncomms7553">10.1038/ncomms7553</a>
H7N9	A/Shanghai/01/2013	5.6	<a href="https://doi.org/10.1128/JVI.01223-19">10.1128/JVI.01223-19</a>
H7N9	A/Shanghai/02/2013	5.76	<a href="https://doi.org/10.1038/emi.2016.139">0.1038/emi.2016.139</a>
H7N9	A/Shanghai/02/2013	5.8	<a href="https://doi.org/10.1038/ncomms7553">10.1038/ncomms7553</a>
H7N9	A/Shanghai/02/2013	5.8	<a href="https://doi.org/10.1016/j.virol.2019.07.012">10.1016/j.virol.2019.07.012</a>
H7N9	A/Taiwan/1/2017	5.4	<a href="https://doi.org/10.1016/j.virol.2019.07.012">10.1016/j.virol.2019.07.012</a>

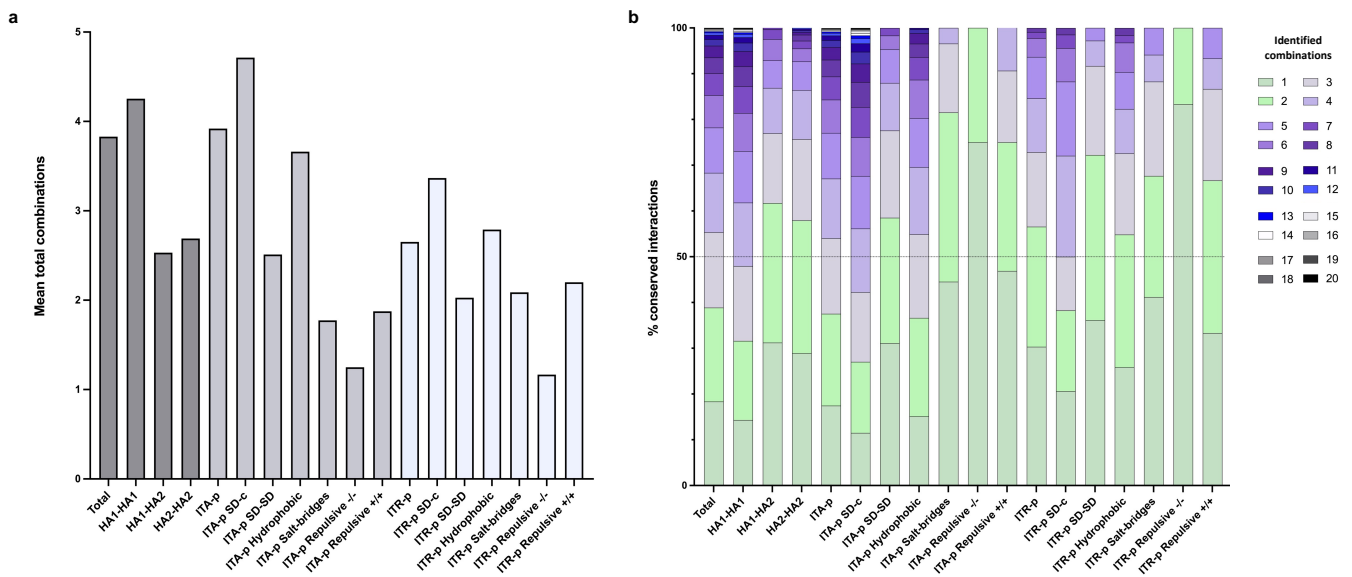
### **3.5 Annex 4: ITR-p interactions are less tolerant to mutations that preserve the interaction type**

Conserved interactions were identified in across HA classifications (**Annex 1 – Extended data table 2**). We explored the ability of residues involved in these conserved interactions to mutate while conserving the interaction type.

Among conserved interactions, we assessed mutational tolerance as the mean number of detected amino acid combinations allowing to preserve the interaction type. HA<sub>1</sub>-HA<sub>2</sub> and HA<sub>2</sub>-HA<sub>2</sub> together with ITR-p interactions showed the lowest tolerance compared to HA<sub>1</sub>-HA<sub>1</sub> and ITA-p interactions (**Annex 4 - figure 1**). Indeed, frequencies of conserved interactions with no more than two detected combinations maintaining the interaction were 31.6%, 37.5% and 38.9% for HA<sub>1</sub>-HA<sub>1</sub>, ITA-p and total interactions, respectively, but 61.7%, 57.9% and 56.6% for HA<sub>1</sub>-HA<sub>2</sub>, HA<sub>2</sub>-HA<sub>2</sub> and ITR-p interactions, indicating that more essential functions are likely to be associated with the latter and more restricted possibilities in terms of mutation preserving the function (**Annex 4 - figure 1**).

Residues involved in ITR-p interactions seem to be less susceptible to mutate than ITA-p interactions, probably due to limited possibilities to meet the constraints imposed at the protomer interface.

Annex 4 - Figure 1. Amino-acids variability amongst conserved interactions.



1

**Annex 4 - Figure 1. Amino-acids variability amongst conserved interactions.**

a. Bar charts representing the mean value of total identified amino-acids combinations per conserved interaction regarding types and their distribution throughout the protein. Interactions between two amino-acids were characterized as conserved if the two amino-acids maintain the precise interaction type, despite potential mutation, and if the interaction is symmetrically retrieved within or between each of the three protomers.

b. Bar charts representing for each interaction types and locations, the number of different combinations of amino acids enabling the interaction type to be conserved as a function of the percentage of total conserved interactions according to the condition.

### 3.6 Annex 5: Influence of the conserved salt-bridges on positioning of the b-loop

Previous structural analyses of various subtypes<sup>3</sup> showed that positioning of the interhelical b-loop and the shape of its c-terminus turn, which is influenced by interactions with the 110-helices, are clade specific factors.

The HAs from group 2 displayed a sharp b-loop turn, but the b-loops were positioned differently for clades 3 and 7 HAs, in relation with variations in the formation of the salt-bridges complex (**Annex 5 - figure 1**). Stabilized by its interaction with E81<sub>2</sub>, R76<sub>2</sub> generates more steric hindrance in clade 3 HAs and constrains the top of the  $\alpha$ -helices to adopt a relaxed conformation which allows formation of the ITR-p HA<sub>2</sub>-HA<sub>2</sub> (E74<sub>2</sub>-R76<sub>2</sub>) salt-bridges, whereas Q76<sub>2</sub> generates less steric hindrance thus preventing the formation of the salt-bridges in clade 7 HAs (**Annex 5 - figure 1b, 1c, 1d**). Therefore, the SB complex seems to influence positioning of the b-loops.

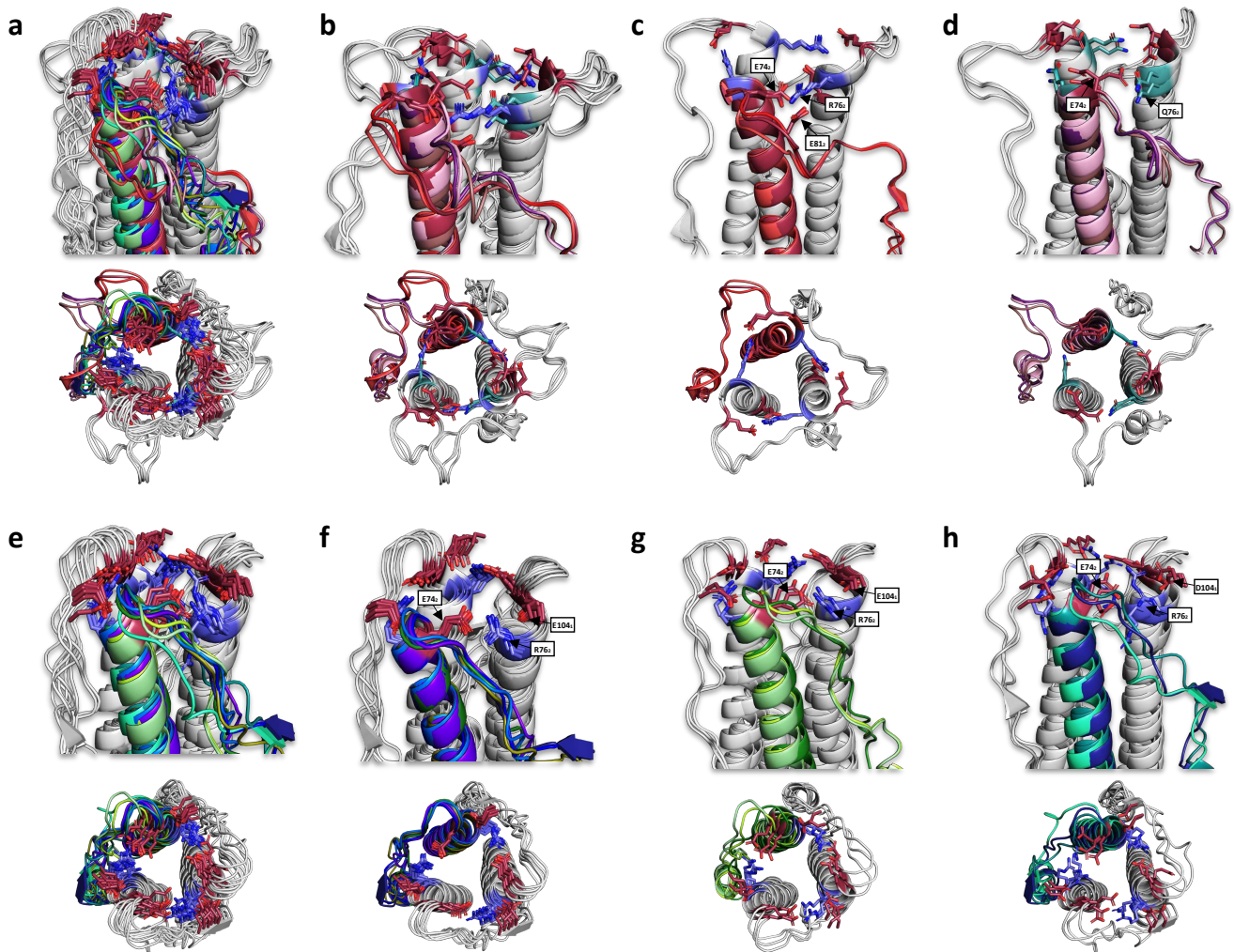
Regarding group 1 HAs, the turn and positioning of the b-loops were identical for clade 1 and clade 9 HAs, for which the ITR-p salt-bridges were identically arranged (**Annex 5 - figure 1f**).

Although the SB complex is formed in all clade 11 HAs, we observed a slight difference in the b-loop positioning in H13 and H16 while H11 (dark green) displayed the same b-loop arrangement as in clade 1 HAs and H8, thus indicating that other factors can influence its position (**Annex 5 - figure 1f, 1h**). Interestingly, clade 17 structures offer additional insights on the impact of the SB complex on the position of the b-loops. Despite the E104<sub>1</sub>D substitution and its subsequent impact on the spatial arrangement of the ITR-p HA<sub>2</sub>-HA<sub>2</sub> (E74<sub>2</sub>-R76<sub>2</sub>) salt-bridge adopted in H18, the turn and positioning of the b-loop was similar to that of in clade 1, clade 9 and H11 HAs (**Annex 5 - figure 1h**). In contrast, the complex was

not formed in H17, and a different positioning of H17 b-loops (cyan) was observed compared to clade 1, clade 9, H11 and H18 (**Annex 5 - figure 1h**).

The formation of the SB complex thus seems to influence the positioning of the b-loop, which thus might influence the CC process.

Annex 5 - Figure 1. Visualization of the b-loop turn and positioning of the b-loop according to classifications.



**Annex 5 - Figure 1. Visualization of the B-loop turn and positioning of the b-loop according to classifications.**

Subtype representatives of trimeric structure of HA<sub>2</sub> sub-units, represented in cartoon mode, were aligned to visualize the influence of the ITR-p salt-bridges complex on the b-loop turn and the positioning of the b-loop. Alignments were performed with the *cealign* command of PyMOL software (v2.4.1). One protomer is colored to distinguish subtypes. Front view and top view are represented for the following structure alignments: a. all HAs, b. group 2 HAs, c. clade 3 HAs, d. clade 7 HAs, e. group 1 HAs, f. clade 1 HAs (+ H8 which is the only resolved structure from clade 9), g. clade 11 HAs, and h. clade 17 HAs. Positions involved in the complex are colored as follows: E104<sub>1</sub> and E74<sub>2</sub> in red, R76<sub>2</sub> in blue and Q76<sub>2</sub> in cyan and indicated in c, d, f, g, h.

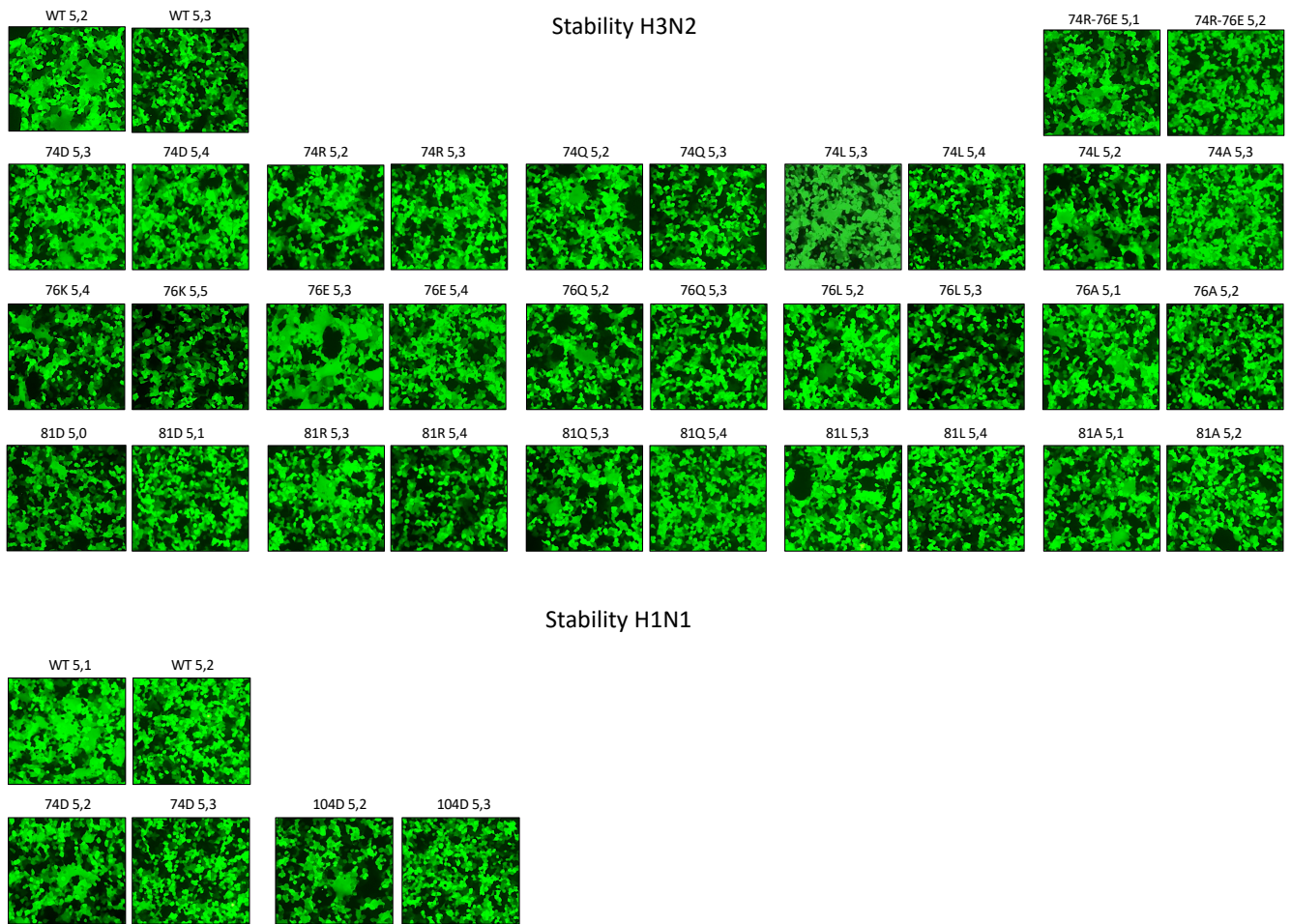


### 3.7 Annex 6: Influence of the SB complex on HA stability

Quantitative fusion assays, upon exposure to highly acidic pH (4.6), showed that mutations of the residues involved in the SB complex had no significant impact on fusion activity in the H3N2 context, whereas, in the H1N1 context, fusion activity was abolished for most mutants. Two conservative mutants, E104D<sub>1</sub> and E74D<sub>2</sub>, were functional but displayed a substantially reduced fusion activity as compared to WT (**Annex 1 - figure 3**).

In order to evaluate the influence of the SB complex on stability of the HA we determined the pH of fusion of the mutant HAs able to induce fusion of membranes and the formation of syncytia (**Annex 6 - figure 1**). HA-expressing cells were treated with exogenous TPCK-trypsin and exposed to pH values ranging from 4.6 to 6.0, with 0.1 increments. For both subtypes, mutations only caused a slight modification of the protein acidic stability compared to the respective WT protein (**Annex 6 - figure 1**). These results indicated that there is no major implication of the SB complex in regulating the pH of fusion, which is in line with the presence of the SB complex in the HA of viruses infecting different hosts with different optimal pH of fusion.

Annex 6 - Figure 1. Stability of viable mutants.



**Annex 6 - Figure 1. Stability of fusion competent HAs with mutations in the SB complex.**

HEK-293T cells were transfected with plasmids expressing the H3 (panel a.) or H1 (panel b.) HAs and co-transfected with a GFP expressing plasmid for visualization. Experiments were performed with WT or mutant HAs as indicated.

Transfected cells were treated with TPCK-trypsin and exposed to acidic pH from 4.6 to 6 with 0.1 increments. The ability to induce fusion was validated by the detection of formed syncytia.

For each mutant, the left panel shows the cells after exposure to the highest pH generating syncytia, and the right panel shows the cells after exposure to the lowest pH not generating syncytia.

### 3.8 Annex 7: Tracking of amino-acid interactions upon conformational change

To gain insights on interaction networks behavior during the CC, we analyzed the described high resolution structures of conformational change (CC) intermediates (H3; PDB accession codes: 6Y5G (neutral pH), 6Y5H (state 1), 6Y5I (state 2), 6Y5J (state 3), 6Y5K (state 4))<sup>114</sup>.

State after state analysis revealed that the total interaction count remained stable throughout the CC (**Annex 1 - Extended data table 3**). The HA<sub>1</sub>-HA<sub>1</sub> interactions count increased slightly (1237-1368), while the HA<sub>1</sub>-HA<sub>2</sub> interactions number decreased substantially (168-56) and that of HA<sub>2</sub>-HA<sub>2</sub> decreased slightly (508-459). It is noteworthy that fewer HA<sub>2</sub>-HA<sub>2</sub> interactions (215) were calculated in state 3 due to disorder preventing structure resolution of the HA<sub>2</sub> membrane-proximal section. This apparent loss was no longer observed in state 4 where the HA<sub>2</sub> structure is entirely resolved. Moreover, the number of ITR-p interactions decreased progressively (200-108) while ITA-p interactions remained essentially stable (1713-1775).

In state 1, the structure seems indistinguishable from the state at neutral pH<sup>114</sup>. However, modifications of the interaction network were detected immediately after acidification (**Annex 1 - Extended data table 3**). For electrostatic interactions, a significant loss of ITA-p HA<sub>1</sub>-HA<sub>1</sub> salt-bridges (-21%) was observed. Moreover, HA<sub>1</sub>-HA<sub>2</sub> links were weakened as ITR-p HA<sub>1</sub>-HA<sub>2</sub> salt-bridges decreased (-50%) while the few ITA-p HA<sub>1</sub>-HA<sub>2</sub> (+200%) and ITR-p HA<sub>2</sub>-HA<sub>2</sub> (+800%) -/- repulsions increased.

In state 2, HA<sub>1</sub> sub-units dilation and slight rotation initiate the CC<sup>114</sup>. We detected the complete loss of ITR-p HA<sub>1</sub>-HA<sub>1</sub> interactions (-100%), likely due to increased solvent access between HA<sub>1</sub> sub-units and its contact with hydrophobic HA<sub>1</sub>-HA<sub>1</sub> interactions, which represent 40% of total interactions. Moreover, HA<sub>1</sub>-HA<sub>2</sub> bonds further decreased (-20%),

especially ITA-p HA<sub>1</sub>-HA<sub>2</sub> salt-bridges (-60%), SD-C (-45%) and SD-SD (-33%) hydrogen interactions (**Annex 1 - Extended data table 3**).

In state 3, just before the major CC and coinciding with prior loss of ITR-p HA<sub>1</sub>-HA<sub>1</sub> interactions, HA<sub>1</sub> subunits are further rotated and dilated thereby increasing acidic solvent access<sup>114</sup>. We observed a transient increase of hydrophilic ITA-p HA<sub>1</sub>-HA<sub>1</sub> hydrogen SD-SD (+12%), hydrogen SD-C (+27%), and salt-bridges (+118%) along with the transient formation of ITR-p HA<sub>1</sub>-HA<sub>2</sub> salt-bridges, probably as hydrogen bonds due to carboxylic groups protonation (**Annex 1 - Extended data table 3**). Still in state 3, solvent penetration inside the trimer applied an outward force<sup>114</sup> and acidified the core, thus inducing major rearrangements of buried interactions, and the HA<sub>2</sub> membrane-proximal section was not resolved due to structural disorder<sup>114</sup>. The number of HA<sub>1</sub>-HA<sub>2</sub> (-72%) and HA<sub>2</sub>-HA<sub>2</sub> (-74%) hydrophobic interactions decreased. They were not (HA<sub>1</sub>-HA<sub>2</sub>), or only partially (HA<sub>2</sub>-HA<sub>2</sub>), restored in state 4 where the membrane-proximal region is resolved, in line with their actual disruption in state 3. Moreover, the number of ITR-p HA<sub>2</sub>-HA<sub>2</sub> (-61%) interactions substantially decreased, with almost no restoration observed in state 4. ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges, which accounted for 32% of ITR-p HA<sub>2</sub>-HA<sub>2</sub> interactions and 46% of salt-bridges in state 2, permanently decreased by -74% (**Annex 1 - Extended data table 3**). In contrast, ITA-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges were restored in state 4, probably as hydrogen bonds. Furthermore, a marked reduction of overall -/- repulsions (-64%) was observed throughout the protein. Our data indicate that, helped by HA<sub>2</sub>-HA<sub>2</sub> +/+ repulsions, disruption of ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges, mostly located at the membrane-proximal domain interface (**Annex 1 - Extended data Fig. 8**), likely initiates the observed disorder within the membrane-proximal region that releases the fusion peptide and induces the CC through hydrophobic forces<sup>114</sup>.

In state 4, the drastic CC has occurred and the membrane-proximal section is resolved<sup>114</sup>. An outward rotation, slight distancing of HA<sub>2</sub> and straightening of the C-helices are observed, concurring with prior decrease in ITR-p HA<sub>2</sub>-HA<sub>2</sub> interactions. We observed an important increase of ITA-p HA<sub>2</sub>-HA<sub>2</sub> -/- repulsions (+300%) along with ITA-p HA<sub>1</sub>-HA<sub>2</sub> (+166%) and ITA-p HA<sub>2</sub>-HA<sub>2</sub> (+260%) salt-bridges, probably as hydrogen interactions due to protonation (Table 2). An increase of ITA-p HA<sub>1</sub>-HA<sub>2</sub> (+110%) and ITA-p HA<sub>2</sub>-HA<sub>2</sub> (86%) hydrogen interactions was also observed. Furthermore, increased numbers of ITA-p HA<sub>2</sub>-HA<sub>2</sub> (+343%) and ITR-p HA<sub>2</sub>-HA<sub>2</sub> (+56%) hydrophobic interactions were detected, consistent with the irreversible extension of the trimeric coiled-coil HA<sub>2</sub> stems<sup>114</sup>. Importantly, ITR-p HA<sub>2</sub>-HA<sub>2</sub> +/+ repulsions were formed at the membrane proximal domain area and maintained the straightening of the C-helices where most ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges were previously found (**Annex 1 - Extended data Fig. 8**).

A second set of H3 intermediates has been reported<sup>106</sup>. Several features make the two sets different: the snapshots, strains, pH, presence of antibody, and detailed states. Nonetheless, the interactions variation trends during the CC are highly similar, as shown by heatmaps, highlighting the reproducibility and significance of interaction modulations during the CC (**Annex 7 – Table 1**)

Overall, our computational tracking of physicochemical interactions provides enhanced comprehension of the critical role of interaction networks as CC molecular drivers.

Annex 7 - Table 1. Comparisons of the interaction modulation trends between the two sets of CC intermediates structures

PDB accession code	6Y5G				6Y5H				6Y5I				6Y5J				7K39				7K37				7K3A				7K3B															
	Pre-fusion (PF)	state 1	state 2	state 3	Pre-fusion (PF)	state 1	state 2	state 3	Pre-fusion (PF)	state 1	state 2	state 3	Pre-fusion (PF)	state 1	state 2	state 3	Pre-fusion (PF)	state 1	state 2	state 3	Pre-fusion (PF)	state 1	state 2	state 3	Pre-fusion (PF)	state 1	state 2	state 3	Pre-fusion (PF)	state 1	state 2	state 3	Pre-fusion (PF)	state 1	state 2	state 3								
State	1237	1270	1286	1430	168	156	125	62	1778	1678	1590	1624	272	252	195	180	743	726	655	259	306	291	231	193	2487	2365	2209	1870	1747	1645	1569	1596	231	222	168	148	509	498	472	126	2793	2856	2440	2463



Chapter **4.** Discussion

Evolution of influenza viruses generated tremendous diversity which led to the existing HA classifications. Although the HA has been extensively studied, research has been mainly focused on antigenic properties, structural comparisons and phylogenetic analyses, whereas inter-amino acid interactions have not been thoroughly analyzed. However, inter-amino acid interactions provide the structural basis for protein function so their arrangement must be precisely adapted. The pre-fusion state of the HA is formed by intra-protomer (ITA-p) and inter-protomer (ITR-p) interactions distributed within or between each of the two sub-units (HA<sub>1</sub> and HA<sub>2</sub>). Moreover, since the HA pre-fusion state is metastable, rearrangements of the structure at acidic pH are driven by the modulation of interaction networks. Therefore, we performed a thorough analysis of HA physicochemical interactions, based on 269 trimeric pre-fusion structures, detailing the distribution, within and between protomers and HA<sub>1</sub>/HA<sub>2</sub> sub-units, of six interaction types at atomic scale: hydrophobic interactions, sidechain-corechain hydrogen (SD-C) interactions, sidechain-sidechain hydrogen (SD-SD), salt-bridges, anion-anion (-/-) repulsions and cation-cation (+/+) repulsions (**Annex 1 - Fig.1**).

Substantial differences were identified between clades regarding interaction networks composition and distribution, especially regarding inter-protomer interactions, hydrophobic, salt-bridges and repulsions. Using principal component analyses, we uncovered clade-dependent physicochemical profiles, with clade 1 and clade 3 HAs displaying distinct interaction networks, and clade 7 HAs exhibiting intermediate features with slight intra-clade segregation (**Annex 1 - Fig.1d**).

Nevertheless, the unveiled distinct physicochemical profiles are compatible with the folding of highly similar trimeric metastable structures carrying identical functions. More precisely, these distinct physicochemical profiles fulfill different requirements such as preserving the overall pre-fusion structure, preserving the metastable properties and triggering CC initiation



upon exposure to acidic pH while also maintaining all other vital properties for function and infectivity (cleavability, attachment to receptors, acidic stability...).

As interactions are inter-dependent, amino acid substitutions must be compatible with interaction networks and the overall physicochemical profile while preserving the overall structure and function of the protein. As a consequence, possibilities for mutation emergence are restricted as the underlying physicochemical profile of interaction networks applies evolutionary constraints. A better understanding of the interaction networks and their modulation may thus allow enhanced prediction of the effect of mutations and better tracking of HA evolution.

Moreover, these distinct physicochemical profiles strongly suggest distinct molecular drivers of the CC.

To gain more insights into these physicochemical strategies, we examined the relationships between interaction networks in pre-fusion structures and HA stability (**Annex 3**). The HAs of human viruses are more stable than HAs of avian viruses and stability has been shown to influence viral properties such as inter-species transmission, host specific adaptation, pathogenicity, and airborne transmission<sup>38,133,134</sup>. We observed a precise balance between destabilizing repulsive interactions and stabilizing ITA-p/ITR-p interactions, that correlates with HA subtype-specific stability values. However, the different profiles did not seem to be restrictive for species-barrier crossing and further adaptation (**Annex 1 - Extended data Fig. 3**). Changes in stability that could contribute to species-barrier crossing and host adaptation thus involve slight and precise adjustments of interactions balances, whose effects must be consistent with the overall physicochemical profile.

The HA conformational change (CC) triggered during endosomal acidification is a key step for virus entry into cells. It was suggested that the CC involves a wide molecular distribution of impacts on interactions in response to acidification<sup>6</sup>. However, no extensive description of interactions behavior upon acidic pH exposure has been available to date.

By monitoring inter-atomic interactions throughout the two available sets of intermediary states structures of the CC (H3)<sup>106,114</sup>, we showed considerable and comparable interactions rearrangements (**Annex 1 - Extended data table 3, Annex 7**).

Previous studies suggested HA<sub>1</sub> sub-units dilation is caused by repulsive forces at the HA<sub>1</sub> interface<sup>118,148</sup>. Analysis of histidine-cation repulsions location (within 4 structures) led to the identification of a cluster lying at the HA<sub>1</sub> interface, that was suggested to be implicated in HA<sub>1</sub> distancing<sup>148</sup>. However, no ITR-p HA<sub>1</sub>-HA<sub>1</sub> repulsions were detected in our analysis, and the total number of repulsions remained stable throughout the CC, supporting their main role in preventing over-stabilization of the pre-fusion state.

Immediately after acidification, we detected the disruption of ITA-p HA<sub>1</sub>-HA<sub>1</sub> and ITR-p HA<sub>1</sub>-HA<sub>2</sub> salt-bridges. Resulting slight rearrangements may have induced the formation of the detected -/- repulsions, especially ITA-p HA<sub>1</sub>-HA<sub>2</sub> -/- repulsions. Assisted by initially present ITA-p HA<sub>1</sub>-HA<sub>1</sub> ++ repulsions, these destabilizations translated in subtle HA<sub>1</sub> restructuring<sup>114</sup>. This correlates with experimental reports demonstrating the disruption of ITR-p HA<sub>1</sub>-HA<sub>2</sub> links and results of structural analyses showing reversible HA<sub>1</sub> rearrangements caused by salt-bridges reorganization within HA<sub>1</sub> and at the HA<sub>1</sub>-HA<sub>2</sub> interface<sup>117,168</sup>. Indeed, both salt-bridges and -/- repulsions can be reversibly disrupted due to carboxyl groups protonation upon acidification and deprotonation upon pH neutralization<sup>148,166-168</sup>. It also correlates with

the recently reported low-pH structure of the HA<sub>0</sub> precursor, which recovered its neutral-pH state upon pH neutralization<sup>6</sup>.

These electrostatic rearrangements likely enabled a slight increase in solvent access at the HA<sub>1</sub> interface<sup>107,114,117</sup>. Hence, solvent contacts with HA<sub>1</sub>-HA<sub>1</sub> hydrophobic interactions, which surprisingly are three times more abundant than hydrophobic HA<sub>2</sub>-HA<sub>2</sub> interactions, caused outward forces that disrupted all ITR-p HA<sub>1</sub>-HA<sub>1</sub> interactions and resulted in the observable distancing of HA<sub>1</sub> sub-units in state 2<sup>114</sup> (**Chapter 1 - Fig. 10**). This is consistent with experimental results demonstrating CC inhibition when ITR-p HA<sub>1</sub>-HA<sub>1</sub> disulfide bonds crosslink the HA<sub>1</sub> sub-units<sup>200</sup>.

Furthermore, we observed progressive disruption of overall HA<sub>1</sub>-HA<sub>2</sub> links, especially hydrophobic interactions, likely due to solvent accessibility, which is consistent with reports of rotation of HA<sub>1</sub> relative to HA<sub>2</sub> and increased solvent accessibility of the b-loop<sup>114,117,168</sup> (**Chapter 1 - Fig. 10**). Indeed, low-pH structure of the HA<sub>0</sub> precursor, indicative of a potential CC intermediate, showed that the b-loop is unleashed upon HA<sub>1</sub> dilation and undergoes its loop-to-helix transition, thus partially extending the HA<sub>2</sub> trimeric stem<sup>6</sup>.

Following HA<sub>1</sub> dilation and rotation, buried areas within the trimer become highly solvent accessible<sup>107,114,117</sup>. This causes acidification of the core and important microenvironmental changes resulting in the protonation of initially negatively charged carboxyl groups<sup>10,160,162,166,167</sup>. Importantly, we detected substantial disruption of ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges, which amplifies destabilization mediated by initially present ITA-p HA<sub>2</sub>-HA<sub>2</sub> +/- repulsions and gives rise to subsequent formation of ITR-p HA<sub>2</sub>-HA<sub>2</sub> ++ repulsions at the same location (**Annex 1 - Extended data Fig. 8**). Together with the disruption of ITR-p HA<sub>2</sub>-HA<sub>2</sub> hydrophobic interactions, these electrostatic rearrangements explain the straightening of the membrane-proximal area of HA<sub>2</sub> sub-units and their distancing at acidic pH, enabling

the complete release of the hydrophobic section of the fusion peptide originally buried at the HA<sub>2</sub>-HA<sub>2</sub> interface<sup>6,106,107,114,117</sup> (**Annex 1 - Extended data Fig. 8**). Moreover, given the role of salt-bridges in restricting flexibility and motion as well as acting as gatekeeper against unfolding, the disruption of ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges correlates with the observed disorder and structural reorganization of the membrane-proximal area<sup>12,114,187,199</sup>. Consecutively, solvent contact with the highly hydrophobic residues of the fusion peptide drives the complete coiled-coil extension<sup>6</sup>. The uncovered critical disruption of ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges is thereby in agreement with a previous study demonstrating that an H3 truncated HA<sub>2</sub> trimer, lacking HA<sub>1</sub> sub-units, retains its pre-fusion form at neutral pH but undergoes the CC at acidic pH<sup>188</sup>. Nonetheless, it is noteworthy that the unveiled distinct physicochemical profiles suggest different CC molecular drivers.

Conservation is most relevant when observed in a highly variable context because the accumulation of substitutions is limited to mutations that preserve protein function. Therefore, it may be possible to take advantage of HA variability to identify markers of vital functions that cannot vary without altering essential properties.

The identification of broadly-neutralizing antibodies targeting conserved epitopes of the stalk domain of HA (HA<sub>2</sub>) provided new hopes for the development of universal treatments and new strategies recently emerged based on antibodies targeting conserved buried epitopes, conserved epitopes at the protomer interface and non-neutralizing antibodies mediating protection via Fc-mediated functions<sup>140,179–181,201–208</sup>. However, resistance mutations have already been reported<sup>72</sup>. In addition, as evidenced by annually recurring influenza epidemics, pandemic risk in the context of zoonotic outbreaks, emergence of antiviral resistant viruses, and escape from broadly neutralizing antibodies, enhanced tools

are needed to effectively tackle viral variability<sup>15,19,82,192,193</sup>. In this context, we decided to study conserved interactions across HA classifications.

A previous study identified ITA-p histidine-cation and -/- repulsions on four HA structures and evaluated the conservation of involved residues based on a sequence alignment<sup>148</sup>. However, the conservation of these residues within sequences does not necessarily ensure their involvement in the described repulsions.

Moreover, invariant amino acids are not inevitably essential for protein function<sup>171,172</sup>. As interaction networks composition and arrangement influence protein folding, and thereby its function, a characterized interaction is not likely to be conserved during evolution if it is not of critical functional importance. We identified conserved interactions, with levels ranging from intra-subtype to globally conserved (**Annex 1 - Extended data table 2**). Strikingly, inter-clade disparities were observed for specific interactions according to type/location, in support of the unveiled clade-dependent physicochemical profiles. Given HA intrinsic variability, our results implied that broadly conserved interactions are likely critical for HA function despite differences in physicochemical profiles.

It was essential to validate the functional importance and limited mutational tolerance of conserved interactions. Therefore, we focused on inter-protomer (ITR-p) salt-bridges interactions due to their: (i) potential influence in folding and maintenance of the quaternary structure of HA trimers; (ii) pH sensor properties enabling them to respond to acidic pH exposure; (iii) role in acting as a gatekeeper against partial unfolding and in restricting protein motion<sup>9-12</sup>.

We identified a buried complex, located in the vicinity of the b-loop, consisting of two connected ITR-p salt bridges that we will refer to as “the SB complex” (**Annex 1 - Fig. 1**). The

first one is the ITR-p HA<sub>2</sub>-HA<sub>2</sub> (E74<sub>2</sub>-R76<sub>2</sub>) salt-bridge, strictly conserved among HAs from all clades except clade 7 HAs and H17 (**Annex 1 - Fig. 1b**). The second is the ITR-p HA<sub>1</sub>-HA<sub>2</sub> (E104<sub>1</sub>-R76<sub>2</sub>) salt-bridge, conserved among almost every subtype of group 1 HAs, whereas the ITR-p HA<sub>2</sub>-HA<sub>2</sub> (R76<sub>2</sub>-E81<sub>2</sub>) salt-bridge is strictly conserved among clade 3 HAs (**Annex 1 - Fig. 1b**). Amino acids involved in the complex were strictly conserved in thousands of sequences of circulating viruses.

In the H1 context, non-conservative mutations at any of the positions involved in the complex resulted in loss of function and infectivity, highlighting the interdependence of these ITR-p salt-bridges and the vital role of the complex (**Annex 1 - Fig. 3, Fig. 4**). The charge-conservative substitutions E104<sub>1</sub>D and E74<sub>2</sub>D were the only viable mutations, but their fusion efficiency was dramatically impaired and the produced viruses showed reduced growth *in vitro*, presumably because of the impact of the shortened sidechain on the spatial arrangement of the complex, as observed in clade 17 HAs. The charge-conservative R76<sub>2</sub>K mutation was however not viable, highlighting the importance of residue-specific physicochemical properties. Indeed, arginine is a key connector that allows for the adoption of various geometrical arrangements and is thus frequently found in complex salt-bridges and at the protomer interface<sup>9,163</sup>. Moreover, arginine is always charged regardless of the environment whereas the pKa of lysine can undergo serious shift, especially when buried, and thus be uncharged at neutral pH<sup>163-165</sup>.

Remarkably, in the H3 context, all mutants were able to induce cell fusion and yielded infectious viruses regardless of the impact of the substitutions on the SB complex (**Annex 1 - Fig. 3, Fig. 4**).

Altogether, assessment of the functional importance of the conserved ITR-p SB complex in different environments highlighted that the underlying clade-dependent physicochemical

profile of interaction networks has an impact on how function is performed, also explaining why mutations have different effects in different physicochemical contexts.

As indicated by the substantially limited mutational tolerance of involved positions in the H1 context, physicochemical profiles influence the mutational tolerance of specific residues, thereby constraining HA evolution.

Our results therefore correlate with the *in vitro* study of the evolutionary pathways of A/WSN/1933 (H1N1) and A/Perth/16/2009 (H3N2) strains<sup>176</sup>. These experiments showed no mutational tolerance of H1 at the positions involved in the ITR-p salt-bridges complex, except for the very minority presence of the R76<sub>2</sub>W mutant which may be a non-functional HA originating from genetic complementation enabled by co-infections<sup>176</sup>. In contrast, in the H3 context, a diversity of residues was tolerated at each of the positions involved in the complex with no preference for charged residues despite their conservation in nature.

The SB complex links the amino termini of the  $\alpha$ -helices of HA<sub>2</sub> and the carboxyl termini of the  $\beta$ -loops, whose rearrangement into helices is one of the major events of the CC triggered at acidic pH. In addition, in the case of group 1 HAs, the second ITR-p salt-bridge (E104<sub>1</sub>–R76<sub>2</sub>) stabilizes contacts between the 110-helices (HA<sub>1</sub>) and the carboxyl termini of the  $\beta$ -loops (HA<sub>2</sub>), which are involved in guiding the fusion peptides toward the cellular membrane bound by HA<sub>1</sub><sup>209</sup>. ITR-p interactions are vital for the folding and maintenance of the symmetrical biological assembly of HA and salt-bridges can be disrupted in the acidic environment of mature endosomes. Moreover, buried salt-bridges complexes provide additional structural stability by electrostatically protecting the charges of involved residues from the hydrophobic environment of the core<sup>10,160</sup>. This may explain the identical geometrical arrangement in space of side chains of residues involved in the buried ITR-p salt-

bridges complex across structures (**Annex 1 - Fig. 1**). Furthermore, complex salt-bridges stabilize transient folding states by acting as a gatekeeper against unfolding that would disrupt the optimal spatial arrangement of the interactions and restrict flexibility and movement of the surrounding area<sup>11,12,156</sup>. Hence, the SB complex might play a role in preventing the loop-to-helix transition of the b-loop at neutral pH in the H1 context.

This was supported by the emergence of the R76<sub>2</sub>G mutation upon passaging of the mutant E104<sub>1</sub>D on A549 cells (**Annex 1 - Fig. 4d**). The E104<sub>1</sub>D mutant multiplied in MDCK cells but was unable to infect A549 cells. This could be attributed to its very low fusion efficiency and the less robust endosomal acidification of A549 cells compared to MDCK cells<sup>139</sup>. Restoration of some growth potential in A549 cells by the R76<sub>2</sub>G mutation might be attributed to the complete disruption of the inefficient mutated salt-bridges complex, thus permanently lifting its pH-dependent restrictive effect on flexibility and motion while bringing additional local flexibility due to the absence of side chain, however at heavy cost on viral fitness. Most importantly, two independent studies also emphasized the role of the ITR-p salt-bridges complex in conferring metastable properties. By exploring early fusion intermediates of H2, which has the same physicochemical profile as H1, these studies suggested that further movement of the b-loop would require the disruption of the ITR-p salt-bridges complex to enable the vital loop-to-helix transition of the b-loop at acidic pH<sup>168,186</sup>.

Given the strict conservation of the SB complex in structures sharing the H1 physicochemical profile and that of involved residues in sequences, the identical geometric arrangement of the complex in space, the highly similar distribution of ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges as well as the necessity to disrupt the complex to undergo the CC in H2 proteins we believe that our findings regarding the essential nature of the complex in H1 are likely to be extrapolated to



the other HAs sharing the H1 physicochemical profile, including viruses with high pandemic potential such as H5N1 viruses.

Together with our experimental results, the distribution of ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges within different subtypes revealed differences between HAs with distinct physicochemical profiles, indicating different behavior of those interactions (**Annex 1 - Extended data Fig. 7**). The only ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges present within HAs sharing the physicochemical profile of H1 were these of the SB complex, which further emphasized its potential importance in these subtypes (**Annex 1 - Extended data Fig. 7**).

The physicochemical profile of clade 7 HAs share characteristics with that of clade 1 and clade 3 HAs, but clade 7 HAs are closer to clade 3 HAs regarding ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges (**Annex 1 - Extended data Fig. 7**). A R76<sub>2</sub>Q substitution prevent the formation of the complex in clade 7 HAs. The non-conservative mutations introduced in H3 disrupt the SB complex and tend to create a similar situation as that in clade 7 HAs, with ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges being limited to the ones located at the membrane-proximal domain (**Annex 1 - Extended data Fig. 7**). Moreover, the positioning of their b-loops should be identical in clade 7 HAs and mutated H3 (**Annex 5**).

Furthermore, an ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges complex of two salt-bridges located at the membrane-proximal domain has been found to be conserved in clade 3 and clade 7 HAs, which emphasized their functional importance. The first salt-bridge of the complex, the ITR-p R127<sub>2</sub>-E131<sub>2</sub> HA<sub>2</sub>-HA<sub>2</sub> salt-bridge, is conserved in 97% and 97% of clade 3 and clade 7 HAs, respectively. The second, the ITR-p E131<sub>2</sub>-R163<sub>2</sub> HA<sub>2</sub>-HA<sub>2</sub> salt-bridge, is conserved in 95% and 92% of clade 3 and clade 7 HAs, respectively.

The importance of these interactions was further emphasized by the modulation of interactions during the CC (H3) (**Annex 1 - Extended data table 3**). Interestingly, the analysis of interaction networks based on fusion intermediate structures (H3) revealed that the membrane-proximal ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges described above are disrupted upon exposure to acidic pH, resulting in rearrangement of the membrane-proximal domain preceding the pivotal stem extension<sup>114</sup>. Moreover, HA<sub>1</sub>-truncated H3 trimers adopt the pre-fusion conformation and undergo the CC at acidic pH, indicating that metastable properties are not solely conferred by HA<sub>1</sub><sup>188</sup>. It thus implies that protonation of the HA<sub>2</sub> sub-units results in the disruption of some metastable properties. Therefore, in clade 3 and probably clade 7 HAs, disruption of the ITR-p HA<sub>2</sub>-HA<sub>2</sub> salt-bridges located at the membrane-proximal domain most likely act as molecular drivers of the CC by inducing the observed straightening of the c-helices and their distancing, thus enabling the complete extrusion of the fusion peptide buried at the HA<sub>2</sub> interface<sup>6,114,188,210</sup> (**Annex 1 - Extended data Fig. 8**). It will therefore be interesting to experimentally assess the functional importance of these salt-bridges in the clade 3 and clade 7 contexts.

Nevertheless, despite the non-essential nature of the SB complex *in vitro* in the H3 context, its strict conservation in circulating H3 viruses (structures and sequences) suggests that it could be important for the infectivity of clade 3 viruses *in vivo*, which remains to be evaluated.

ITR-p interactions are crucial to confer metastable properties as indicated by our experimental results obtained by mutating the SB complex in the H1 context but also by the progressive disruption of ITR-p HA<sub>1</sub>-HA<sub>2</sub> and HA<sub>2</sub>-HA<sub>2</sub>, including the conserved salt-bridges at

the membrane-proximal domain, and the complete disruption of ITR-p HA<sub>1</sub>-HA<sub>1</sub> during the CC in the H3 context.

In accordance, the fusion potential of the HA was shown to be inhibited when disulfide bonds crosslink the HA<sub>1</sub> sub-units<sup>200</sup>. Furthermore, CT149 and FI6 are two antibodies that were shown to neutralize both group 1 and group 2 HAs by cross-linking HA<sub>2</sub> sub-units<sup>211,212</sup>. Initially buried areas may prove accessible upon modulation of ITR-p interactions during acidification, which is why we did not filter out interactions based on their solvent-exposure in our study. Indeed, the increase of extracellular pH can trigger incomplete transient conformational change of HAs at the extracellular surface of infected cells. Due to acidic solvent exposure and subsequent ITR-p interactions modulation, the resulting “breathing” of activated HA (dilation of HA<sub>1</sub> sub-units) allows antibodies to bind initially buried cryptic epitopes that are transiently exposed<sup>213–216</sup>. In addition, immunization with HA in its post-fusion conformation generates cross-group protective antibodies that target initially occluded and highly conserved epitopes<sup>181</sup>. Seasonal vaccination induces the production of similar cross-reactive antibodies, targeting buried epitopes at the interprotomer interface, that are protective upon passive transfer in mice against lethal challenge infections by H1N1pdm09 or H3N2 viruses<sup>179</sup>. Moreover, it has been recently demonstrated that HA<sub>0</sub> precursors at the plasma membrane can undergo an incomplete CC upon exposure to acidic pH. The transition to this state is also reversible, which means that buried areas of HA<sub>1</sub> but also HA<sub>2</sub>, including their protomer interface, can become accessible<sup>6</sup>.

Pathogenic infection and the immune response could both mediate the necessary extracellular acidification to trigger the reversible breathing of activated HAs or incomplete CC of precursors. Activated immune cells such as neutrophils are reported to be able to release protons, thus lowering the extracellular medium<sup>183</sup>. In addition, as a consequence of

glycolysis, conversion of glucose to lactic acid at neutral pH generates extracellular acidification owing to the release of protons<sup>182</sup>. Several viruses, including viruses infecting airway epithelial cells, hijack the cell metabolism regulation machinery and cause an increase of the glycolysis rate<sup>217–219</sup>. More specifically, IAV infection induces increased extracellular glucose uptake, glycolysis rate and lactic acid release from the cell at the early stages of infection and thereby generate extracellular acidification, especially around the plasma membrane<sup>185,220–222</sup>. Additional effects of immune cells could also influence the extracellular pH as macrophages and dendritic cells undergo a strong increase of glycolysis following pathogens recognition by Toll-like receptors (TLRs)<sup>223,224</sup>. Furthermore, it has been recently shown that HAs located at the infected cell surface undergo the conformational change, even at neutral pH, due to the binding of Zn<sup>2+</sup> divalent cations<sup>225</sup>. It is known that divalent cations are vital for innate and adaptative immune responses as they are involved in immune cells maturation, proliferation and cytokine production, *inter alia*<sup>226,227</sup>. More specifically, zinc is known to regulate both innate and adaptative immune cell functions through signal transduction, leading to cell survival, proliferation and differentiation processes<sup>228,229</sup>. Thus, immune cell death at the vicinity of other infected cells, harboring cleaved HA at their surface, could release enough divalent cations to trigger the conformational change of some proteins.

Finally, initially buried areas of the HA can also be accessed inside the endosome by sialic acid-coupled drugs that are co-endocytosed with the virus and by drugs that bind the pre-fusion state of HA and relocate at a new binding site under acidic conditions<sup>184,230</sup>.

Our computational analysis was based on six interaction types. Yet, less canonical interactions could be analyzed to refine the described data such as  $\pi$ - $\pi$ , cation- $\pi$ , anion- $\pi$ ,

together with long-range interactions, and other features could also bring additional information such as interaction geometry, solvent accessibility, and steric hindrance<sup>159,231–234</sup>. Moreover, the location of the two partners of an interaction could be further specified by indicating the specific secondary structures to which the involved amino acids belong, rather than simply indicating their subunits. The analysis of such data could be helpful to refine the described physicochemical profiles of interaction networks to emphasize further differences and evolutionary constraints. Moreover, analysis of the behavior of these interaction types upon exposure to acidic pH could highlight additional molecular drivers of the CC, especially the anion- $\pi$  interactions that may be disrupted at acidic pH. Because of their pKa and complete protonation at acidic pH, it might also be of interest to analyze, upon acidification, the formation of electrostatic interactions involving histidine residues independently of other electrostatic interactions<sup>235</sup>.

Importantly, our results were limited by the dataset used in our study, specifically regarding the heterogeneous number of structures resolved per subtype. For instance, the three structures of bat viruses H17 and H18 are insufficient to compare their physicochemical profile of interaction networks with that of other subtypes and to obtain relevant data on conserved interactions. Moreover, results were also limited by the heterogeneity in strain diversity among structures of different subtypes. For example, the A/Aichi/2/1968 strain was rather overrepresented for the H3 subtype and resulted in greater conservation of interactions compared to the H1 subtype, which was more diversified in terms of strains for which HA is resolved.

Generalized access to single particle cryo-electron microscopy at atomic resolution together with artificial intelligence-based structural predictions (AlphaFold) will enable to provide

exponentially more data that will be crucial to increase size and diversity of the datasets and to identify new intermediate states of various subtypes to analyze<sup>117,189–191</sup>.

Monitoring the emergence of mutations in circulating viruses and their consequences on interaction networks according to the distinct physicochemical profiles would allow enhanced understanding of the different physicochemical strategies employed to preserve the structure and function. Moreover, further experimental data could deepen the process by characterizing any mutations that would preserve or disrupt HA function depending on the distinct physicochemical contexts. A high-throughput method of directed mutagenesis coupled with fusion assays or reverse genetics experiments would be relevant to accumulate such data. Eventually, it might be possible that artificial intelligence-based models trained on sufficiently large and high-quality datasets could effectively predict the impact of mutations on HA structure and function in the context of different physicochemical profiles.

Moreover, experimental data accumulation will also be critical to validate the functional importance of conserved interactions and the inability of the virus to mutate at involved positions. Hence, understanding how structure folding, metastable property and CC are enabled by distinct physicochemical strategies along with the functional validation of a wide range of such conserved interactions, covering the distinct physicochemical profiles, will not only be beneficial to predict viral evolution but could also be used for the development of antivirals or epitope-based vaccines with broad specificity against highly variable viruses and with reduced potential for resistance and escape.

Docking small molecules onto conserved interactions of validated functional importance could inhibit HA function through various mechanisms depending on the role of targeted interactions. For example, upon binding of such molecules, mediated destabilization or overstabilization of conserved salt-bridges that act as molecular drivers of the CC would

prevent their disruption at acidic pH. As a result, either the infectivity of the virus should be inhibited or the virus should mutate and become non-viable. A combination of molecules targeting several conserved and crucial interactions of the HA could therefore provide broad and non-escapable protection against IAVs. The characterization of HA-targeting molecules would also intrinsically decrease the potential for emergence of resistance as the virus would not be able to enter the cell and so to replicate and potentially mutate.

In addition, the screening of known B and T epitopes enriched in residues involved in pivotal conserved interactions, whose functional importance and low mutational tolerance have been validated by experiments, should be beneficial for the design of universal and non-escapable peptide-based vaccines. New vaccine strategies aim to elicit a strong humoral and/or cellular immune response against the HA<sub>2</sub> subunits rather than against the immunodominant HA<sub>1</sub> subunits. Triggering such immune response against key and conserved specific epitopes with low mutational tolerance could allow antibodies to inhibit HA function of a wide range of subtypes, likely by overstabilizing the fusion domain, but also to generate a strong cellular response against the most conserved epitopes with the lowest mutational tolerance. A combination of different peptides would be required for optimal effectiveness and further reduction of potential for escape.

In the context of continued antigenic drift of influenza viruses, the potential for emergence of resistance against antivirals, and pandemic risk, we need new solutions to efficiently tackle viral variability. Our results highlighted that HA evolution is limited by restricted possibilities in terms of viable mutations fitting interaction networks, according to distinct physicochemical profiles. Hence, based on further experimental data, enhanced predictions of functional consequences of mutations on the distinct HA physicochemical profiles of

interaction networks and on protein function should constitute an additional tool for surveillance of the evolution of IAVs.

Moreover, taking advantage of high variability to identify conserved interactions and validating their functional importance and mutational tolerance by extensive mutagenesis could offer new opportunities for the characterization of therapeutic targets and vaccine epitopes with broad specificity and low escape potential. It could also benefit pandemic preparedness by identifying critical targets in subtypes of concern beforehand emergencies.



# Bibliography

1. Skehel, J. J. & Wiley, D. C. Receptor Binding and Membrane Fusion in Virus Entry: The Influenza Hemagglutinin. *Annu. Rev. Biochem.* **69**, 531–569 (2000).
2. Salinas, V. H. & Ranganathan, R. Coevolution-based inference of amino acid interactions underlying protein function. *Elife* **7**, e34300 (2018).
3. Russell, R. J. *et al.* H1 and H7 influenza haemagglutinin structures extend a structural classification of haemagglutinin subtypes. *Virology* **325**, 287–296 (2004).
4. Nachbagauer, R. *et al.* A chimeric hemagglutinin-based universal influenza virus vaccine approach induces broad and long-lasting immunity in a randomized, placebo-controlled phase I trial. *Nat Med* **27**, 106–114 (2021).
5. Air, G. M. Sequence relationships among the hemagglutinin genes of 12 subtypes of influenza A virus. *Proc. Natl. Acad. Sci. U.S.A.* **78**, 7639–7643 (1981).
6. Garcia-Moro, E. *et al.* Reversible structural changes in the influenza hemagglutinin precursor at membrane fusion pH. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2208011119 (2022).
7. Wang, W.-C., Sayedahmed, E. E., Sambhara, S. & Mittal, S. K. Progress towards the Development of a Universal Influenza Vaccine. *Viruses* **14**, 1684 (2022).
8. Lee, N. & Hurt, A. C. Neuraminidase inhibitor resistance in influenza: a clinical perspective. *Current Opinion in Infectious Diseases* **31**, 520–526 (2018).
9. Janin, J., Miller, S. & Chothia, C. Surface, subunit interfaces and interior of oligomeric proteins. *J Mol Biol* **204**, 155–164 (1988).
10. Pahari, S., Sun, L. & Alexov, E. PKAD: a database of experimentally measured pKa values of ionizable groups in proteins. *Database* **2019**, baz024 (2019).
11. Meuzelaar, H., Vreede, J. & Woutersen, S. Influence of Glu/Arg, Asp/Arg, and

- Glu/Lys Salt Bridges on  $\alpha$ -Helical Stability and Folding Kinetics. *Biophysical Journal* **110**, 2328–2341 (2016).
12. Hinzman, M. W., Essex, M. E. & Park, C. Salt bridge as a gatekeeper against partial unfolding. *Protein Sci* **25**, 999–1009 (2016).
  13. World Health Organization. *Global influenza strategy 2019-2030*. (World Health Organization, 2019).
  14. Kilbourne, E. D. Influenza pandemics of the 20th century. *Emerg Infect Dis* **12**, 9–14 (2006).
  15. Taubenberger, J. K. & Morens, D. M. Influenza: The Once and Future Pandemic. *Public Health Rep* **125**, 16–26 (2010).
  16. Lampejo, T. Influenza and antiviral resistance: an overview. *Eur J Clin Microbiol Infect Dis* **39**, 1201–1208 (2020).
  17. Kalil, A. C. & Thomas, P. G. Influenza virus-related critical illness: pathophysiology and epidemiology. *Critical Care* **23**, 258 (2019).
  18. Teijaro, J. R. *et al.* Endothelial cells are central orchestrators of cytokine amplification during influenza virus infection. *Cell* **146**, 980–991 (2011).
  19. Krammer, F. *et al.* Influenza. *Nat Rev Dis Primers* **4**, 1–21 (2018).
  20. Paterson, D. & Fodor, E. Emerging Roles for the Influenza A Virus Nuclear Export Protein (NEP). *PLoS Pathog* **8**, e1003019 (2012).
  21. Zhang, Y. *et al.* Targeting Hemagglutinin: Approaches for Broad Protection against the Influenza A Virus. *Viruses* **11**, (2019).
  22. Wu, Y., Wu, Y., Tefsen, B., Shi, Y. & Gao, G. F. Bat-derived influenza-like viruses H17N10 and H18N11. *Trends in Microbiology* **22**, 183–191 (2014).
  23. Long, J. S., Mistry, B., Haslam, S. M. & Barclay, W. S. Host and viral determinants of influenza A virus species specificity. *Nat Rev Microbiol* **17**, 67–81 (2019).

24. Neumann, G., Chen, H., Gao, G. F., Shu, Y. & Kawaoka, Y. H5N1 influenza viruses: outbreaks and biological properties. *Cell Res* **20**, 51–61 (2010).
25. Shi, W., Shi, Y., Wu, Y., Liu, D. & Gao, G. F. Origin and molecular characterization of the human-infecting H6N1 influenza virus in Taiwan. *Protein Cell* **4**, 846–853 (2013).
26. Fouchier, R. A. M. *et al.* Avian influenza A virus (H7N7) associated with human conjunctivitis and a fatal case of acute respiratory distress syndrome. *Proc Natl Acad Sci U S A* **101**, 1356–1361 (2004).
27. Lam, T. T.-Y. *et al.* The genesis and source of the H7N9 influenza viruses causing human infections in China. *Nature* **502**, 241–244 (2013).
28. Peiris, M., Yam, W. C., Chan, K. H., Ghose, P. & Shortridge, K. F. Influenza A H9N2: Aspects of Laboratory Diagnosis. *J Clin Microbiol* **37**, 3426–3427 (1999).
29. Chen, H. *et al.* Clinical and epidemiological characteristics of a fatal case of avian influenza A H10N8 virus infection: a descriptive study. *The Lancet* **383**, 714–721 (2014).
30. Janke, B. H. Influenza A virus infections in swine: pathogenesis and diagnosis. *Vet Pathol* **51**, 410–426 (2014).
31. Ma, W., Kahn, R. E. & Richt, J. A. The pig as a mixing vessel for influenza viruses: Human and veterinary implications. *J Mol Genet Med* **3**, 158–166 (2008).
32. Koçer, Z. A., Jones, J. C. & Webster, R. G. Emergence of Influenza Viruses and Crossing the Species Barrier. *Microbiol Spectr* **1**, (2013).
33. Kuiken, T. *et al.* Host Species Barriers to Influenza Virus Infections. *Science* **312**, 394–397 (2006).
34. Ito, T. *et al.* Molecular basis for the generation in pigs of influenza A viruses with pandemic potential. *J Virol* **72**, 7367–7373 (1998).
35. Imai, M. & Kawaoka, Y. The role of receptor binding specificity in interspecies transmission of influenza viruses. *Curr Opin Virol* **2**, 160–167 (2012).

36. Webster, R. G., Sharp, G. B. & Claas, E. C. Interspecies transmission of influenza viruses. *Am J Respir Crit Care Med* **152**, S25-30 (1995).
37. Di Lella, S., Herrmann, A. & Mair, C. M. Modulation of the pH Stability of Influenza Virus Hemagglutinin: A Host Cell Adaptation Strategy. *Biophysical Journal* **110**, 2293–2301 (2016).
38. Russell, C. J., Hu, M. & Okda, F. A. Influenza Hemagglutinin Protein Stability, Activation, and Pandemic Risk. *Trends in Microbiology* **26**, 841–853 (2018).
39. Gambaryan, A. S. & Matrosovich, M. N. What adaptive changes in hemagglutinin and neuraminidase are necessary for emergence of pandemic influenza virus from its avian precursor? *Biochemistry Moscow* **80**, 872–880 (2015).
40. Neumann, G. & Kawaoka, Y. Transmission of influenza A viruses. *Virology* **479–480**, 234–246 (2015).
41. Subbarao, K. The Critical Interspecies Transmission Barrier at the Animal–Human Interface. *Trop Med Infect Dis* **4**, 72 (2019).
42. McKellar, J., Rebendenne, A., Wencker, M., Moncorgé, O. & Goujon, C. Mammalian and Avian Host Cell Influenza A Restriction Factors. *Viruses* **13**, 522 (2021).
43. Kim, H., Webster, R. G. & Webby, R. J. Influenza Virus: Dealing with a Drifting and Shifting Pathogen. *Viral Immunology* **31**, 174–183 (2018).
44. Schweiger, B., Zadow, I. & Heckler, R. Antigenic drift and variability of influenza viruses. *Med Microbiol Immunol* **191**, 133–138 (2002).
45. Barbezange, C. *et al.* Seasonal Genetic Drift of Human Influenza A Virus Quasispecies Revealed by Deep Sequencing. *Front Microbiol* **9**, 2596 (2018).
46. Nobusawa, E. & Sato, K. Comparison of the mutation rates of human influenza A and B viruses. *J Virol* **80**, 3675–3678 (2006).
47. Lyons, D. M. & Lauring, A. S. Mutation and Epistasis in Influenza Virus Evolution.

*Viruses* **10**, E407 (2018).

48. Yamayoshi, S. & Kawaoka, Y. Current and future influenza vaccines. *Nat Med* **25**, 212–220 (2019).

49. Steel, J. & Lowen, A. C. Influenza A virus reassortment. *Curr Top Microbiol Immunol* **385**, 377–401 (2014).

50. Russell, C. J. & Webster, R. G. The genesis of a pandemic influenza virus. *Cell* **123**, 368–371 (2005).

51. Taubenberger, J. K., Morens, D. M. & Fauci, A. S. The next influenza pandemic: can it be predicted? *JAMA* **297**, 2025–2027 (2007).

52. Gust, I. D., Hampson, A. W. & Lavanchy, D. Planning for the next pandemic of influenza. *Reviews in Medical Virology* **11**, 59–70 (2001).

53. Wendel, I., Matrosovich, M. & Klenk, H. D. SnapShot: Evolution of human influenza A viruses. *Cell Host Microbe* **17**, 416-416.e1 (2015).

54. Johnson, N. P. A. S. & Mueller, J. Updating the accounts: global mortality of the 1918-1920 ‘Spanish’ influenza pandemic. *Bull Hist Med* **76**, 105–115 (2002).

55. Oxford, J. S. Influenza A pandemics of the 20th century with special reference to 1918: virology, pathology and epidemiology. *Rev Med Virol* **10**, 119–133 (2000).

56. Wood, J. M. The 2009 influenza pandemic begins. *Influenza Other Respir Viruses* **3**, 197–198 (2009).

57. Smith, G. J. D. *et al.* Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature* **459**, 1122–1125 (2009).

58. Mostafa, A., Abdelwhab, E. M., Mettenleiter, T. C. & Pleschka, S. Zoonotic Potential of Influenza A Viruses: A Comprehensive Overview. *Viruses* **10**, 497 (2018).

59. World Health Organization. *Avian Influenza Weekly Update Number 769*. (World Health Organization, 2020).

60. World Health Organization. *Influenza at the human-animal interface*. vol. 211 (2015).
61. Li, Q. *et al.* Epidemiology of Human Infections with Avian Influenza A(H7N9) Virus in China. *New England Journal of Medicine* **370**, 520–532 (2014).
62. Cowling, B. J. *et al.* Comparative epidemiology of human infections with avian influenza A H7N9 and H5N1 viruses in China: a population-based study of laboratory-confirmed cases. *Lancet* **382**, 129–137 (2013).
63. Kain, T. & Fowler, R. Preparing intensive care for the next pandemic influenza. *Critical Care* **23**, 337 (2019).
64. Herfst, S. *et al.* Airborne transmission of influenza A/H5N1 virus between ferrets. *Science* **336**, 1534–1541 (2012).
65. Linster, M. *et al.* Identification, Characterization, and Natural Selection of Mutations Driving Airborne Transmission of A/H5N1 virus. *Cell* **157**, 329–339 (2014).
66. Imai, M. *et al.* Experimental adaptation of an influenza H5 HA confers respiratory droplet transmission to a reassortant H5 HA/H1N1 virus in ferrets. *Nature* **486**, 420–428 (2012).
67. Boikos, C. *et al.* Relative Effectiveness of the Cell-derived Inactivated Quadrivalent Influenza Vaccine Versus Egg-derived Inactivated Quadrivalent Influenza Vaccines in Preventing Influenza-related Medical Encounters During the 2018-2019 Influenza Season in the United States. *Clin Infect Dis* **73**, e692–e698 (2021).
68. Imran, M. *et al.* Relative Effectiveness of Cell-based Versus Egg-based Quadrivalent Influenza Vaccines in Children and Adolescents in the United States During the 2019–2020 Influenza Season. *The Pediatric Infectious Disease Journal* **41**, 769–774 (2022).
69. Wong, S.-S. & Webby, R. J. Traditional and New Influenza Vaccines. *Clin Microbiol Rev* **26**, 476–492 (2013).
70. Yoshida, R. *et al.* Cross-Protective Potential of a Novel Monoclonal Antibody

Directed against Antigenic Site B of the Hemagglutinin of Influenza A Viruses. *PLOS Pathogens* **5**, e1000350 (2009).

71. Corti, D. *et al.* Heterosubtypic neutralizing antibodies are produced by individuals immunized with a seasonal influenza vaccine. *J Clin Invest* **120**, 1663–1673 (2010).
72. Nath Neerukonda, S., Vassell, R. & Weiss, C. D. Neutralizing Antibodies Targeting the Conserved Stem Region of Influenza Hemagglutinin. *Vaccines* **8**, 382 (2020).
73. Impagliazzo, A. *et al.* A stable trimeric influenza hemagglutinin stem as a broadly protective immunogen. *Science* **349**, 1301–1306 (2015).
74. Yassine, H. M. *et al.* Hemagglutinin-stem nanoparticles generate heterosubtypic influenza protection. *Nat Med* **21**, 1065–1070 (2015).
75. Y, L., Jp, W. & Jr, S. Production and stabilization of the trimeric influenza hemagglutinin stem domain for potentially broadly protective influenza vaccines. *Proceedings of the National Academy of Sciences of the United States of America* **111**, (2014).
76. Liao, H.-Y. *et al.* Chimeric hemagglutinin vaccine elicits broadly protective CD4 and CD8 T cell responses against multiple influenza strains and subtypes. *PNAS* **117**, 17757–17763 (2020).
77. Krammer, F., Pica, N., Hai, R., Margine, I. & Palese, P. Chimeric Hemagglutinin Influenza Virus Vaccine Constructs Elicit Broadly Protective Stalk-Specific Antibodies. *Journal of Virology* **87**, 6542–6550 (2013).
78. Wang, T. T. *et al.* Vaccination with a synthetic peptide from the influenza virus hemagglutinin provides protection against distinct viral subtypes. *Proc Natl Acad Sci U S A* **107**, 18979–18984 (2010).
79. Mallajosyula, V. V. A. *et al.* Influenza hemagglutinin stem-fragment immunogen elicits broadly neutralizing antibodies and confers heterologous protection. *Proc Natl Acad*

*Sci U S A* **111**, E2514-2523 (2014).

80. Eggink, D., Goff, P. H. & Palese, P. Guiding the immune response against influenza virus hemagglutinin toward the conserved stalk domain by hyperglycosylation of the globular head domain. *J Virol* **88**, 699–704 (2014).
81. Koszalka, P., Subbarao, K. & Baz, M. Preclinical and clinical developments for combination treatment of influenza. *PLOS Pathogens* **18**, e1010481 (2022).
82. Hussain, M., Galvin, H. D., Haw, T. Y., Nutsford, A. N. & Husain, M. Drug resistance in influenza A virus: the epidemiology and management. *Infect Drug Resist* **10**, 121–134 (2017).
83. Blaising, J., Polyak, S. J. & Pécheur, E.-I. Arbidol as a broad-spectrum antiviral: an update. *Antiviral Res* **107**, 84–94 (2014).
84. Wu, N. C. & Wilson, I. A. Structural Biology of Influenza Hemagglutinin: An Amaranthine Adventure. *Viruses* **12**, 1053 (2020).
85. Uehara, T. *et al.* Treatment-Emergent Influenza Variant Viruses With Reduced Baloxavir Susceptibility: Impact on Clinical and Virologic Outcomes in Uncomplicated Influenza. *The Journal of Infectious Diseases* **221**, 346–355 (2020).
86. Jones, J. C., Pascua, P. N. Q., Harrington, W. N., Webby, R. J. & Govorkova, E. A. Multiple polymerase acidic (PA) I38X substitutions in influenza A(H1N1)pdm09 virus permit polymerase activity and cause reduced baloxavir inhibition. *J Antimicrob Chemother* **76**, 957–960 (2021).
87. Sedeyn, K. & Saelens, X. New antibody-based prevention and treatment options for influenza. *Antiviral Res* **170**, 104562 (2019).
88. Goldhill, D. H. *et al.* The mechanism of resistance to favipiravir in influenza. *Proceedings of the National Academy of Sciences* **115**, 11613–11618 (2018).
89. Lao, J. & Vanet, A. A New Strategy to Reduce Influenza Escape: Detecting



- Therapeutic Targets Constituted of Invariance Groups. *Viruses* **9**, (2017).
90. Perrier, A., Eluard, M., Petitjean, M. & Vanet, A. In Silico Design of New Inhibitors Against Hemagglutinin of Influenza. *J. Phys. Chem. B* **123**, 582–592 (2019).
91. Goder, V. & Spiess, M. Topogenesis of membrane proteins: determinants and dynamics. *FEBS Letters* **504**, 87–93 (2001).
92. Skehel, J. J. & Wiley, D. C. Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Annu Rev Biochem* **69**, 531–569 (2000).
93. Copeland, C. S., Doms, R. W., Bolzau, E. M., Webster, R. G. & Helenius, A. Assembly of influenza hemagglutinin trimers and its role in intracellular transport. *J Cell Biol* **103**, 1179–1191 (1986).
94. Gething, M.-J., McCammon, K. & Sambrook, J. Expression of wild-type and mutant forms of influenza hemagglutinin: The role of folding in intracellular transport. *Cell* **46**, 939–950 (1986).
95. Hurlley, S. M., Bole, D. G., Hoover-Litty, H., Helenius, A. & Copeland, C. S. Interactions of misfolded influenza virus hemagglutinin with binding protein (BiP). *J Cell Biol* **108**, 2117–2126 (1989).
96. Wilson, I. A., Skehel, J. J. & Wiley, D. C. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. *Nature* **289**, 366–373 (1981).
97. Wiley, D. C., Wilson, I. A. & Skehel, J. J. Structural identification of the antibody-binding sites of Hong Kong influenza haemagglutinin and their involvement in antigenic variation. *Nature* **289**, 373–378 (1981).
98. Zheng, Z., Paul, S. S., Mo, X., Yuan, Y.-R. A. & Tan, Y.-J. The Vestigial Esterase Domain of Haemagglutinin of H5N1 Avian Influenza A Virus: Antigenicity and Contribution to Viral Pathogenesis. *Vaccines (Basel)* **6**, (2018).
99. Gamblin, S. J. *et al.* Hemagglutinin Structure and Activities. *Cold Spring Harb*

*Perspect Med* a038638 (2020) doi:10.1101/cshperspect.a038638.

100. Pelkmans, L. & Helenius, A. Insider information: what viruses tell us about endocytosis. *Current Opinion in Cell Biology* **15**, 414–422 (2003).
101. Gamblin, S. J. & Skehel, J. J. Influenza Hemagglutinin and Neuraminidase Membrane Glycoproteins. *J Biol Chem* **285**, 28403–28409 (2010).
102. Bullough, P. A., Hughson, F. M., Skehel, J. J. & Wiley, D. C. Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature* **371**, 37–43 (1994).
103. Sauter, N. K. *et al.* Binding of influenza virus hemagglutinin to analogs of its cell-surface receptor, sialic acid: analysis by proton nuclear magnetic resonance spectroscopy and X-ray crystallography. *Biochemistry* **31**, 9609–9621 (1992).
104. Harrison, S. C. Viral membrane fusion. *Nat Struct Mol Biol* **15**, 690–698 (2008).
105. Benton, D. J., Gamblin, S. J., Rosenthal, P. B. & Skehel, J. J. Structural transitions in influenza haemagglutinin at membrane fusion pH. *Nature* **583**, 150–153 (2020).
106. Gao, J., Gui, M. & Xiang, Y. Structural intermediates in the low pH-induced transition of influenza hemagglutinin. *PLOS Pathogens* **16**, e1009062 (2020).
107. Böttcher, C., Ludwig, K., Herrmann, A., van Heel, M. & Stark, H. Structure of influenza haemagglutinin at neutral and at fusogenic pH by electron cryo-microscopy. *FEBS Letters* **463**, 255–259 (1999).
108. Carr, C. M. & Kim, P. S. A spring-loaded mechanism for the conformational change of influenza hemagglutinin. *Cell* **73**, 823–832 (1993).
109. Doms, R. W. & Helenius, A. Quaternary structure of influenza virus hemagglutinin after acid treatment. *J Virol* **60**, 833–839 (1986).
110. Stegmann, T. Membrane Fusion Mechanisms: The Influenza Hemagglutinin Paradigm and its Implications for Intracellular Fusion. *Traffic* **1**, 598–604 (2000).
111. Hamilton, B. S., Whittaker, G. R. & Daniel, S. Influenza virus-mediated membrane

fusion: determinants of hemagglutinin fusogenic activity and experimental approaches for assessing virus fusion. *Viruses* **4**, 1144–1168 (2012).

112. Kemble, G. W., Danieli, T. & White, J. M. Lipid-anchored influenza hemagglutinin promotes hemifusion, not complete fusion. *Cell* **76**, 383–391 (1994).

113. Chernomordik, L. V., Frolov, V. A., Leikina, E., Bronk, P. & Zimmerberg, J. The Pathway of Membrane Fusion Catalyzed by Influenza Hemagglutinin: Restriction of Lipids, Hemifusion, and Lipidic Fusion Pore Formation. *J Cell Biol* **140**, 1369–1382 (1998).

114. Benton, D. J., Gamblin, S. J., Rosenthal, P. B. & Skehel, J. J. Structural transitions in influenza haemagglutinin at membrane fusion pH. *Nature* **583**, 150–153 (2020).

115. Garcia, N. K., Guttman, M., Ebner, J. L. & Lee, K. K. Dynamic changes during acid-induced activation of influenza hemagglutinin. *Structure* **23**, 665–676 (2015).

116. Lin, X., Noel, J. K., Wang, Q., Ma, J. & Onuchic, J. N. Lowered pH Leads to Fusion Peptide Release and a Highly-dynamic Intermediate of Influenza Hemagglutinin. *J Phys Chem B* **120**, 9654–9660 (2016).

117. Benhaim, M. A., Prasad, V. M., Garcia, N. K., Guttman, M. & Lee, K. K. Structural monitoring of a transient intermediate in the hemagglutinin fusion machinery on influenza virions. *Science Advances* **6**, eaaz8822 (2020).

118. Huang, Q. *et al.* Early steps of the conformational change of influenza virus hemagglutinin to a fusion active state: Stability and energetics of the hemagglutinin. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1614**, 3–13 (2003).

119. Huang, Q., Opitz, R., Knapp, E.-W. & Herrmann, A. Protonation and stability of the globular domain of influenza virus hemagglutinin. *Biophys J* **82**, 1050–1058 (2002).

120. Gambaryan, A. S. *et al.* Specification of Receptor-Binding Phenotypes of Influenza Virus Isolates from Different Hosts Using Synthetic Sialylglycopolymers: Non-Egg-Adapted Human H1 and H3 Influenza A and Influenza B Viruses Share a Common High Binding

- Affinity for 6'-Sialyl(N-acetyl)lactosamine). *Virology* **232**, 345–350 (1997).
121. Rogers, G. N. & D'Souza, B. L. Receptor binding properties of human and animal H1 influenza virus isolates. *Virology* **173**, 317–322 (1989).
122. Connor, R. J., Kawaoka, Y., Webster, R. G. & Paulson, J. C. Receptor Specificity in Human, Avian, and Equine H2 and H3 Influenza Virus Isolates. *Virology* **205**, 17–23 (1994).
123. Matrosovich, M. N. *et al.* Avian Influenza A Viruses Differ from Human Viruses by Recognition of Sialyloligosaccharides and Gangliosides and by a Higher Conservation of the HA Receptor-Binding Site. *Virology* **233**, 224–234 (1997).
124. van Riel, D. *et al.* Human and Avian Influenza Viruses Target Different Cells in the Lower Respiratory Tract of Humans and Other Mammals. *Am J Pathol* **171**, 1215–1223 (2007).
125. Short, K. R. *et al.* One health, multiple challenges: The inter-species transmission of influenza A virus. *One Health* **1**, 1–13 (2015).
126. Böttcher-Friebertshäuser, E., Klenk, H.-D. & Garten, W. Activation of influenza viruses by proteases from host cells and bacteria in the human airway epithelium. *Pathog Dis* **69**, 87–100 (2013).
127. Böttcher-Friebertshäuser, E. *et al.* Cleavage of Influenza Virus Hemagglutinin by Airway Proteases TMPRSS2 and HAT Differs in Subcellular Localization and Susceptibility to Protease Inhibitors. *Journal of Virology* **84**, 5605–5614 (2010).
128. Zhirnov, O. P., Ikizler, M. R. & Wright, P. F. Cleavage of influenza a virus hemagglutinin in human respiratory epithelium is cell associated and sensitive to exogenous antiproteases. *J Virol* **76**, 8682–8689 (2002).
129. Stieneke-Gröber, A. *et al.* Influenza virus hemagglutinin with multibasic cleavage site is activated by furin, a subtilisin-like endoprotease. *The EMBO Journal* **11**, 2407–2414 (1992).

130. Schrauwen, E. J. A. *et al.* The Multibasic Cleavage Site in H5N1 Virus Is Critical for Systemic Spread along the Olfactory and Hematogenous Routes in Ferrets. *Journal of Virology* **86**, 3975–3984 (2012).
131. Klenk, H.-D. & Garten, W. Host cell proteases controlling virus pathogenicity. *Trends in Microbiology* **2**, 39–43 (1994).
132. Byrd-Leotis, L., Galloway, S. E., Agbogu, E. & Steinhauer, D. A. Influenza hemagglutinin (HA) stem region mutations that stabilize or destabilize the structure of multiple HA subtypes. *J Virol* **89**, 4504–4516 (2015).
133. Galloway, S. E., Reed, M. L., Russell, C. J. & Steinhauer, D. A. Influenza HA subtypes demonstrate divergent phenotypes for cleavage activation and pH of fusion: implications for host range and adaptation. *PLoS Pathog* **9**, e1003151 (2013).
134. Russier, M. *et al.* Molecular requirements for a pandemic influenza virus: An acid-stable hemagglutinin protein. *Proc Natl Acad Sci U S A* **113**, 1636–1641 (2016).
135. Mellman, I., Fuchs, R. & Helenius, A. Acidification of the endocytic and exocytic pathways. *Annu. Rev. Biochem.* **55**, 663–700 (1986).
136. Daidoji, T. *et al.* Avian Influenza Virus Infection of Immortalized Human Respiratory Epithelial Cells Depends upon a Delicate Balance between Hemagglutinin Acid Stability and Endosomal pH. *J Biol Chem* **290**, 10627–10642 (2015).
137. Murakami, S. *et al.* Enhanced Growth of Influenza Vaccine Seed Viruses in Vero Cells Mediated by Broadening the Optimal pH Range for Virus Membrane Fusion. *Journal of Virology* **86**, 1405–1410 (2012).
138. Rybak, S. L. & Murphy, R. F. Primary cell cultures from murine kidney and heart differ in endosomal pH. *Journal of Cellular Physiology* **176**, 216–222 (1998).
139. Hensen, L., Matrosovich, T., Roth, K., Klenk, H.-D. & Matrosovich, M. HA-Dependent Tropism of H5N1 and H7N9 Influenza Viruses to Human Endothelial Cells Is

Determined by Reduced Stability of the HA, Which Allows the Virus To Cope with Inefficient Endosomal Acidification and Constitutively Expressed IFITM3. *Journal of Virology* **94**, (2019).

140. Sui, J. *et al.* Structural and functional bases for broad-spectrum neutralization of avian and human influenza A viruses. *Nat Struct Mol Biol* **16**, 265–273 (2009).

141. Lakadamyali, M., Rust, M. J., Babcock, H. P. & Zhuang, X. Visualizing infection of individual influenza viruses. *PNAS* **100**, 9280–9285 (2003).

142. Yoshimura, A. & Ohnishi, S. Uncoating of influenza virus in endosomes. *J Virol* **51**, 497–504 (1984).

143. Marvin, S. A., Russier, M., Huerta, C. T., Russell, C. J. & Schultz-Cherry, S. Influenza Virus Overcomes Cellular Blocks To Productively Replicate, Impacting Macrophage Function. *J Virol* **91**, (2017).

144. Zaraket, H. *et al.* Increased acid stability of the hemagglutinin protein enhances H5N1 influenza virus growth in the upper respiratory tract but is insufficient for transmission in ferrets. *J Virol* **87**, 9911–9922 (2013).

145. Kellis, J. T., Nyberg, K., Sali, D. & Fersht, A. R. Contribution of hydrophobic interactions to protein stability. *Nature* **333**, 784–786 (1988).

146. Makhatadze, G. I., Loladze, V. V., Ermolenko, D. N., Chen, X. & Thomas, S. T. Contribution of Surface Salt Bridges to Protein Stability: Guidelines for Protein Engineering. *Journal of Molecular Biology* **327**, 1135–1148 (2003).

147. Nick Pace, C., Scholtz, J. M. & Grimsley, G. R. Forces stabilizing proteins. *FEBS Lett* **588**, 2177–2184 (2014).

148. Harrison, J. S. *et al.* Role of Electrostatic Repulsion in Controlling pH-Dependent Conformational Changes of Viral Fusion Proteins. *Structure* **21**, 1085–1096 (2013).

149. Mm, G. & S, S. Inter-residue interactions in protein folding and stability. *Prog*

*Biophys Mol Biol* **86**, 235–277 (2004).

150. Hou, Q., Bourgeas, R., Pucci, F. & Rooman, M. Computational analysis of the amino acid interactions that promote or decrease protein solubility. *Scientific Reports* **8**, 14661 (2018).
151. Freitas, R. F. de & Schapira, M. A systematic analysis of atomic protein–ligand interactions in the PDB. *Med. Chem. Commun.* **8**, 1970–1981 (2017).
152. Barlow, D. J. & Thornton, J. M. Ion-pairs in proteins. *Journal of Molecular Biology* **168**, 867–885 (1983).
153. Pace, C. N., Shirley, B. A., McNutt, M. & Gajiwala, K. Forces contributing to the conformational stability of proteins. *FASEB J* **10**, 75–83 (1996).
154. Horovitz, A., Serrano, L., Avron, B., Bycroft, M. & Fersht, A. R. Strength and cooperativity of contributions of surface salt bridges to protein stability. *Journal of Molecular Biology* **216**, 1031–1044 (1990).
155. Anderson, D. E., Becktel, W. J. & Dahlquist, F. W. pH-Induced denaturation of proteins: a single salt bridge contributes 3-5 kcal/mol to the free energy of folding of T4 lysozyme. *Biochemistry* **29**, 2403–2408 (1990).
156. Kumar, S. & Nussinov, R. Close-range electrostatic interactions in proteins. *Chembiochem* **3**, 604–617 (2002).
157. Kumar, S. & Nussinov, R. Salt bridge stability in monomeric proteins | Edited by J. M. Thornton. *Journal of Molecular Biology* **293**, 1241–1255 (1999).
158. Meuzelaar, H., Vreede, J. & Woutersen, S. Influence of Glu/Arg, Asp/Arg, and Glu/Lys Salt Bridges on  $\alpha$ -Helical Stability and Folding Kinetics. 35.
159. Donald, J. E., Kulp, D. W. & DeGrado, W. F. Salt bridges: geometrically specific, designable interactions. *Proteins* **79**, 898–915 (2011).
160. Baumgart, M. *et al.* Design of buried charged networks in artificial proteins. *Nat*

*Commun* **12**, 1895 (2021).

161. Kohn, W. D., Kay, C. M. & Hodges, R. S. Protein destabilization by electrostatic repulsions in the two-stranded alpha-helical coiled-coil/leucine zipper. *Protein Sci* **4**, 237–250 (1995).
162. Harris, T. K. & Turner, G. J. Structural Basis of Perturbed pKa Values of Catalytic Groups in Enzyme Active Sites. *IUBMB Life* **53**, 85–98 (2002).
163. Musafia, B., Buchner, V. & Arad, D. Complex Salt Bridges in Proteins: Statistical Analysis of Structure and Function. *Journal of Molecular Biology* **254**, 761–770 (1995).
164. Isom, D. G., Castañeda, C. A., Cannon, B. R. & García-Moreno E., B. Large shifts in pKa values of lysine residues buried inside a protein. *Proceedings of the National Academy of Sciences* **108**, 5260–5265 (2011).
165. Fitch, C. A., Platzer, G., Okon, M., Garcia-Moreno E, B. & McIntosh, L. P. Arginine: Its pKa value revisited. *Protein Sci* **24**, 752–761 (2015).
166. Harms, M. J. *et al.* The pKa Values of Acidic and Basic Residues Buried at the Same Internal Location in a Protein Are Governed by Different Factors. *Journal of Molecular Biology* **389**, 34–47 (2009).
167. Di Russo, N. V., Estrin, D. A., Martí, M. A. & Roitberg, A. E. pH-Dependent conformational changes in proteins and their effect on experimental pK(a)s: the case of Nitrophorin 4. *PLoS Comput Biol* **8**, e1002761 (2012).
168. Xu, R. & Wilson, I. A. Structural characterization of an early fusion intermediate of influenza virus hemagglutinin. *J Virol* **85**, 5172–5182 (2011).
169. Weis, W. I., Brünger, A. T., Skehel, J. J. & Wiley, D. C. Refinement of the influenza virus hemagglutinin by simulated annealing. *J Mol Biol* **212**, 737–761 (1990).
170. Chen, J., Skehel, J. J. & Wiley, D. C. N- and C-terminal residues combine in the fusion-pH influenza hemagglutinin HA(2) subunit to form an N cap that terminates the triple-



- stranded coiled coil. *Proc Natl Acad Sci U S A* **96**, 8967–8972 (1999).
171. Angelova, K., de Jonge, H., Granneman, J. C. M., Puett, D. & Bogerd, J. Functional Differences of Invariant and Highly Conserved Residues in the Extracellular Domain of the Glycoprotein Hormone Receptors. *J Biol Chem* **285**, 34813–34827 (2010).
172. Liu, Y. *et al.* Impact of Mutations in Highly Conserved Amino Acids of the HIV-1 Gag-p24 and Env-gp120 Proteins on Viral Replication in Different Genetic Backgrounds. *PLoS One* **9**, e94240 (2014).
173. Dyson, H. J., Wright, P. E. & Scheraga, H. A. The role of hydrophobic interactions in initiation and propagation of protein folding. *Proceedings of the National Academy of Sciences* **103**, 13057–13061 (2006).
174. Hubbard, R. E. & Kamran Haider, M. Hydrogen Bonds in Proteins: Role and Strength. in *eLS* (John Wiley & Sons, Ltd, 2010). doi:10.1002/9780470015902.a0003011.pub2.
175. Sticke, D. F., Presta, L. G., Dill, K. A. & Rose, G. D. Hydrogen bonding in globular proteins. *Journal of Molecular Biology* **226**, 1143–1159 (1992).
176. Lee, J. M. *et al.* Deep mutational scanning of hemagglutinin helps predict evolutionary fates of human H3N2 influenza variants. *Proceedings of the National Academy of Sciences* **115**, E8276–E8285 (2018).
177. Qiao, H. *et al.* Specific Single or Double Proline Substitutions in the “Spring-loaded” Coiled-Coil Region of the Influenza Hemagglutinin Impair or Abolish Membrane Fusion Activity. *Journal of Cell Biology* **141**, 1335–1347 (1998).
178. Antanasijevic, A. *et al.* Mutagenesis Studies of the H5 Influenza Hemagglutinin Stem Loop Region. *J Biol Chem* **289**, 22237–22245 (2014).
179. Lee, J. *et al.* Molecular-level analysis of the serum antibody repertoire in young adults before and after seasonal influenza vaccination. *Nat Med* **22**, 1456–1464 (2016).
180. Yu, F. *et al.* A Potent Germline-like Human Monoclonal Antibody Targets a pH-

- Sensitive Epitope on H7N9 Influenza Hemagglutinin. *Cell Host Microbe* **22**, 471-483.e5 (2017).
181. Adachi, Y. *et al.* Exposure of an occluded hemagglutinin epitope drives selection of a class of cross-protective influenza antibodies. *Nat Commun* **10**, 3883 (2019).
182. Mookerjee, S. A., Goncalves, R. L. S., Gerencser, A. A., Nicholls, D. G. & Brand, M. D. The contributions of respiration and glycolysis to extracellular acid production. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1847**, 171–181 (2015).
183. van Zwieten, R., Wever, R., Hamers, M. N., Weening, R. S. & Roos, D. Extracellular proton release by stimulated neutrophils. *J. Clin. Invest.* **68**, 310–313 (1981).
184. Guan, S. *et al.* Exploration of binding and inhibition mechanism of a small molecule inhibitor of influenza virus H1N1 hemagglutinin by molecular dynamics simulation. *Scientific Reports* **7**, 3786 (2017).
185. Liu, H., Maruyama, H., Masuda, T., Honda, A. & Arai, F. The Influence of Virus Infection on the Extracellular pH of the Host Cell Detected on Cell Membrane. *Front Microbiol* **7**, 1127 (2016).
186. Zhou, Y., Wu, C., Zhao, L. & Huang, N. Exploring the early stages of the pH-induced conformational change of influenza hemagglutinin. *Proteins* **82**, 2412–2428 (2014).
187. Sinha, N., Kumar, S. & Nussinov, R. Interdomain Interactions in Hinge-Bending Transitions. *Structure* **9**, 1165–1181 (2001).
188. Eller, M. W., Siaw, H. M. H. & Dyer, R. B. Stability of HA2 Prefusion Structure and pH-Induced Conformational Changes in the HA2 Domain of H3N2 Hemagglutinin. *Biochemistry* **60**, 2623–2636 (2021).
189. Yip, K. M., Fischer, N., Paknia, E., Chari, A. & Stark, H. Atomic-resolution protein structure determination by cryo-EM. *Nature* **587**, 157–161 (2020).
190. Nakane, T. *et al.* Single-particle cryo-EM at atomic resolution. *Nature* **587**, 152–156

(2020).

191. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
192. Doud, M. B., Lee, J. M. & Bloom, J. D. How single mutations affect viral escape from broad and narrow antibodies to H1 influenza hemagglutinin. *Nat Commun* **9**, 1386 (2018).
193. Andrews, S. F., Graham, B. S., Mascola, J. R. & McDermott, A. B. Is It Possible to Develop a “Universal” Influenza Virus Vaccine? *Cold Spring Harb Perspect Biol* **10**, a029413 (2018).
194. O’Sullivan, O., Suhre, K., Abergel, C., Higgins, D. G. & Notredame, C. 3DCoffee: combining protein sequences and structures within multiple sequence alignments. *J Mol Biol* **340**, 385–395 (2004).
195. Konagurthu, A. S., Whisstock, J. C., Stuckey, P. J. & Lesk, A. M. MUSTANG: A multiple structural alignment algorithm. *Proteins: Structure, Function, and Bioinformatics* **64**, 559–574 (2006).
196. Onofrio, A. *et al.* Distance-dependent hydrophobic–hydrophobic contacts in protein folding simulations. *Phys. Chem. Chem. Phys.* **16**, 18907–18917 (2014).
197. Wit, E. de *et al.* Efficient generation and growth of influenza virus A/PR/8/34 from eight cDNA fragments. *Virus Research* **103**, 155–161 (2004).
198. Matrosovich, M., Matrosovich, T., Garten, W. & Klenk, H.-D. New low-viscosity overlay medium for viral plaque assays. *Virology Journal* **3**, 63 (2006).
199. Rachakonda, P. S. *et al.* The relevance of salt bridges for the stability of the influenza virus hemagglutinin. *FASEB J* **21**, 995–1002 (2007).
200. Godley, L. *et al.* Introduction of intersubunit disulfide bonds in the membrane-distal region of the influenza hemagglutinin abolishes membrane fusion activity. *Cell* **68**, 635–645 (1992).

201. Henry Dunand, C. J. *et al.* Both Neutralizing and Non-Neutralizing Human H7N9 Influenza Vaccine-Induced Monoclonal Antibodies Confer Protection. *Cell Host & Microbe* **19**, 800–813 (2016).
202. Jegaskanda, S., Weinfurter, J. T., Friedrich, T. C. & Kent, S. J. Antibody-dependent cellular cytotoxicity is associated with control of pandemic H1N1 influenza virus infection of macaques. *J Virol* **87**, 5512–5522 (2013).
203. He, W. *et al.* Alveolar macrophages are critical for broadly-reactive antibody-mediated protection against influenza A virus in mice. *Nat Commun* **8**, 846 (2017).
204. Mullarkey, C. E. *et al.* Broadly Neutralizing Hemagglutinin Stalk-Specific Antibodies Induce Potent Phagocytosis of Immune Complexes by Neutrophils in an Fc-Dependent Manner. *mBio* **7**, e01624-16.
205. Krammer, F. Novel universal influenza virus vaccine approaches. *Curr Opin Virol* **17**, 95–103 (2016).
206. Ekiert, D. C. *et al.* Antibody recognition of a highly conserved influenza virus epitope. *Science* **324**, 246–251 (2009).
207. Darricarrère, N. *et al.* Broad neutralization of H1 and H3 viruses by adjuvanted influenza HA stem vaccines in nonhuman primates. *Science Translational Medicine* **13**, (2021).
208. Ekiert, D. C. *et al.* A highly conserved neutralizing epitope on group 2 influenza A viruses. *Science* **333**, 843–850 (2011).
209. Gray, C. & Tamm, L. K. Structural studies on membrane-embedded influenza hemagglutinin and its fragments. *Protein Sci* **6**, 1993–2006 (1997).
210. Gao, J., Gui, M. & Xiang, Y. Structural intermediates in the low pH-induced transition of influenza hemagglutinin. *PLOS Pathogens* **16**, e1009062 (2020).
211. Wu, Y. *et al.* A potent broad-spectrum protective human monoclonal antibody

- crosslinking two haemagglutinin monomers of influenza A virus. *Nat Commun* **6**, 7708 (2015).
212. Corti, D. *et al.* A Neutralizing Antibody Selected from Plasma Cells That Binds to Group 1 and Group 2 Influenza A Hemagglutinins. *Science* **333**, 850–856 (2011).
213. Bangaru, S. *et al.* A Site of Vulnerability on the Influenza Virus Hemagglutinin Head Domain Trimer Interface. *Cell* **177**, 1136–1152.e18 (2019).
214. Turner, H. L. *et al.* Potent anti-influenza H7 human monoclonal antibody induces separation of hemagglutinin receptor-binding head domains. *PLoS Biol* **17**, e3000139 (2019).
215. Watanabe, A. *et al.* Antibodies to a Conserved Influenza Head Interface Epitope Protect by an IgG Subtype-Dependent Mechanism. *Cell* **177**, 1124–1135.e16 (2019).
216. Bajic, G. *et al.* Influenza antigen engineering focuses immune responses to a subdominant but broadly protective viral epitope. *Cell Host Microbe* **25**, 827–835.e6 (2019).
217. Sanchez, E. L. & Lagunoff, M. Viral activation of cellular metabolism. *Virology* **479–480**, 609–618 (2015).
218. Martín-Vicente, M. *et al.* Metabolic changes during respiratory syncytial virus infection of epithelial cells. *PLoS One* **15**, e0230844 (2020).
219. Zhao, Y., Chahar, H. S., Komaravelli, N., Dossumentkova, A. & Casola, A. Human metapneumovirus infection of airway epithelial cells is associated with changes in core metabolic pathways. *Virology* **531**, 183–191 (2019).
220. Klemperer, H. Glucose breakdown in chick embryo cells infected with influenza virus. *Virology* **13**, 68–77 (1961).
221. Ritter, J. B., Wahl, A. S., Freund, S., Genzel, Y. & Reichl, U. Metabolic effects of influenza virus infection in cultured animal cells: Intra- and extracellular metabolite profiling. *BMC Syst Biol* **4**, 61 (2010).
222. Keshavarz, M. *et al.* Metabolic host response and therapeutic approaches to influenza

- infection. *Cellular & Molecular Biology Letters* **25**, 15 (2020).
223. Tannahill, G. M. *et al.* Succinate is an inflammatory signal that induces IL-1 $\beta$  through HIF-1 $\alpha$ . *Nature* **496**, 238–242 (2013).
224. Krawczyk, C. M. *et al.* Toll-like receptor-induced changes in glycolytic metabolism regulate dendritic cell activation. *Blood* **115**, 4742–4749 (2010).
225. Seok, J. H. *et al.* Divalent cation-induced conformational changes of influenza virus hemagglutinin. *Scientific Reports* **10**, 15457 (2020).
226. Chaigne-Delalande, B. & Lenardo, M. J. Divalent cation signaling in immune cells. *Trends Immunol* **35**, 332–344 (2014).
227. Feske, S., Wulff, H. & Skolnik, E. Y. Ion channels in innate and adaptive immunity. *Annu Rev Immunol* **33**, 291–353 (2015).
228. Bonaventura, P., Benedetti, G., Albarède, F. & Miossec, P. Zinc and its role in immunity and inflammation. *Autoimmun Rev* **14**, 277–285 (2015).
229. Maywald, M., Wessels, I. & Rink, L. Zinc Signals and Immunity. *Int J Mol Sci* **18**, 2222 (2017).
230. Kong, B. *et al.* Virucidal nano-perforator of viral membrane trapping viral RNAs in the endosome. *Nat Commun* **10**, 185 (2019).
231. Quiñonero, D. *et al.* Interplay Between Cation- $\pi$ , Anion- $\pi$  and  $\pi$ - $\pi$  Interactions. *ChemPhysChem* **7**, 2487–2491 (2006).
232. Childers, M. C., Towse, C.-L. & Daggett, V. The effect of chirality and steric hindrance on intrinsic backbone conformational propensities: tools for protein design. *Protein Eng Des Sel* **29**, 271–280 (2016).
233. Babu, M. M. NCI: a server to identify non-canonical interactions in protein structures. *Nucleic Acids Res* **31**, 3345–3348 (2003).
234. Gromiha, M. M. & Selvaraj, S. Importance of long-range interactions in protein

1This article is dedicated to our teacher, Professor P.K. Ponnuswamy on the occasion of his 60th birthday.1. *Biophysical Chemistry* **77**, 49–68 (1999).

235. Trost, J. F. *et al.* A conserved histidine in Group-1 influenza subtype hemagglutinin proteins is essential for membrane fusion activity. *Virology* **536**, 78–90 (2019).