The Host Heat Shock Response, Viral Immune Escape and Viral Replication

by

Anna Ponomarenko

Specialist in Chemistry
Lomonosov Moscow State University, 2014

SUBMITTED TO THE DEPARTMENT OF CHEMISTRY IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN CHEMISTRY
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2020

Signature of the Author: _____
Department of Chemistry
August 4, 2020

Certified by: _____
Matthew D. Shoulders
Whitehead Career Development Associate Professor
Thesis supervisor

Accepted by: _____
Adam Willard
Associate Professor
Graduate Officer

This doctoral thesis has been examined by a committee of the Department of Chemistry as follows:

_____

Bradley L. Pentelute
Pfizer-Laubach Career Development Associate Professor
Department of Chemistry, MIT
Thesis committee chair

_____

Matthew D. Shoulders
Whitehead Career Development Associate Professor
Department of Chemistry, MIT
Thesis supervisor

_____

Leonid A. Mirny
Professor of Medical Engineering and Science, and Physics
Institute of Medical Engineering and Science, and Department of Physics, MIT
Thesis committee member

The Host Heat Shock Response, Viral Immune Escape and Viral Replication

by

Anna Ponomarenko

Submitted to the Department of Chemistry on August 4, 2020 in Partial Fulfilment of the Requirements for the Degree of Doctor of Philosophy in Chemistry

ABSTRACT

Viral outbreaks follow human history, as illustrated by past and ongoing pandemics. The continuous nature of viral threats to global health is dictated by viral evolution, which enables the rise of resistance to therapeutic and preventive measures, emergence of new strains, and host switching. Addressing all of these challenges requires a deep understanding of all aspects of viral evolution at the host–pathogen interface. The underlying root of viral evolution is the high mutation rate coupled with relatively short replication times. Rapid accumulation of mutations produces mutant viral proteins with new biological functions, such as escape from the host immune system. However, such mutations are generally deleterious to protein structure or folding and thus confer a high biophysical cost, which most viruses are not equipped to address on their own. In contrast, host cells have an extensive chaperone network, which assists folding and resolves misfolding of endogenous cellular proteins. Chaperone assistance also facilitates evolution of host proteins, and, recent evidence suggests, of viral proteins as well, arising from the extensive involvement of chaperones in the viral lifecycle. In this thesis, I address the implications of host chaperones potentiating viral protein evolution. First, I describe how influenza virus resolves the biophysical cost of adaptive mutations by exploiting host chaperones. High-throughput profiling of influenza nucleoprotein mutational tolerance revealed the dependence of a key 1918 Spanish Flu adaptive variant on the availability of host chaperones. Limited access to host chaperone assistance, especially at fever temperatures, restricts accessibility of this structurally deleterious innate immune escape variant. Next, I address molecular details of chaperone folding assistance required for efficient propagation of this adaptive mutant. This work provides the first experimental evidence of host chaperones defining the accessibility of evolutionary important protein variants. Such dependence on host chaperone assistance highlights the vulnerability of rapidly mutating viral proteins. Finally, I demonstrate that excessive upregulation of the host proteostasis network can also render viruses vulnerable and restrict HIV-1 replication. Overall, interaction with the host chaperones is a critical determinant of viral evolution, and the extent of such interaction has to be carefully regulated throughout the viral life cycle.

Thesis Supervisor: Matthew D. Shoulders
Title: Whitehead Career Development Associate Professor

Chapter Abstracts

Chapter 1. The need, the cost and the tools for evolution of viral proteins

As illustrated by past and ongoing pandemics, viruses continue to impose a global health threat. The root of such a threat is the inherent high mutation rate coupled with small genomes and short replication times. These properties translate into the ability of viruses to efficiently evade the host immune response and antiviral treatment within a short amount of time. However, adaptive mutations, that drive viral evolution towards drug and immune resistance, are often biophysically deleterious, increasing the propensity for viral proteins to misfold and thereby hampering viral replication. Host protein folding factors called chaperones interact extensively with viral proteins, and are thus theoretically well-poised to assist the folding of biophysically defective adaptive variants. An important implication of such assistance is the access of viral proteins to evolution buffering function of the host chaperones. Recent work provided the first experimental evidence of chaperones defining the evolutionary trajectories and mutational tolerance of viral proteins. These fundamental discoveries provide the foundation for investigating the direct role of host chaperones in viral adaptation to restrictive replication environments.

Chapter 2. Host chaperones address the biophysical cost of influenza innate immune escape

The threat of viral pandemics demands a comprehensive understanding of evolution at the host–pathogen interface. We show that adaptive mutations responsible for innate immune system escape by influenza are made possible by the hijacking of host chaperones to resolve serious protein folding defects caused by the mutations. Particularly noteworthy, we observe that the Pro283 nucleoprotein variant, which (1) is conserved across human influenza strains, (2) confers resistance to the MxA restriction factor, and (3) critically contributed to adaptation to humans in the 1918 pandemic influenza strain, is rendered unfit by heat shock factor 1 inhibition-mediated host chaperone depletion at febrile temperatures. Thus, influenza subverts host chaperones to uncouple the biophysically deleterious consequences of viral protein variants from the benefits of immune escape. In summary, host proteostasis plays a central role in shaping influenza adaptation, with implications for the evolution of other viruses, for viral host-switching, and for antiviral drug development.

Chapter 3. HSF1 activation can restrict HIV-1 replication

Host protein folding stress responses can play important roles in RNA virus replication and evolution. Prior work suggested a complicated interplay between the cytosolic proteostasis stress response, controlled by the transcriptional master regulator heat shock factor 1 (HSF1), and human immunodeficiency virus-1 (HIV-1). We sought to uncouple HSF1 transcription factor activity from cytotoxic proteostasis stress and thereby better elucidate the proposed role(s) of HSF1 in the HIV-1 lifecycle. To achieve this objective, we used chemical genetic, stress-independent control of HSF1 activity to establish whether and how HSF1 influences HIV-1 replication. Stress-independent HSF1 induction decreased both the total quantity and infectivity of HIV-1 virions. Moreover, HIV-1 was unable to escape HSF1-mediated restriction over the course of several serial passages. These results clarify the interplay between the host's heat shock response and HIV-1 infection, and promote continued consideration of chaperones as potential antiviral therapeutic targets.

Chapter 4. Conclusions and future directions

I briefly review and conclude the results described in the thesis and outline my perspective on future directions in elucidating the molecular mechanisms of chaperone-mediated viral protein evolution and understanding the role of these mechanisms in viral pathogenesis and emergence of resistant variants.

Appendix A. Elucidating the host factors essential for assisting the folding of the influenza nucleoprotein Pro283 variant

In the Appendix, I review the development of experimental system for elucidating the individual HSF1-controlled host factors, which assist folding of influenza nucleoprotein Pro283 variant. We have optimized the method for NP gene delivery and pull-down, and successfully performed the first mass-spectrometry based proteomics experiment to investigate NP host interaction partners in the absence of the host proteostasis perturbation. The successful pilot NP interactome screen provided initial insights into the prospective important host interactors. We are now well-positioned to pursue a systematic investigation of NP host interactome in different host proteostasis environments.

Acknowledgements

The work presented in this thesis would be impossible without the mentorship, guidance, help and support of many people.

First of all, I am incredibly grateful to my advisor Professor Matthew Shoulders for the opportunity to perform a groundbreaking research in his lab at MIT. This opportunity has literally changed my life. Moving to the U.S. for graduate school allowed me to grow not only professionally, but also personally, and Matt was always there for me with lots of support and inspiring enthusiasm. In good times, he celebrated my successes with sincere excitement and taught me to appreciate the small scientific victories much more than I used to. In not so good times, he was always willing to do everything he could to help me solve problems and direct me to the right resources so I could succeed. This incredible support gave me strength to return to the bench after yet another failed experiment and start over with some modifications to both my protocol and my attitude. I am also very thankful that Matt encouraged my growing interest in immune aspects of viral evolution and allowing me to step into this field, thus making the project even more cross-disciplinary. Matt's positive and open-minded attitude has fundamentally shaped the whole lab and created a warm and highly-collaborative work environment. I could not wish for a better place, community and advisor for my graduate journey.

Next, I would like to thank my other mentors who helped me navigate through the graduate school. Thanks to Professor Leonid Mirny for supporting and providing a valuable feedback on my work from a different angle, initiating interesting discussions and genuine interest in my research. Thanks to Professor Bradley Pentelute for sharing the excitement about my work, helping me to perceive a bigger picture while not losing the sight of important details. I would also like to thank Professor John Dolhun for being my very first mentor at MIT. John helped me to become a better teacher, promoting my deep engagement into teaching 5.310 to non-chemistry majors simultaneously with learning something new from each of my students. It is thanks to all of my great mentors that I was able to become a mentor myself and continue to grow professionally and personally.

My appreciation for a collaborative approach in research has risen dramatically during my Ph.D. I had the pleasure to pursue a lot of productive collaborations. Professor Jesse Bloom and the work of his lab had a great impact on own work. They were always very enthusiastic about new ideas and projects that we came up with and always gladly shared their reagents and time for valuable discussions. I am very grateful to Professor Yu-Shan Lin and her lab for their valuable contribution to the nucleoprotein story, described in Chapter 2, by performing the molecular dynamics simulations and assisting us with the data analysis and troubleshooting our analysis pipeline codes. The MIT BioMicro center team, and especially Vincent Butty and Stuart Levine not only made our sequencing projects possible, but also provided an enormous help with troubleshooting our experiments and analyzing the deep sequencing data.

The Shoulders lab was a fun, exciting and creative environment to work in and I am proud of being a part of this warm scientific family. First, I would like to thank the viral evolution project team, with whom I worked shoulder to shoulder while trying to understand the role of chaperones in viral protein evolution. I am forever grateful for my fellow graduate student mentor Angela Phillips. She has patiently taught me all biochemistry, cell biology and virology skills, set an example of excellent organizational skills and cultivated my growing interest in influenza evolution. Angela has always been not only an excellent resource to address to with multiple scientific questions, but also all other questions as well, sharing her experience as a

more senior graduate student and helping me navigate the Ph.D. program with her wisdom and energy. Having such a great mentor during the start of my scientific career shaped all of my graduate research journey. I was also pleased to work with Emmanuel Nekongo and collaborate on the HIV research project, described in Chapter 4. Emmanuel was the most friendly, welcoming and kind colleague that I ever had. He was always very thoughtful of others and was always ready to help his teammates. I was excited to welcome and mentor the newer members of the project, Jimin Yoon and Jessica Patrick. Jimin brought a lot of energy, tons of excellent questions and amazing persistence that eventually allowed her to solve the problem of nucleoprotein expression and start addressing a lot of interesting biophysical questions. Jessica brought a strong enthusiasm and scientific curiosity, which will, without any doubt, greatly advance the project in the nearest future. I had a lot to learn from them as well, for example how to remain positive in challenging times and balance my love for the great outdoors with my passion for research.

Despite working full-time on the viral evolution project, I have never felt isolated from the rest of the Shoulders lab community. From the very first day when I joined the lab I was surrounded by the very intelligent and caring people, who became my scientific family in the U.S. They were always there for me, from helping me to locate the reagents and teaching me how to use new equipment, to explaining me the U.S. culture and listening to me bragging about the Moscow Subway. When starting as a first-year in lab, I had a lot of senior students to look up to and follow their lead. Chris Moore was the person who seemed to have an answer for any question, knowing how to do literally anything. Madeline always very kind, patient and willing to teach new tricks and techniques. Andrew and Azade have always had hilarious stories, that made any day feel better. Duc was a great sympathetic ear to discuss the struggles of immigration. Chet always made sure that the lab does not stay quiet and stays informed about the most recent New England Patriots news. Chris Richardson usually had strong opinions on everything, and initiated interesting debates in the lab, which were fascinating to watch and participate in. Louis, the genome artist of the Shoulders lab, was always a great help with any molecular biology questions, as well as gardening concerns. The impressive jungle on his desk inspired me to start in-lab gardening as well, including an experiment of growing a morning glory plant in lab. Also I have to give a special thanks to Louis for helping me deal with the BL2+ freezer emergency and urgently evacuate all of our HIV and flu stocks and stuff them into other lab's freezers across MIT.

The more junior generation of the lab helped me to become a mentor myself and I tried my best to be as friendly and helpful to them, as the more senior generation was to me. I have a lot of respect for Chichi, for dealing with a lot of organizational lab work and keeping it running smoothly. She was also a great neighbor in lab, who I used to have random short conversations with when we were both working on different sides of the bench, and we could complain about our experiments not working like good comrades. Sam was usually the initiator of discussions in the lunch room, which derailed pretty fast and I was fascinated to watch how much the topic could transform by the end of lunch. Rebecca was always kind to share her HSF1 expertise whenever I had any spontaneous questions and to calm my panic about parking on campus. Kathryn, Jimin and Mani brought a lot of energy, contagious enthusiasm and bright ideas to lab.

I have to thank my best friends for being with me, both physically and remotely, and supporting me throughout my graduate journey. I am deeply grateful for meeting and becoming friends with Kenny even before both of us officially joined the Shoulders lab. We navigated the graduate school together, overcoming all of the challenges and supporting each other, while also scouting campus in search of free food from time to time. Kenny was always open to discuss absolutely anything and share his personal experiences. I am very grateful for having such a great friend

throughout the graduate school. Next, I am thankful for my friends in Russia, Irina Strelkova and Svetlana Tochilina, for preserving our friendship even while being a thousand miles apart. I cherish all of those rare moments when we got to get together in Moscow, update each other on the latest personal news, play some board games and enjoy our time as if nothing has changed much since we last saw each other a year prior to each of our meetings. Even though some aspects of my new life were not always easy to understand for them, I always felt their support and care.

Finally, none of my successes and accomplishments would be possible without my family. I owe everything to my mother, Nataliya Ponomarenko, who worked incredibly hard and sacrificed a lot to help me get where I am now. From pre-school English classes to choosing my scientific career direction, she was always ready to do everything she could to help me either herself, or to find someone who could provide the necessary expertise. It is thanks to my mom I have the chance to live my life the way I want to, feel happy coming to lab every morning and seeing an exciting future straight ahead. I would also like to thank my aunt and uncle, Irina and Sergey, and my cousin Nadiya, for their support and warm welcome whenever I get a chance to go back home, and all of the nice things that they are always willing to do for me.

And last, but not least, I would not make it through the graduate school without my husband Eric Basile and his family. His daily support, speeches of encouragement, both good and slightly weird jokes and gourmet dinners gave me strength to go through yet another busy day in lab without giving up when nothing seemed to work. His enthusiasm, open-mindedness and thirst for knowledge inspires me every day to be not only a better researcher, but also a better self. I am also very grateful to the Basile family for accepting me, supporting all my steps, listening patiently whenever I had something to rant about and offering a thoughtful advice when I really needed it.

Thanks to all of you, who made this scientific journey possible, stayed with me and helped me to go through it to the end!

Although we can continue to marvel at the adaptability of RNA viruses… rather than thinking about what RNA viruses can do in their evolution, we should concentrate on their limitations. RNA viruses might be more at the mercy of their mutation rates than we think.

Edward C. Holmes. *Trends Microbiol.* **2003**.

Table of Contents

List of Figures

Chapter 3: HSF1 activation can restrict HIV replication

Appendix A: Elucidating the host factors essential for assisting the folding of the influenza nucleoprotein Pro283 variant

List of Tables

List of Abbreviations

° C	Degrees Celsius
ΔG_K	Kinetic free energy
ΔG_T	Thermodynamic free energy
µg	microgram
µL	microliter
µm	micrometer
17-AAG	17-N-allylamino-17-demethoxygeldanamycin
A280	Absorbance at 280 nm
A549	Alveolar basal epithelial cells
AHSA1	Activator Of Hsp90 ATPase Activity 1
AIDS	Acquired immunodeficiency syndrome
Ala	Alanine
ANOVA	Analysis of variance
ARL4C	ADP-ribosylation factor-like 4C
ATF6	Activating transcription factor 6
BAG3	BLC2 associated anathogene 3
BiP	Binding immunoglobulin protein
BSA	Bovine serum albumin
CACYBP	Calcyclin Binding Protein
CD	Circular dichroism
CDK3	Cyclin-dependent kinase 3
cDNA	Complementary deoxynucleic acid
CEM	T lymphoblastoid cell line
CFP	Cyan fluorescent protein
CHAC1	ChaC glutathione specific gamma-glutamylcyclotransferase 1
ChIP	Chromatin immunoprecipitation
CHOP	CCAAT-enhancer-binding protein homologous protein
CHORDC1	Cysteine And Histidine Rich Domain Containing 1
cHSF1	Constitutive heat shock factor 1
CLU	Clusterin
cm	Centimeter
CMV	Cytomegalovirus
COVID-19	Coronavirus disease 2019

COX3	Cytochrome C oxidase subunit 3
CRYAB	Crystallin alpha B
CSPD	[3-(1-chloro-3'-methoxyspiro[adamantane-4,4'-dioxetane]-3'-yl)phenyl] dihydrogen phosphate
CTL	Cytotoxic T-lymphocyte
CYP1B1	Cytochrome P450 family 1 subfamily B member 1
CypA	Cyclophilin A
DDIT4	DNA damage inducible transcript 4
DHFR	Dihydrofolate reductase
diffsel	Differential selection
DMEM	Dulbecco's modified eagle medium
DMS	Deep mutational scanning
DMSO	Dimethyl sulfoxide
DNA	Deoxynucleic acid
DNAJA1	DnaJ heat shock protein family (Hsp40) member A1
DNAJB1	DnaJ heat shock protein family (Hsp40) member B1
DNAJB11	DnaJ heat shock protein family (Hsp40) member B11
DNAJB2	DnaJ heat shock protein family (Hsp40) member B2
DNAJC10	DnaJ heat shock protein family (Hsp40) member C2
dn-cHSF1	Dominant negative constitutively active heat shock factor 1
dox	Doxycycline
dsDNA	Double-stranded deoxynucleic acid
DSP	Dithiobis-(succinimidyl propionate)
E. coli	Escherichia coli
EDTA	Ethylenediaminetetraacetic acid
EIF5	Eukaryotic translation initiation factor-5
eIF5A	Eukaryotic translation initiation factor 5A-1
ELISA	Enzyme-linked immunosorbent assay
ENCODE	Encyclopedia of DNA Elements
ER	Endoplasmic reticulum
ERDJ4	Endoplasmic Reticulum DNA J Domain-Containing Protein 4
FBS	Fetal bovine serum
FDR	False discovery rate
FGF21	Fibroblast growth factor 21
FKBP4	FK506 binding protein 4
FLAG	Peptide sequence DYKDDDK

FPLC	Fast protein liquid chromatography
fwd	Forward
G	Glycine
Gag	Group specific antigen protein
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GJA1	Gap junction protein alpha 1
Gly	Glycine
GO	Gene ontology
GRP94	Glucose-regulated protein 94 kDa
GSEA	Gene set enrichment analysis
GSG	Glycine-serine-glycine
h	Hour(s)
HA	Influenza hemagglutinin
HCV	Hepatitis C virus
HEK	Human embryonic kidney
HIV	Human immunodeficiency virus type 1
H1N1	Hemagglutinin subtype 1, neuraminidase subtype 1 influenza
H3N2	Hemagglutinin subtype 3, neuraminidase subtype 2 influenza
hpi	Hours post infection
Hsc70	Heat shock cognate 71 kDa protein
HSF1	Heat shock factor 1
HSF1i	Heat shock factor 1-inhibited environment
Hsp40	Heat shock protein 40 kDa
Hsp70	Heat shock protein 70 kDa
Hsp90	Heat shock protein 90 kDa
HSP90AA1	Heat shock protein 90 alpha family class A member 1
HSP90AB1	Heat shock protein 90 alpha family class B member 1
HSPA1A	Heat shock protein family A (Hsp70) member 1A
HSPA1B	Heat shock protein family A (Hsp70) member 1B
HSPA1L	Heat shock protein family A (Hsp70) member 1 like
HSPA4	Heat shock protein family A (Hsp70) member 4
HSPA6	Heat shock protein family A (Hsp70) member 6
HSPA8	Heat Shock Protein Family A (Hsp70) Member 8
HSPB1	Heat Shock Protein Family B (Small) Member 1
HSPB8	Heat Shock Protein Family B (Small) Member 8
HSPD1	Heat Shock Protein Family D (Hsp60) Member 1

HSPH1	Heat shock protein family H (Hsp110) member 1
HSR	Heat shock response
HTLV-1	Human T-cell leukemia virus type 1
HYOU1	Hypoxia up-regulated 1
IDO1	Indoleamine-pyrrole 2,3-dioxygenase
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IRE1	Inositol-requiring enzyme 1
IRES	Internal ribosomal entry site
ISG	Interferon stimulated gene
K	Lysine
kb	Kilobase(s)
KRT27	Keratin 27
LB	Lysogeny broth
LCN10	Lipocalin 10
Leu	Leucine
MDCK	Madin Darby canine kidney cells
MDCK-SIAT1	Madin Darby canine kidney cells expressing α-2,6-sialic acid receptors
MHC	Major histocompatibility complex
min	Minute(s)
mL	Milliliter(s)
mM	Millimolar
MMP13	Matrix metallopeptidase 13
MOI	Multiplicity of infection
MOPS	3-(N-morpholino)propanesulfonic acid
mRNA	Messenger ribonucleic acid
MS	Mass spectrometry
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
mut	Mutant
MW	Molecular weight
MWCO	Molecular weight cutoff
MxA	Myxovirus resistance protein 1
NHBE	Normal human bronchial epithelial cells
Ni-NTA	Nickel2+ ion coupled nitrilotriacetic acid
nm	Nanometer(s)
nM	Nanomolar
NMR	Nuclear magnetic resonance

NP	Nucleoprotein
NQO1	NAD(P)H quinone dehydrogenase 1
ns	Not significant
NS1	Non-structural protein 1
NUDC	Nuclear migration protein
PARP13	Poly(ADP-ribose) polymerase-13
P	Proline
PBS	Phosphate buffered saline
PCDH7	Protocadherin 7
PCR	Polymerase chain reaction
PDB	Protein database
PEI	Polyethyleneimine
pH	Potential of hydrogen
Pin1	Peptidylprolyl cis-trans isomerase NIMA-interacting 1
PLOD1	Procollagen-Lysine, 2-Oxoglutarate 5-Dioxygenase 1
PME	Particle mesh Ewald
POU2AF1	POU Class 2 Homeobox Associating Factor 1
PPIA	Peptidylprolyl isomerase A (Cyclophilin A)
PPIase	Peptidylprolyl isomerase
PPIB	Peptidylprolyl isomerase B
PPIC	Peptidylprolyl isomerase C
PPIE	Peptidylprolyl isomerase E
PPIG	Peptidylprolyl isomerase G
PPIH	Peptidylprolyl isomerase H
PPIL1	Peptidylprolyl isomerase like 1
Pro	Proline
PRR	Pattern recognition receptor
ps	Picosecond(s)
PTSG2	Prostaglandin-endoperoxide synthase
p-value	Probability value
qPCR	Quantitative polymerase chain reaction
r	Pearson correlation coefficient
rev	Reverse
RIG-I	Retinoic acid inducible gene I
RIPA	Radioimmunoprecipitation assay buffer
RFP	Red fluorescent protein

RNA	Ribonucleic acid
RNA-seq	Ribonucleic acid sequencing
RPLP2	Ribosomal protein lateral stalk subunit P2
RPMI	Roswell park memorial institute medium
RSEM	RNA-seq by expectation maximization
S	Serine
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
sec	Second(s)
SEC	Size exclusion chromatography
SEC24D	SEC24 homolog D, COPII coat complex component
Ser	Serine
SERPINH1	Serpin family H member 1
SPINK5	Serine peptidase inhibitor, kazal type 5
ST13	Suppression of tumorigenicity 13 protein
STAR	Spliced transcripts alignment to a reference
STIP1	Stress induced phosphoprotein 1
STR	Short tandem repeat
syn	Synonymous
TCID50	Tissue culture infectious dose 50
TCP1	T-complex protein 1
TF	Trigger factor
Tm	Apparent melting temperature
TMP	Trimethoprim
TNFRSF10D	TNF receptor superfamily member 10 D
TNFSF14	TNF superfamily member 14
Tris	Tris(hydroxymethyl)aminomethane
UPR	Unfolded protein response
Vif	Viral infectivity factor
Vpu	Viral protein U
WSN	Influenza virus A/WSN/1933
XBP1s	Spliced X-box like protein 1
X-Gal	5-bromo-4-chloro-3-indolyl-p-D-galactopyranoside
Y	Tyrosine
YFP	Yellow fluorescent protein
ZAP	Zinc Finger Antiviral Protein
ZC3HAV1	Zinc finger CCCH-type, antiviral 1

Chapter 1: The need, the cost and the tools for evolution of viral proteins

1.1   Overview

Viral infections constitute a continuous threat to global health despite active advances in healthcare and disease control and prevention measures[1]. Minimalistic structure, fast replication and effective transmission routes equip viruses for causing global outbreaks such as the 1918 Spanish flu pandemic, acquired immunodeficiency syndrome (AIDS) epidemic and the ongoing coronavirus (COVID-19) pandemic, which has already infected more than 18.4 million people worldwide and claimed more than 0.7 million of lives[2]. The persistence of existing pathogens and emergence of novel viruses dictate the urgent need for a comprehensive view of the principles and determinants of viral evolution.

Mutations fuel the evolution of viruses, allowing them to quickly adapt to environmental changes with minimal sacrifice of structural integrity and function of viral proteins. Many RNA viruses, such as influenza and human immunodeficiency virus type 1 (HIV-1) are

Figure 1.1. Mutation rates of different biological entities.
Adapted from Gago, S. *Science* **2009**[3].

particularly well-positioned for rapid adaptation, enabled by their high mutation rate relative to other organisms[3] (**Figure 1.1**). Rapid acquisition of mutations drives emergence and fixation of escape variants, resistant to treatment. Influenza is a prototypical example. The majority of currently circulating influenza strains are already resistant to available antiviral drugs[4]. The efficacy of the influenza vaccine is also limited, thus imposing the need for annual vaccine adjustment, repeated global population vaccination and development of alternative vaccine formulations[5]. Overcoming current treatment and prevention shortcomings requires a thorough understanding of the complex interactions occurring at the host-pathogen interface.

Viral replication involves extensive interactions between viral and host components[6-7]. While the virus is striving to efficiently generate and spread its progeny using the host's resources[8], the host attempts to defend itself with the available immune system arsenal. This standoff implies multiple intermolecular interactions, both enabling and restricting viral replication. These interactions set the direction of viral protein evolution: variants that are more susceptible to host restriction are eliminated while mutants that decrease restrictive molecular interactions with the host are enriched. However, in addition to functional benefits or disadvantages, mutations also impose a biophysical cost to protein stability and folding.

Protein biophysics fundamentally restricts protein evolution[9]. Mutations typically destabilize protein structure[10]. If mutant proteins fail to adapt their functional conformation, such mutations will be purged from the population because of compromised vital functions of those proteins. Thus, even if a mutation theoretically could confer a potential phenotypic benefit, such as superior immune escape, it will remain inaccessible if it drastically perturbs protein folding or stability (**Figure 1.2**)[11]. Viruses are generally not equipped with molecular tools to address protein stability and folding issues on their own. In some cases, deleterious effects of mutations
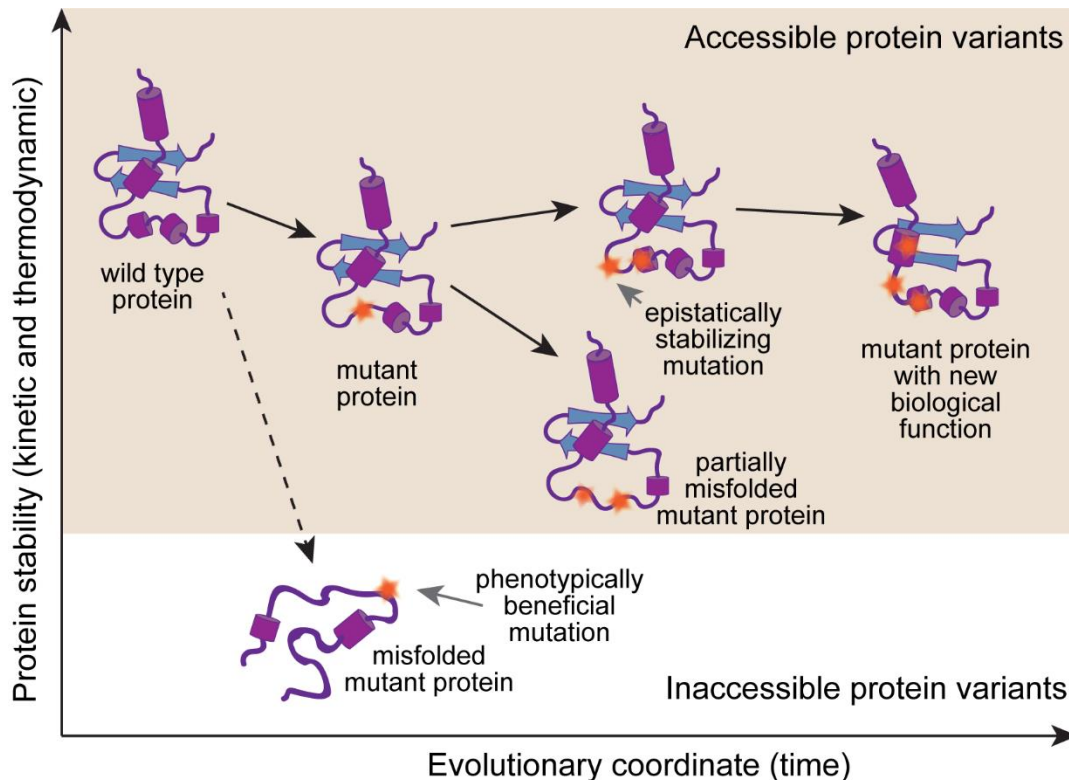
Figure 1.2. Protein evolution is fundamentally constrained by stability.
During evolution, a protein acquires mutations, which drive its amino acid sequence further away from the wild type towards developing a new biological function (for example, immune escape for viral proteins). Mutations generally negatively impact protein folding and stability. In rare cases, destabilizing effects of a certain mutation can be partially mitigated through epistasis. Mutations that drastically perturb protein structure and cause misfolding cannot persist in the population. Adapted from DePristo *et al. Nat. Rev. Genet.* **2005**[11].

can be partially alleviated by the acquisition of additional epistatically stabilizing mutations (**Figure 1.2**)[12-13]. In contrast, host cells have a complex proteostasis network positioned to solve these types of mutation-induced protein folding and stability problems by assisting the folding of endogenous cellular proteins through the functions of numerous chaperones[14]. Viruses actively interact with diverse host cell proteostasis components[6, 15], and inhibition of individual host chaperones can restrict viral proliferation[16-17].

Dependence on host proteostasis assistance is a pharmacologically under-exploited vulnerable step in the viral life cycle. Importantly, host proteostasis is not only one more host resource that viruses exploit during replication[8], but it is also a potential mechanism for guiding viral evolution. Host chaperones facilitate evolution of their endogenous cellular protein clients[18] and define evolutionary trajectories of viral proteins[19]. The implications of chaperone-guided viral evolution can span emergence of novel viral strains, host switching and adaptation, as well as the development of viral resistance mechanisms refractory immune surveillance and drug treatment. Below, I review known forces that drive and restrict evolution of viral proteins, followed by discussing the roles that the host proteostasis can potentially play in the evolution of viral proteins, as well as the epidemiological relevance of these phenomena.

1.2 Viruses evolve under constant selection pressure by mutating rapidly

Viruses are obligatory parasites proliferating at the expense of their host[20]. Coexistence of a virus and a host implies dynamic adaptation of both parties, driving the arms race of the host to reduce the burden of viral replication and evolution of the virus to bypass the host defense strategies[21-22]. Such coevolution directs viral genetic variation[23] towards changes that maximize viral fitness in a given environment[24]. Understanding the forces that facilitate and constrain viral evolutionary paths is necessary for predicting evolutionary dynamics with critical implications in medicine and global health.

A high inherent rate of genetic variation enables rapid viral evolution[3, 25]. While there are two main sources of genetic variation: mutation and recombination[23], this work focuses exclusively on mutations. Viruses with small genomes, especially RNA viruses, possess the highest mutation rates relative to other viruses and cellular organisms (**Figure 1.1**)[26]. RNA viruses replicate with extremely low fidelity originating from the function of error-prone polymerases[27] and reverse transcriptases[28], both of which lack proofreading activity. For example, mutation rates of influenza and human immunodeficiency virus type 1 (HIV-1) are 4.5×10^{-5} and 4.9×10^{-5} substitutions per nucleotide per replication cycle, respectively[29]. Accounting for the genome size, these values translate into 0.6 and 0.4 mutations per genome per replication cycle. Depending on the measurement method used, these mutation rates can be even higher than values derived from genome sequencing estimates, mentioned above. For instance, a recent more accurate method of measuring the influenza mutation rate revealed an average of 2 – 3 mutations per genome per generation[30]. Similarly, a mutation rate of 4 mutations per genome per replication cycle for HIV-1 was recently estimated *in vivo*, attributing the 10-fold increase over cell-based assay estimates to the large contribution of mutagenic host antiviral restriction factors[31]. In contrast to the above mentioned examples, the mutation rates of other organisms are multiple orders of magnitudes lower. For example, mutation rates for *Escherichia coli* and *Caenorhabditis elegans* are 4.1×10^{-10} and 2.2×10^{-10} substitutions per nucleotide per generation, respectively[32-33]. Thus, high mutation rates coupled with small genomes and short replication times allow viruses to rapidly explore genotypic space and identify beneficial mutations. This work focuses on mutations that alter amino acid sequences of viral proteins (or nonsynonymous mutations) and are subject to phenotype based selection.

Mutations that alter viral phenotype are beneficial if they provide a survival advantage for the virus. Mutant viral populations are subject to selection pressure from the host defense systems, targeted on recognition and clearance of invading pathogens. Viable mutant viruses

27

that successfully bypass multiple levels of the host antiviral response can continue replication and spread the infection. Thus, viruses evolve under continuous pressure from their host.

The first line of the host defense system that imposes selection on evolving viruses is the innate immune system[34]. In response, viruses developed strategies to evade the innate immune response by either adapting viral proteins to interfere with the host's antiviral pathways[35] or leveraging viral genetic variation to escape from undesired intermolecular interactions with individual restriction factors[36]. An example of the former strategy is interferon response deactivation[37] by production of viral-encoded interferon antagonists[38]. Interferons are cytokines that are produced in response to sensing viral genomes by specialized classes of host cell receptors (**Figure 1.3A**)[39]. The interferon molecules are then secreted to alert the adjacent cells[40], to limit the viral spread by activating the adaptive immune system[41]. Viral proteins can antagonize multiple stages of the host interferon response. For example, Dengue virus reduces type I interferon production in human dendritic cells[42]. Non-structural proteins of influenza[43] and other viruses[44-45] prevent interferon signaling. Human adenovirus protein E1A compromises interferon-induced gene expression, designated to mount the downstream antiviral response[46]. The HIV-1 accessory proteins ,Vif and Vpu, counteract antiviral protein products of the interferon response[47]. However, in contrast to generally established antagonism of viral proteins towards conserved host's targets[48], genetic variation remains a continuous source of viral diversity[25], leading to emergence of novel viral protein variants, conferring immune escape.

Intermolecular interactions between individual host restriction factors and viral proteins promote enrichment of mutations that abrogate innate immune replication restriction. A striking example is introduction of point mutations into influenza nucleoprotein that allow viral genome complexes with nucleoprotein and polymerase to evade recognition by the host restriction factor MxA (**Figure 1.3B**)[49]. Acquisition and fixation of such mutations played a key role in the emergence of pandemic human influenza strains in both 1918 and 2009. DNA viruses can also evade host innate immunity by acquiring mutations, despite their lower mutation rate compared to RNA viruses[3]. For instance, a murine cytomegalovirus gains mutations in the *m157* open reading frame during a single infection of immunodeficient mice, which lack the adaptive immune response and only possess innate immunity[50]. These mutations abrogate recognition of the virus by the natural killer cells receptor. Mutations can not only relieve the virus of inhibitory interactions with the host, but also assist in developing novel escape mechanisms by establishing additional host-pathogen interactions. For example, a mutation in Zika virus NS1

28
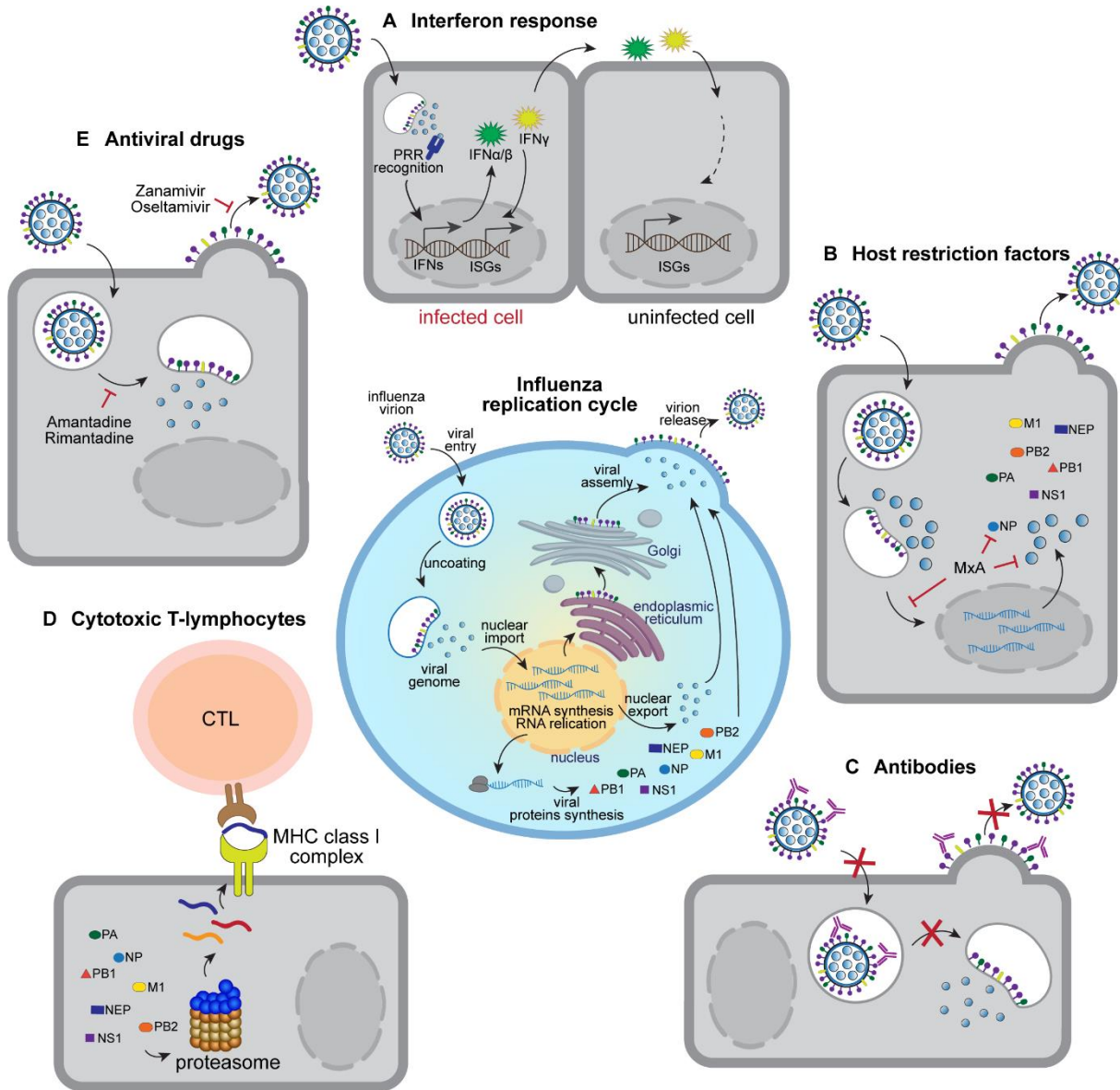
Figure 1.3. Influenza replication cycle (center) and sources of selection pressure.
(**A**) The host cells sense incoming viral genomes by pattern recognition receptors (PRR)[39] and activate interferon response. Interferon molecules are produced, secreted to alert the adjacent cells and trigger activation of interferon stimulated genes (ISGs)[40]. (**B**) Host restriction factors, such as MxA, inhibit viral replication by establishing intermolecular interactions with viral proteins[49]. (**C**) Antibodies are produced by the adaptive immune system and target a specific set of residues of the viral surface proteins. Antibody binding blocks viral surface proteins from performing their functions[53, 54]. (**D**) Proteasomal degradation of viral proteins produces short peptides (antigens), which can be displayed on the cell surface by major histocompatibility complex (MHC) class I. Such antigen presentation attracts specialized cytotoxic T-lymphocytes (CTLs), which recognize and destroy infected cells[61]. (**E**) Antiviral drugs typically target viral surface proteins, bind to distinct pockets and interfere with viral protein function[69].

29

gene was introduced in 2012 that allowed the mutant NS1 protein to bind the cellular kinase required for interferon regulation and to antagonize the interferon-β induction[51]. Thus, mutations functionally complement the antagonism strategy of viruses in evasion from the host innate immunity.

Following the initial non-specific response to viral infection, the innate immune system activates the specialized adaptive immune response, which is tailored to recognize and select for a narrow set of residues of the viral surface proteins. This response is the production of virus-specific antibodies by B lymphocytes[52]. Neutralization is achieved by direct binding of antibodies to their epitopes on viral surface proteins (**Figure 1.3C**)[53-54]. Thus, antibodies recognize and exert selection pressure on those protein residues that they can sufficiently access and bind. Mutation of residues at the antigen-antibody interface is sufficient to abolish neutralization and restore the effective viral replication. Mapping all possible point mutations of different viruses that enable escape from neutralizing antibodies is important for efficient vaccine design and can be achieved using recently developed high throughput screening methods[55-59].

In addition to viral surface proteins, internal proteins are also targeted by the adaptive immune system through cytotoxic T lymphocytes (CTLs)[60]. During virus replication, some of the viral proteins are degraded into peptides, which can bind to the major histocompatibility complex (MHC) molecules and be presented on the cell surface, thus recruiting CTLs to kill infected cells (**Figure 1.3D**)[61]. In contrast to antibodies that recognize generally highly variable surface proteins, CTLs predominantly recognize more conserved internal viral proteins[62-66]. This observation prompted vaccine development targeted at eliciting not only antibody, but also CTL responses[61]. However, point mutations can still enable CTL escape[67-68]. Nonetheless, both antibody-producing B-lymphocytes and CTLs provide pathogen-specific defense and can persist in the human body for years after initial exposure to pathogen, thus creating immune memory. Such memory provides both a much more rapid response to a repeated infection with the same pathogen and a continuous selection pressure on immunogenic epitopes in viral proteins.

Finally, antiviral drugs bind a small group of residues in viral proteins and thus select for point mutations that disrupt binding and restore viral replication (**Figure 1.3E**). Small molecule-based antiviral treatments are currently available for a variety of viruses including influenza[69], HIV-1[70], hepatitis B[71] and C[72], and typically block certain steps in the viral life cycle. However, after initial success of reducing viral loads, these treatments often lose efficacy due to the emergence of resistance[73]. Influenza is an example illustrating the rapid accumulation of

resistance conferring mutations[74], rendering the majority of current circulating strains resistant to available medications. The problem of antiviral drug resistance is exacerbated by the slow rates of novel drug discovery and approval, which can require decades. Altogether, functionally beneficial mutations in viral proteins pose a serious challenge for treatment and constitute a threat to global human health.

However, despite the phenotypic advantage of certain mutant viral protein variants, the vast majority of mutations are deleterious to the virus. First, mutations can not only confer resistance, but also make the virus more vulnerable to host restriction. For example, despite the viral protein antagonist strategy of suppressing the interferon response, point mutations in viral proteins can still render the virus sensitive to interferon restriction and cut short viral replication. Hence, the recent discovery of mutations across the entire influenza genome that both induce interferon sensitivity and increase immunogenicity by provoking a stronger interferon response than the wild type virus[75]. While such mutations serve as invaluable targets for vaccine design, mutant viruses face significant replication challenges *in vivo*. Similarly, mutations in influenza polymerase can disrupt the integrity of the viral genome-polymerase complex and expose the viral RNA panhandle – a viral transcription and replication promoter, created by partially complementary 5′ and 3′ sequences. The exposed panhandle can be then recognized and bound to the host RIG-I RNA helicase, which inhibits viral replication[76]. Thus, not all mutations are beneficial for evolving viruses and, in fact, the majority of mutations are deleterious to viral proteins, conferring a significant fitness cost. In addition to environmental selection pressure, discussed above, fundamental protein biophysics is the critical determinant of protein evolution.

1.3 Functional mutations often have biophysical costs for viral proteins

Mutations that provide phenotypic advantage to viruses often also impair viral fitness. Surface viral proteins are standout examples as they remain under stringent immune selection pressure and need to maintain their functions essential for viral entry into the host cells and subsequent replication. For instance, influenza hemagglutinin (HA) can escape neutralizing antibodies by acquiring new glycosylation sites through point mutations. However, such escape mutants decrease receptor binding avidity[77]. Similarly, escape mutations in HIV-1 envelope glycoprotein allow it to evade broadly neutralizing antibodies[78] and CTLs[79], but also impair viral replication. These fitness costs might be so high that the pathogen might revert back to the wild-type strain[80-81] or adjust its lifecycle according to selection patterns, such as timing of drug therapy[82]. In that regard, the viral mutation rate becomes a critical determinant of viral evolution, which has to be finely tuned to prevent virus replication failure due to excessive accumulation of destabilizing mutations.

Viruses have to balance their adaptation to a continuously changing environment and genetic fidelity, both of which are necessary for survival. Error-prone replication causes viruses to exist as a quasispecies – a dominant wild-type sequence, surrounded by a group (cloud) of closely related mutants[83]. The mutant cloud as a whole, and not individual viral variants, is subject to evolutionary selection. The dynamic heterogeneous population can be potentially advantageous for viral adaptation to environmental changes. However, the rate of accumulating and diversifying this pool of mutants cannot be increased indefinitely, because it will eventually lead to accumulation of lethal errors and loss of biological information once the mutation error threshold is crossed[84]. This dichotomy is especially serious for RNA viruses, possessing the highest mutation rates among viruses[29]. Any additional increase of the mutation rate, for example by using chemical[85-86] or cellular[87] mutagens, can render viral population non-viable by pushing it beyond the error threshold. Thus, replication fidelity is finely tuned, and its perturbation by point mutations in viral replication machinery can alter viral growth[88]. However, even if the mutation acquisition rate is optimal, the landscape of mutations available for a virus to explore is limited by fundamental protein biophysics.

Each mutant protein variant is characterized by its functional performance, or fitness, contributing to the ability of a virus as a whole to adapt and produce infectious progeny in a given environment. The relative viral fitness can be experimentally determined by performing viral growth competitions between genetically distinct viruses[89-90]. Linking fitness to individual
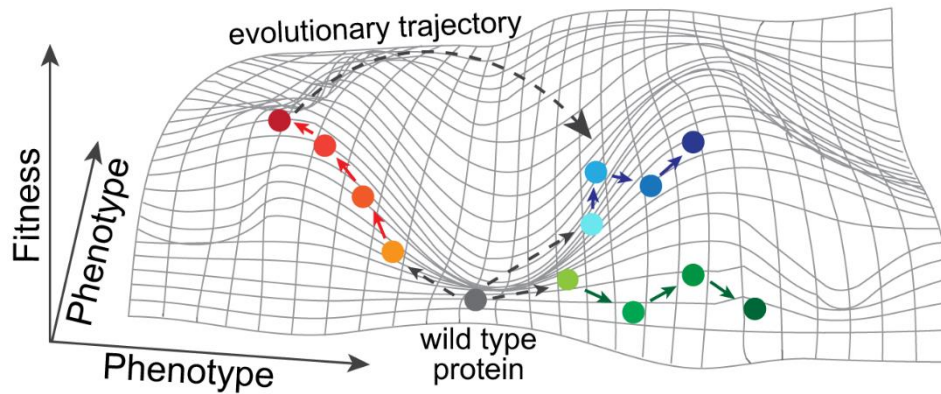
Figure 1.4. Fitness landscape visualization.
The X and Y axes determine viral protein phenotype and Z axis defines viral variant fitness, measuring the performance of mutant viral proteins in a viral life cycle. Each sphere represents a certain amino acid sequence of an evolving protein. Arrows represent evolutionary steps – single changes between the two neighboring sequences. Typically, evolutionary changes are small (short arrows), however in rare cases the change can be drastic (long arrow). Three alternative evolutionary paths originating from the same starting sequence (grey sphere) are illustrated with arrows and colored spheres.

genotypes informs the fitness landscape, comprised of fitness peaks of the most fit variants, valleys of relatively neutral variants and pits of poorly performing mutants[91] (**Figure 1.4**). During viral adaptation to a new environment, mutated viral proteins explore the sequence space and thus move along the fitness landscape. The fitness landscape is useful for predicting viral evolution[92] and thus informing design of novel vaccines[93], and can be inferred experimentally[94] and computationally[93, 95]. However, the virus explores only a small part of the fitness landscape due to the destabilizing impact of mutations, which restrict the accessible protein evolutionary trajectories. The composition of viral quasispecies and individual biophysical properties of viral variants at a given time direct the course of viral evolution.

The biophysical effect of mutations is the critical determinant of protein evolution. Mutations that are represented in viral quasispecies at a given time can be deleterious, neutral and beneficial to the viral fitness. However, mutations are generally destabilizing to protein structure[10]. Thermodynamic stability of proteins is typically low, with the free energy (ΔG) ranging between −3 and −10 kcal/mol (**Figure 1.5**). Mutations can change the protein free energy by as much as 0.5 – 5 kcal/mol on average and thus drastically perturb protein stability[11]. Kinetic stability,
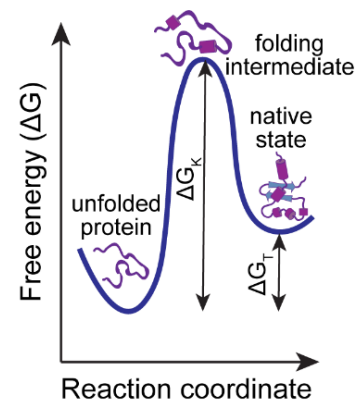
Figure 1.5. Free energy diagram, illustrating thermodynamic (ΔG_T) and kinetic (ΔG_K) protein stability.
Adapted from Colón, W. *Biochemistry* **2017**[96].

defined by protein folding (**Figure 1.5**), also has liabilities. While the marginal stability of proteins allows for their dynamic turnover within living systems, it also imposes a low barrier for stability, structure and function disruption by destabilizing mutations[96]. The extent of protein ability to remain functional in spite of mutation, termed mutational tolerance or mutational robustness, defines which mutations can persist in the viral quasispecies population.

Mutational robustness is an important protein property, which depends on protein structure and can be predicted computationally[97-98] or quantified experimentally[99]. RNA viruses generally have a high mutational robustness, which allows them to rapidly accumulate multiple adaptive mutations while maintaining their biological integrity. However, mutational tolerance of individual viral proteins can vary greatly. The two striking examples of viral proteins with drastically different mutational tolerances are influenza hemagglutinin and measles virus glycoproteins. Influenza and measles are both RNA viruses with similar mutation rates[29, 100]. However, evolution of influenza HA drives the rapid emergence of novel antigenic variants[101], while measles virus surface proteins lack such antigenic variation[102]. An explanation for this discrepancy lies in the structural flexibility of the two surface proteins' structures. Influenza HA , and especially its head domain, is structurally flexible, tolerating numerous mutations and insertions[99, 103]. This property promotes the emergence and accumulation of mutant influenza viruses with altered HA proteins. Such alterations might be not drastic enough to perturb influenza HA biological functions, but sufficient to abolish interactions with the host's antibodies. Accumulation of such mutations fuels antigenic variation and antibody escape[55-56, 104]. In contrast, surface proteins of the measles virus are structurally rigid and do not tolerate mutagenesis[105]. The lack of structural flexibility prevents measles virus from accumulating mutations in surface glycoproteins and thus driving antigenic variation in a way similar to influenza HA. Instead, the structures of measles virus internal proteins are more accommodating to mutations and insertions, making internal proteins major sources of variation between measles strains[106]. Thus, mutational tolerance of the viral surface protein structure defines the long-term effectiveness of acquired immunity. However, fitness effects of individual mutations are not independent of each other, and their genetic context is important to consider.

The effect of a new mutation on a protein can depend on the context of other mutations – a phenomenon known as epistasis[12]. Epistasis is likely to arise during adaptive protein evolution[107]. For example, the acquisition of five epistatically stabilizing mutations in influenza neuraminidase allowed acquisition and fixation of the critical H275Y amino acid substitution, conferring oseltamivir resistance during 2007 – 2009 H1N1 influenza season (**Figure 1.6A**)[108].

Similarly, influenza nucleoprotein evolution marked fixation of mildly stabilizing mutations just before the acquisition of destabilizing mutations (**Figure 1.6B**), necessary for immune escape[109]. Moreover, epistasis is not constrained to direct and short evolutionary paths on the fitness landscape. Indirect trajectories involving gradual acquisition and loss of mutations can act as alternative routes of adaptation[110]. However, epistatic stabilization of protein structure has its own limits. Strongly stabilizing mutations are exceptionally rare. Moreover, such epistasis is incomplete because it is antagonistic[111], causing smaller effects in combination than expected from effects of individual mutations[112]. Finally, the complexity of protein folding can dramatically affect epistasis[13].
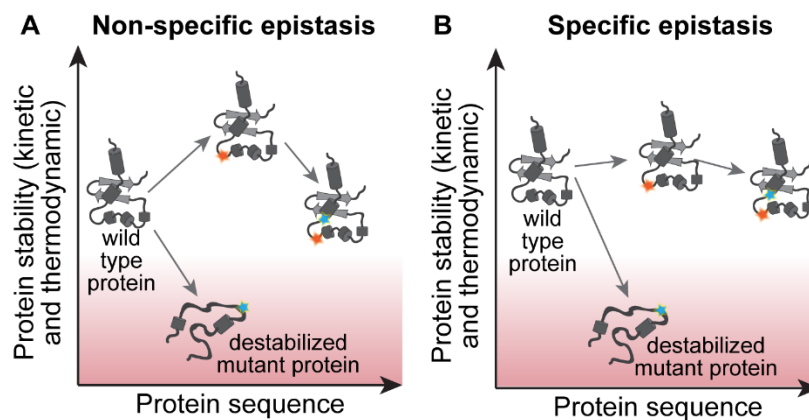
Figure 1.6. Two mechanisms of epistasis during protein evolution.
(**A**) Non-specific epistasis allows fixation of a functionally beneficial mutation (blue star) after acquisition of a strongly stabilizing mutation (orange star). (**B**) Specific epistasis requires initial acquisition of a certain mutation, which by itself does not drastically perturb protein stability but is required for emergence of a functionally beneficial mutation. Specific epistasis enables less mutations than non-specific. In the absence of any epistasis, the phenotypically beneficial mutation drastically destabilizes the protein, causing protein misfolding. Adapted from Starr, T.N. and Thornton, J.W. *Protein Sci*. **2016**[12].

Mutations affect not only the final folded protein structure, but also the folding process itself, which might require external assistance. Mutations can alter the stability and structure of folding intermediates, as well as the folding kinetics[113]. In extreme scenarios, a mutation can drastically change the fold of the protein[114]. In a crowded cellular environment, mutations can also disrupt the essential molecular interactions, as well as cause undesired interactions. Molecular chaperones can address such detrimental effects of mutations by providing folding assistance[113]. Such assistance allows proteins to explore the sequence space more efficiently, hence facilitating protein evolution[18].

1.4 Host proteostasis is positioned to mitigate the protein folding cost of viral evolution

Protein folding is critical to protein function. Viruses encode a limited subset of proteins, each of which typically has more than one function in the viral life cycle. Interruption of any of the vital functions due to mutation-provoked misfolding can be lethal for the virus, even if the mutation could theoretically provide an adaptive benefit. Previous experiments demonstrated that adaptive mutations conferring a beneficial evolutionary change for the virus perturb protein folding stability and have to be accompanied by structure-stabilizing secondary mutations to persist in the population[109, 115]. Thus, the deleterious effect of mutations to protein folding imposes a significant constraint on viral protein evolution. The evolving viruses have to resolve the protein folding issues of mutated viral proteins in order to access the functional advantage of acquired mutations.
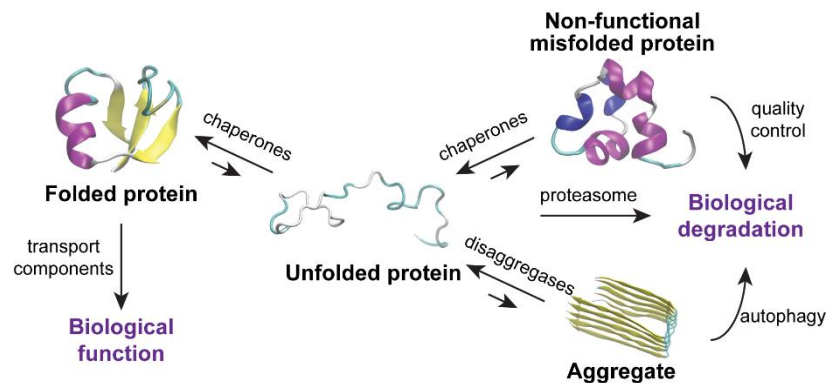
Figure 1.7. The proteostasis network maintains cellular proteome.
The proteostasis network is responsible for protein folding, trafficking, quality control and a timely degradation.

The dynamic nature of a nascent polypeptide chain, marginal stability of folded proteins and risk of protein misfolding imposed by internal and environmental factors justify the need for specialized assistance in protein folding and structural maintenance. Host cells have developed a dynamic system termed the proteostasis network for assisting and regulating folding for thousands of its client proteins[14] (**Figure 1.7**). The proteostasis network is comprised of multiple chaperones, quality control molecular machinery, trafficking and degradation components and regulatory signaling pathways, enabling protein folding in a complex cellular milieu. Different cellular compartments are equipped with the proteostasis network branches and components that provide specialized assistance for their client proteomes[116-118].

The proteostasis network is responsible for not only routine maintenance of the cellular proteome, but also for a rapid cellular response to environmental changes. A variety of internal

and external stimuli, such as heat, oxidative stress, invading pathogens and errors in protein synthesis steps can cause proteome stress. Stress-driven accumulation of misfolded or aggregated proteins triggers the specialized cellular stress responses, targeted to clear the burden of compromised protein homeostasis. Such stress responses include the heat shock response (HSR) in the cytosol[116], and the unfolded protein responses (UPR) in the endoplasmic reticulum (ER)[117] and mitochondria[119] (**Figure 1.8**). These responses promote expression of molecular chaperones, which address the mutational and environmental damages to protein structure and restore optimal protein homeostasis.
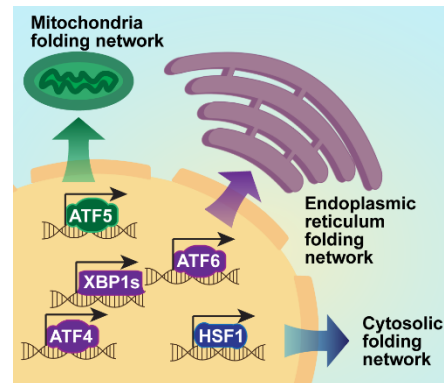
Figure 1.8. Protein misfolding stress responses. Accumulation of misfolded proteins activates proteostasis network branches in cytosol, endoplasmic reticulum and mitochondria.

The ability to address and resolve deleterious impacts of genotypic changes on protein structure enables chaperones to facilitate adaptive protein evolution. The pioneering work by S. Lindquist on the Hsp90 molecular chaperone demonstrated that Hsp90 buffers evolution of its client proteins[120]. Deleterious mutations in Hsp90 clients condition those proteins for prolonged Hsp90 binding, thus masking the phenotypic effect of such mutations. Inhibition of Hsp90 uncovers the pool of mutant proteins, or 'cryptic variation', which can then lead the evolution towards novel phenotypical traits[121]. Another example of chaperone buffering of deleterious mutations was observed for the GroEL/GroES prokaryotic system, which assists in folding ~10% of all soluble *E. coli* proteins. Overexpression of these chaperonins causes the increased accumulation of mutations[122]. The groundbreaking work on Hsp90 and GroEL/GroES triggered extensive interest in exploring the chaperone potential in mediating protein evolution[19, 123-126].

The essential functions of the proteostasis network, discussed above for the host cells, mark its potential significance in the viral lifecycle as well. Rapid accumulation of mutations by evolving viruses possesses a high risk of producing misfolded proteins with compromised functions. The limited number of virus-encoded proteins typically implies their versatility. Interruption of even one of viral protein functions can be lethal[127]. Although a few viruses do encode their own chaperone-like proteins[128-130], majority of viruses lack protein folding machinery. Host chaperones are thus well-positioned to address the viral protein misfolding problems in a fashion similar to their endogenous clients.

Viruses upregulate and extensively interact with host chaperones during replication. Multiple previous studies highlighted the intimate interplay between viral proteins and host proteostasis network components (**Figure 1.9**)[6-7, 131]. Known folding mechanisms of some viral proteins directly involve host chaperones assistance in reaching functional protein conformation[132-135]. For example, influenza hemagglutinin folding intermediates bind calnexin and

Figure 1.9. Viral proteins actively engage host chaperones.
Each line represents a specific interaction between influenza proteins (colored rectangles) and host chaperones (grey ovals)[6-7, 130].

calreticulin chaperones, which prevent HA aggregation and incorrect folding[132]. In addition, host chaperones also enable important post-translational modifications, such as addition of glycans to viral surface proteins and thus masking them from the host immune system surveillance[136-137]. Finally, multiple viruses also provoke host stress responses to reprogram the host cells to produce large amounts of viral proteins within a short time period[138-141].

The role of chaperones in viral replication spans beyond just protein folding assistance and expands into multiple stages of a virus life cycle[15]. Additional vital functions of cellular chaperones include assistance in cellular entry[142], viral capsid uncoating[143], nuclear entry[144], genome replication[145], gene expression[146], capsid assembly[147] and maturation of viral particles[148]. Restricted availability and activity of the host chaperones can endanger viral survival[16]. Hsp70 and Hsp90 are the most extensively studied viral protein interactors, and their pharmacological inhibition suppresses viral replication[17, 149]. However, in contrast to extensively demonstrated roles of chaperones in sustaining viral infection, their possible role in promoting evolution of viral proteins gained significant attention only recently.

Exploitation of the host chaperones provides viruses with access to chaperone-mediated buffering of genetic variation. The ability of chaperones to buffer destabilized mutant proteins enables them to accelerate evolution of their client proteins[150-152]. Such property positions chaperones as invaluable host cell assets for promoting more extensive exploration of the amino acid sequence space by evolving viral proteins. Recent development of tools for perturbing the host proteostasis network (**Figure 1.10A**)[153-154] and high-throughput profiling of

mutational landscape available to viral proteins (**Figure 1.10B**)[56-57, 109, 155] allowed a thorough investigation of proteostasis influence on viral evolution. Our laboratory recently reported the first experimental evidence of proteostasis network shaping influenza evolutionary trajectories[19] and defining viral protein mutational tolerance[156]. Next, the molecular role of Hsp90 chaperone in balancing the protein stability and aggregation propensity of evolving poliovirus protein was demonstrated[125]. Taken together, these reports provide initial insights into the roles of host proteostasis as one of determinants of viral protein evolution. However, the epidemiological and therapeutic relevance of such interaction is yet to be determined. Application of the latest experimental tools (**Figure 1.10**), guided by the known viral escape routes from the host's defensive strategies[55-56, 104, 157] may reveal previously unknown mechanisms that viruses use for developing resistance, adaptation and host-switching.
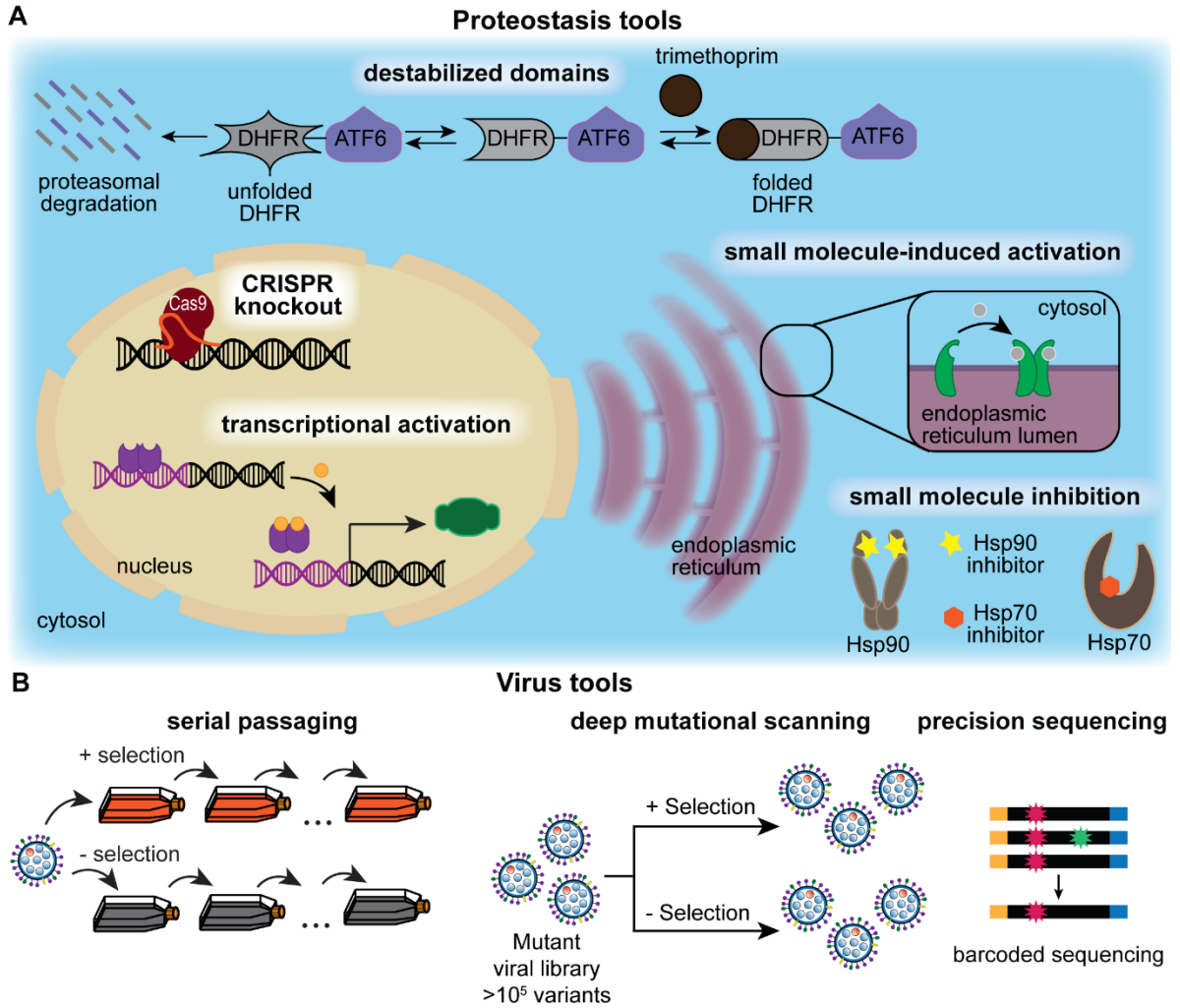
Figure 1.10. Tools for investigating the role of chaperones in viral protein evolution.
(**A**) Proteostasis tools include: chemical, genetic and chemical genetics tools for manipulating activity and levels of individual chaperones and whole proteostasis branches[152-153]. (**B**) Tools for studying viral evolution include classical serial passaging experiments for assessing the long-term evolutionary effects and deep mutational scanning coupled with low-error sequencing for high-throughput mutational profiling of viral proteins[19, 56-57, 108, 154].

1.5   Summary

Rapid and successful adaptation of viruses to changing environmental conditions is the source of a continuous global health risk. The inherent low replication fidelity enables viral proteins to efficiently overcome immune and therapeutic restriction without compromising vital protein functions. Due to resistance problem, therapeutic efforts start to expand their scope from targeting constantly changing viral proteins[29] to more evolutionary stable host factors, which are critical for viral replication[15]. However, despite extensive interactome studies in the past, a comprehensive view of how the host mechanisms guide evolution of viral proteins is still missing. Understanding this concept is a critical milestone in evolutionary biology with direct therapeutic implications.

Acquisition of new protein functions is driven by mutations, which come at a cost to protein structure. Mutations that alter protein-protein or protein-small molecule interactions can confer phenotypic benefits, such as immune escape and drug resistance. However, mutations generally destabilize protein structure and endanger proper protein folding[10-11, 85]. Compromised protein folding can have lethal consequences for viruses, which heavily depend on the optimal performance of their proteins with complex secondary and tertiary structures. Being unable to address the folding issues with a limited set of their own proteins, viruses might seek help from their vast host interactome partners. Among those, host chaperones are well-positioned to assist folding of destabilized viral proteins. Although the essential role of chaperones in viral life cycle was previously supported by limiting viral replication with genetic and pharmacological inhibition of chaperones[16-17, 149], molecular understanding of how viruses profit from interaction with chaperones is yet to be established.

Host chaperone exploitation can provide assistance in not only folding mutated viral proteins, but also in more efficient exploration of the amino acid sequence space, lifting certain restrictions imposed by protein stability. Pioneering studies on individual chaperones discovered the mutation-buffering and evolution-accelerating roles of chaperones applicable to their endogenous client proteins[120-122, 124, 150, 158]. Recent investigations reported the first observation of composition and activity of the host proteostasis affecting the balance between viral protein stability and aggregation[125], extent of mutational tolerance[156] and mutational trajectories, available to viral proteins[19]. The next important steps are to (1) uncover the detailed molecular mechanism of how chaperones resolve mutation-induced misfolding of viral proteins and (2) investigate the evolutionary implications of such interaction.

In Chapter 2, I describe how host chaperones assist influenza innate immune escape. Our high-throughput screen of influenza nucleoprotein (NP) mutant variants indicated an important adaptive mutation, linked with a high biophysical cost to the protein. Acquisition of the mutation conferring innate immune escape in 1918 Spanish flu pandemic strain destabilized the NP structure. The destabilized variant fails to efficiently propagate in the absence of external folding assistance. The unresolved structural defect outweighs the phenotypic benefit, thus compromising viral replication even in presence of the positive selection pressure from innate immune restriction factor. We demonstrate that the branch of the host proteostasis network under control of the Heat Shock Factor 1 (HSF1) transcription factor provides the necessary folding assistance and allows this variant persist and thrive in the population. In Appendix, I discuss the possible individual host factors that are likely to be exploited by the virus to resolve its biophysical defect and provide the experimental system for elucidating the molecular origins and mechanism of such folding assistance. We believe that this finding is generalizable to other viral proteins and other viruses and has important implications in viral evolution.

In Chapter 3, I demonstrate that heat shock response can restrict HIV-1 replication. We provide the first experimental evidence of decreased HIV-1 proliferation in the context of increased levels of functionally active HSF1, achieved independently of global cellular stress. We discuss the possible mechanisms underlying the observed inhibition to chart the future of detailed investigations required in this direction.

In Chapter 4, I conclude by discussing the multi-faceted role of the host chaperones in viral replication and evolution. I provide my perspective on ongoing and future directions in the field.

In Appendix, I review the efforts towards elucidating the individual host factors, important for assisting NP Pro283 variant folding by assessing the interactome of Pro283 and Ser283 variants using mass-spectrometry based proteomics. Our initial transfection-based approach for NP gene delivery proved to be unreliable due to poor efficiency and reproducibility issues. However, engineering the NP-encoding human adenovirus type 5 for NP resolved those issues and allowed us to successfully perform the first qualitative mass-spectrometry based proteomics experiment to investigate NP host interaction partners in the absence of the host proteostasis perturbation. The successful pilot interactome screen highlighted a few host factors which might be important for assisting the NP Pro283 variant folding. We are now fully prepared to systematically investigate and quantitatively compare the host interactomes of Pro283 and Ser283 in different host proteostasis environments. Such investigation would identify the HSF1-

controlled host factors, which assist in resolving the biophysical defect introduced into NP by acquisition of adaptive Pro283 mutation.

1.6   References

1.   Bloom, D. E.; Cadarette, D. Infectious disease threats in the 21st century: strengthening the global response. *Front. Immunol.* **2019,** *10*, 549.

2.   World Health Organisation (accessed August 5, 2020) Coronavirus Disease (COVID-19) Dashboard. **2020**, https://covid19.who.int/

3.   Gago, S.; Elena, S. F.; Flores, R.; Sanjuán, R. Extremely high mutation rate of a hammerhead viroid. *Science* **2009,** *323* (5919), 1308-1308.

4.   Hussain, M.; Galvin, H.; Haw, T.; Nutsford, A.; Husain, M. Drug resistance in influenza A virus: the epidemiology and management. *Infect. Drug Resist.* **2017,** *10*, 121–134.

5.   Wei, C.-J.; Crank, M. C.; Shiver, J.; Graham, B. S.; Mascola, J. R.; Nabel, G. J. Next-generation influenza vaccines: opportunities and challenges. *Nat. Rev. Drug Discov.* **2020**, 1-14.

6.   García-Dorival, I.; Wu, W.; Armstrong, S. D.; Barr, J. N.; Carroll, M. W.; Hewson, R.; Hiscox, J. A. Elucidation of the cellular interactome of Ebola virus nucleoprotein and identification of therapeutic targets. *J. Proteome Res.* **2016,** *15* (12), 4290-4303.

7.   Marques, M.; Ramos, B.; Soares, A. R.; Ribeiro, D. Cellular proteostasis during influenza A virus infection—friend or foe? *Cells* **2019,** *8* (3), 228.

8.   Mayer, K. A.; Stöckl, J.; Zlabinger, G. J.; Gualdoni, G. A. Hijacking the supplies: metabolism as a novel facet of virus-host interaction. *Front. Immunol.* **2019,** *10*, 1533.

9.   Kurahashi, R.; Sano, S.; Takano, K. Protein evolution is potentially governed by protein stability: directed evolution of an esterase from the hyperthermophilic archaeon Sulfolobus tokodaii. *J. Mol. Evol.* **2018,** *86* (5), 283-292.

10.   Tokuriki, N.; Tawfik, D. S. Stability effects of mutations and protein evolvability. *Curr. Opin. Struct. Biol.* **2009,** *19* (5), 596-604.

11.   DePristo, M. A.; Weinreich, D. M.; Hartl, D. L. Missense meanderings in sequence space: a biophysical view of protein evolution. *Nat. Rev. Genet.* **2005,** *6* (9), 678-687.

12.   Starr, T. N.; Thornton, J. W. Epistasis in protein evolution. *Prot. Sci.* **2016,** *25* (7), 1204-1218.

13.   Dasmeh, P.; Serohijos, A. W. Estimating the contribution of folding stability to nonspecific epistasis in protein evolution. *Proteins: Struct., Funct., Bioinform.* **2018,** *86* (12), 1242-1250.

14.   Jayaraj, G. G.; Hipp, M. S.; Hartl, F. U. Functional modules of the proteostasis network. *Cold Spring Habor Perspect. Biol.* **2020,** *12* (1), a033951.

15.   Aviner, R.; Frydman, J. Proteostasis in viral infection: unfolding the complex virus–chaperone interplay. *Cold Spring Harbor Perspect. Biol.* **2020,** *12* (3), a034090.

16.   Taguwa, S.; Yeh, M.-T.; Rainbolt, T. K.; Nayak, A.; Shao, H.; Gestwicki, J. E.; Andino, R.; Frydman, J. Zika virus dependence on host hsp70 provides a protective strategy against infection and disease. *Cell Rep.* **2019,** *26* (4), 906-920. e3.

17.   Geller, R.; Taguwa, S.; Frydman, J. Broad action of Hsp90 as a host chaperone required for viral replication. *BBA-Mol. Cell Res.* **2012,** *1823* (3), 698-706.

18.   Bogumil, D.; Dagan, T. Cumulative impact of chaperone-mediated folding on genome evolution. *Biochemistry* **2012,** *51* (50), 9941-9953.

19.   Phillips, A. M.; Gonzalez, L. O.; Nekongo, E. E.; Ponomarenko, A. I.; McHugh, S. M.; Butty, V. L.; Levine, S. S.; Lin, Y.-S.; Mirny, L. A.; Shoulders, M. D. Host proteostasis modulates influenza evolution. *eLife* **2017,** *6*, e28652.

20.   Raoult, D.; Forterre, P. Redefining viruses: lessons from mimivirus. *Nat. Rev. Microbiol.* **2008,** *6* (4), 315-319.

21.   Simmonds, P.; Aiewsakun, P.; Katzourakis, A. Prisoners of war—host adaptation and its constraints on virus evolution. *Nat. Rev. Microbiol.* **2019,** *17* (5), 321-328.

22.   Sharp, P. M.; Simmonds, P. Evaluating the evidence for virus/host co-evolution. *Curr. Opin. Virol.* **2011,** *1* (5), 436-441.

23.   Fleischmann Jr, W. R. Viral genetics. In *Medical Microbiology. 4th edition*, University of Texas Medical Branch at Galveston: **1996**.

24.   Dolan, P. T.; Whitfield, Z. J.; Andino, R. Mapping the evolutionary potential of RNA viruses. *Cell Host Microbe* **2018,** *23* (4), 435-446.

25.   Retel, C.; Märkle, H.; Becks, L.; Feulner, P. G. Ecological and evolutionary processes shaping viral genetic diversity. *Viruses* **2019,** *11* (3), 220.

26.   Duffy, S.; Shackelton, L. A.; Holmes, E. C. Rates of evolutionary change in viruses: patterns and determinants. *Nat. Rev. Genet.* **2008,** *9* (4), 267-276.

27.   Choi, K. H. Viral polymerases. In *Viral Molecular Machines*, Springer: **2012**; pp 267-304.

28.   Menéndez-Arias, L.; Sebastián-Martín, A.; Álvarez, M., Viral reverse transcriptases. *Virus Res.* **2017,** *234*, 153-176.

29.   Sanjuán, R.; Nebot, M. R.; Chirico, N.; Mansky, L. M.; Belshaw, R. Viral mutation rates. *J. Virol.* **2010,** *84* (19), 9733-9748.

30.   Pauly, M. D.; Procario, M. C.; Lauring, A. S. A novel twelve class fluctuation test reveals higher than expected mutation rates for influenza A viruses. *eLife* **2017,** *6*, e26437.

31.   Cuevas, J. M.; Geller, R.; Garijo, R.; López-Aldeguer, J.; Sanjuán, R. Extremely high mutation rate of HIV-1 in vivo. *PLoS Biol.* **2015,** *13* (9), e1002251.

32.   Jee, J.; Rasouly, A.; Shamovsky, I.; Akivis, Y.; Steinman, S. R.; Mishra, B.; Nudler, E. Rates and mechanisms of bacterial mutagenesis from maximum-depth sequencing. *Nature* **2016,** *534* (7609), 693-696.

33.   Drake, J. W.; Charlesworth, B.; Charlesworth, D.; Crow, J. F. Rates of spontaneous mutation. *Genetics* **1998,** *148* (4), 1667-1686.

34.   Takeuchi, O.; Akira, S. Innate immunity to virus infection. *Immunol. Rev.* **2009,** *227* (1), 75-86.

35.   Lee, H.-C.; Chathuranga, K.; Lee, J.-S. Intracellular sensing of viral genomes and viral evasion. *Exp. Mol. Med.* **2019,** *51* (12), 1-13.

36.   Beachboard, D. C.; Horner, S. M. Innate immune evasion strategies of DNA and RNA viruses. *Curr. Opin. Microbiol.* **2016,** *32*, 113-119.

37.   García-Sastre, A., Ten strategies of interferon evasion by viruses. *Cell Host Microbe* **2017,** *22* (2), 176-184.

38.   Versteeg, G. A.; García-Sastre, A. Viral tricks to grid-lock the type I interferon system. *Curr. Opin. Microbiol.* **2010,** *13* (4), 508-516.

39.   Kawai, T.; Akira, S. Innate immune recognition of viral infection. *Nat. Immunol.* **2006,** *7* (2), 131-137.

40.   Schoggins, J. W.; Wilson, S. J.; Panis, M.; Murphy, M. Y.; Jones, C. T.; Bieniasz, P.; Rice, C. M. A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature* **2011,** *472* (7344), 481-485.

41.   Ivashkiv, L. B.; Donlin, L. T.,Regulation of type I interferon responses. *Nat. Rev. Immunol.* **2014,** *14* (1), 36-49.

42.   Rodriguez-Madoz, J. R.; Belicha-Villanueva, A.; Bernal-Rubio, D.; Ashour, J.; Ayllon, J.; Fernandez-Sesma, A. Inhibition of the type I interferon response in human dendritic cells by dengue virus infection requires a catalytically active NS2B3 complex. *J. Virol.* **2010,** *84* (19), 9760-9774.

43.   Marc, D. Influenza virus non-structural protein NS1: interferon antagonism and beyond. *J. Gen. Virol.* **2014,** *95* (12), 2594-2611.

44.   Hollidge, B. S.; Weiss, S. R.; Soldan, S. S. The role of interferon antagonist, non-structural proteins in the pathogenesis and emergence of arboviruses. *Viruses* **2011,** *3* (6), 629-658.

45.   García-Sastre, A. Inhibition of interferon-mediated antiviral responses by influenza A viruses and other negative-strand RNA viruses. *Virol. J.* **2001,** *279* (2), 375-384.

46.   Fonseca, G.; Thillainadesan, G.; Yousef, A.; Ablack, J.; Mossman, K.; Torchia, J.; Mymryk, J. Adenovirus evasion of interferon-mediated innate immunity by direct antagonism of a cellular histone posttranslational modification. *Cell Host Microbe* **2012,** *11* (6), 597-606.

47.   Doyle, T.; Goujon, C.; Malim, M. H. HIV-1 and interferons: who's interfering with whom? *Nat. Rev. Microbiol.* **2015,** *13* (7), 403-413.

48.   Duggal, N. K.; Emerman, M. Evolutionary conflicts between viruses and restriction factors shape immunity. *Nat. Rev. Immunol.* **2012,** *12* (10), 687-695.

49.   Mänz, B.; Dornfeld, D.; Götz, V.; Zell, R.; Zimmermann, P.; Haller, O.; Kochs, G.; Schwemmle, M. Pandemic influenza A viruses escape from restriction by human MxA through adaptive mutations in the nucleoprotein. *PLoS Pathog.* **2013,** *9* (3), e1003279.

50.   French, A. R.; Pingel, J. T.; Wagner, M.; Bubic, I.; Yang, L.; Kim, S.; Koszinowski, U.; Jonjic, S.; Yokoyama, W. M. Escape of mutant double-stranded DNA virus from innate immune control. *Immunity* **2004,** *20* (6), 747-756.

51.   Xia, H.; Luo, H.; Shan, C.; Muruato, A. E.; Nunes, B. T.; Medeiros, D. B.; Zou, J.; Xie, X.; Giraldo, M. I.; Vasconcelos, P. F. An evolutionary NS1 mutation enhances Zika virus evasion of host interferon induction. *Nat. Comm.* **2018,** *9* (1), 1-13.

52.   Janeway, C. A.; Travers, P.; Walport, M.; Shlomchik, M. J. Immunobiology: the immune system in health and disease. 5th ed.; Garland Science: New York, **2001**.

53.   Law, M.; Hangartner, L. Antibodies against viruses: passive and active immunization. *Curr. Opin. Immunol.* **2008,** *20* (4), 486-492.

54.   Murin, C. D.; Wilson, I. A.; Ward, A. B. Antibody responses to viral infections: a structural perspective across three different enveloped viruses. *Nat. Microbiol.* **2019,** *4* (5), 734-747.

55.   Lee, J. M.; Eguia, R.; Zost, S. J.; Choudhary, S.; Wilson, P. C.; Bedford, T.; Stevens-Ayers, T.; Boeckh, M.; Hurt, A. C.; Lakdawala, S. S. et al. Mapping person-to-person variation in viral mutations that escape polyclonal serum targeting influenza hemagglutinin. *eLife* **2019,** *8*, e49324.

56.   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. Comm.* **2018,** *9* (1), 1-12.

57.   Dingens, A. S.; Arenz, D.; Weight, H.; Overbaugh, J.; Bloom, J. D. An antigenic atlas of HIV-1 escape from broadly neutralizing antibodies distinguishes functional and structural epitopes. *Immunity* **2019,** *50* (2), 520-532. e3.

58.   Sourisseau, M.; Lawrence, D. J.; Schwarz, M. C.; Storrs, C. H.; Veit, E. C.; Bloom, J. D.; Evans, M. J. Deep mutational scanning comprehensively maps how Zika envelope protein mutations affect viral growth and antibody escape. *J. Virol.* **2019,** *93* (23), e01291-19.

59.   Wu, N. C.; Qi, H. Application of deep mutational scanning in hepatitis C virus. In *Hepatitis C Virus Protocols*, Springer: **2019**; pp 183-190.

60.   Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. The adaptive immune system. *Mol. Biol. Cell* **2002,** *4*, 1363-1402.

61.   Rosendahl Huber, S.; van Beek, J.; de Jonge, J.; Luytjes, W.; van Baarle, D. T cell responses to viral infections–opportunities for peptide vaccination. *Front. Immunol.* **2014,** *5*, 171.

62.   Yewdell, J. W.; Bennink, J. R.; Smith, G. L.; Moss, B. Influenza A virus nucleoprotein is a major target antigen for cross-reactive anti-influenza A virus cytotoxic T lymphocytes. *Proc. Natl. Acad. Sci. U.S.A.* **1985,** *82* (6), 1785-1789.

63.   Onion, D.; Crompton, L. J.; Milligan, D. W.; Moss, P. A.; Lee, S. P.; Mautner, V. The CD4+ T-cell response to adenovirus is focused against conserved residues within the hexon protein. *J. Gen. Virol.* **2007,** *88* (9), 2417-2425.

64.   Assarsson, E.; Bui, H.-H.; Sidney, J.; Zhang, Q.; Glenn, J.; Oseroff, C.; Mbawuike, I. N.; Alexander, J.; Newman, M. J.; Grey, H. Immunomic analysis of the repertoire of T-cell specificities for influenza A virus in humans. *J. Virol.* **2008,** *82* (24), 12241-12251.

65.   Wentworth, P. A.; Sette, A.; Celis, E.; Sidney, J.; Southwood, S.; Crimi, C.; Stitely, S.; Keogh, E.; Wong, N. C.; Livingston, B. Identification of A2-restricted hepatitis C virus-specific cytotoxic T lymphocyte epitopes from conserved regions of the viral genome. *Int. Immunol.* **1996,** *8* (5), 651-659.

66.   Liu, Y.; McNevin, J.; Rolland, M.; Zhao, H.; Deng, W.; Maenza, J.; Stevens, C. E.; Collier, A. C.; McElrath, M. J.; Mullins, J. I. Conserved HIV-1 epitopes continuously elicit subdominant cytotoxic T-lymphocyte responses. *J. Infect. Dis.* **2009,** *200* (12), 1825-1833.

67.   Bowen, D. G.; Walker, C. M. Mutational escape from CD8+ T cell immunity: HCV evolution, from chimpanzees to man. *J. Exp. Med.* **2005,** *201* (11), 1709-1714.

68.   Streeck, H.; Li, B.; Poon, A. F.; Schneidewind, A.; Gladden, A. D.; Power, K. A.; Daskalakis, D.; Bazner, S.; Zuniga, R.; Brander, C. Immune-driven recombination and loss of control after HIV superinfection. *J. Exp. Med.* **2008,** *205* (8), 1789-1796.

69.   Lehnert, R.; Pletz, M.; Reuss, A.; Schaberg, T. Antiviral medications in seasonal and pandemic influenza: a systematic review. *Dtsch. Ärztebl. Int.* **2016,** *113* (47), 799.

70.   Gulick, R. M.; Flexner, C. Long-acting HIV drugs for treatment and prevention. *Ann. Rev. Med.* **2019,** *70*, 137-150.

71.   Hanazaki, K. Antiviral therapy for chronic hepatitis B: a review. *Curr. Drug Targets-Inflamm. Allergy* **2004,** *3* (1), 63-70.

72.   Holmes, J. A.; Rutledge, S. M.; Chung, R. T. Direct-acting antiviral treatment for hepatitis C. *Lancet* **2019,** *393* (10179), 1392-1394.

73.   Irwin, K. K.; Renzette, N.; Kowalik, T. F.; Jensen, J. D. Antiviral drug resistance as an adaptive process. *Virus Evol.* **2016,** *2* (1), vew014.

74.   van der Vries, E.; Ison, M. G. Antiviral resistance in influenza viruses: clinical and epidemiological aspects. In *Microb. Drug Resist.*, Springer: **2017**; pp 1165-1183.

75.   Du, Y.; Xin, L.; Shi, Y.; Zhang, T.-H.; Wu, N. C.; Dai, L.; Gong, D.; Brar, G.; Shu, S.; Luo, J. Genome-wide identification of interferon-sensitive mutations enables influenza vaccine design. *Science* **2018,** *359* (6373), 290-296.

76.   Weber, M.; Sediri, H.; Felgenhauer, U.; Binzen, I.; Bänfer, S.; Jacob, R.; Brunotte, L.; García-Sastre, A.; Schmid-Burgk, J. L.; Schmidt, T. Influenza virus adaptation PB2-627K modulates nucleocapsid inhibition by the pathogen sensor RIG-I. *Cell Host Microbe* **2015,** *17* (3), 309-319.

77.   Das, S. R.; Hensley, S. E.; David, A.; Schmidt, L.; Gibbs, J. S.; Puigbò, P.; Ince, W. L.; Bennink, J. R.; Yewdell, J. W. Fitness costs limit influenza A virus hemagglutinin glycosylation as an immune evasion strategy. *Proc. Natl. Acad. Sci. U.S.A.* **2011,** *108* (51), E1417-E1422.

78.   Lynch, R. M.; Wong, P.; Tran, L.; O'Dell, S.; Nason, M. C.; Li, Y.; Wu, X.; Mascola, J. R. HIV-1 fitness cost associated with escape from the VRC01 class of CD4 binding site neutralizing antibodies. *J. Virol.* **2015,** *89* (8), 4201-4213.

79.   Miura, T.; Brockman, M. A.; Schneidewind, A.; Lobritz, M.; Pereyra, F.; Rathod, A.; Block, B. L.; Brumme, Z. L.; Brumme, C. J.; Baker, B. HLA-B57/B* 5801 human immunodeficiency virus type 1 elite controllers select for rare gag variants associated with reduced viral replication capacity and strong cytotoxic T-lymphotye recognition. *J. Virol.* **2009,** *83* (6), 2743-2755.

80.   Leslie, A.; Pfafferott, K.; Chetty, P.; Draenert, R.; Addo, M.; Feeney, M.; Tang, Y.; Holmes, E.; Allen, T.; Prado, J. HIV evolution: CTL escape mutation and reversion after transmission. *Nat. Med.* **2004,** *10* (3), 282-289.

81.   Maisnier-Patin, S.; Andersson, D. I. Adaptation to the deleterious effects of antimicrobial drug resistance mutations by compensatory evolution. *Res. Microbiol.* **2004,** *155* (5), 360-369.

82.   Neagu, I. A.; Olejarz, J.; Freeman, M.; Rosenbloom, D. I.; Nowak, M. A.; Hill, A. L. Life cycle synchronization is a viral drug resistance mechanism. *PLoS Comp. Biol.* **2018,** *14* (2).

83.   Eigen, M. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **1971,** *58* (10), 465-523.

84.	Domingo, E.; Sheldon, J.; Perales, C. Viral quasispecies evolution. *Microbiol. Mol. Biol. Rev.* **2012,** *76* (2), 159-216.

85.	Crotty, S.; Cameron, C. E.; Andino, R. RNA virus error catastrophe: direct molecular test by using ribavirin. *Proc. Natl. Acad. Sci. U.S.A.* **2001,** *98* (12), 6895-6900.

86.	Baranovich, T.; Wong, S.-S.; Armstrong, J.; Marjuki, H.; Webby, R. J.; Webster, R. G.; Govorkova, E. A. T-705 (favipiravir) induces lethal mutagenesis in influenza A H1N1 viruses in vitro. *J. Virol.* **2013,** *87* (7), 3741-3751.

87.	Okada, A.; Iwatani, Y. APOBEC3G-mediated G-to-A hypermutation of the HIV-1 genome: the missing link in antiviral molecular mechanisms. *Front. Microbiol.* **2016,** *7,* 2027.

88.	Rai, D. K.; Diaz-San Segundo, F.; Campagnola, G.; Keith, A.; Schafer, E. A.; Kloc, A.; de los Santos, T.; Peersen, O.; Rieder, E. Attenuation of foot-and-mouth disease virus by engineered viral polymerase fidelity. *J. Virol.* **2017,** *91* (15), e00081-17.

89.	Domingo, E.; Holland, J. RNA virus mutations and fitness for survival. *Ann. Rev. Microbiol.* **1997,** *51* (1), 151-178.

90.	Rice, D. P.; Good, B. H.; Desai, M. M. The evolutionarily stable distribution of fitness effects. *Genetics* **2015,** *200* (1), 321-329.

91.	Wright, S. The roles of mutation, inbreeding, crossbreeding, and selection in evolution. *Proc. 6th Int. Cong. Genet.* **1932,** *1*, 356-366.

92.	Chéron, N.; Serohijos, A. W.; Choi, J. M.; Shakhnovich, E. I. Evolutionary dynamics of viral escape under antibodies stress: a biophysical model. *Prot. Sci.* **2016,** *25* (7), 1332-1340.

93.	Louie, R. H.; Kaczorowski, K. J.; Barton, J. P.; Chakraborty, A. K.; McKay, M. R. Fitness landscape of the human immunodeficiency virus envelope protein that is targeted by antibodies. *Proc. Natl. Acad. Sci. U. S. A.* **2018,** *115* (4), E564-E573.

94.	Lauring, A. S.; Andino, R. Exploring the fitness landscape of an RNA virus by using a universal barcode microarray. *J. Virol.* **2011,** *85* (8), 3780-3791.

95.	Rodrigues, J. V.; Bershtein, S.; Li, A.; Lozovsky, E. R.; Hartl, D. L.; Shakhnovich, E. I. Biophysical principles predict fitness landscapes of drug resistance. *Proc. Natl. Acad. Sci. U.S.A.* **2016,** *113* (11), E1470-E1478.

96.	Colón, W.; Church, J.; Sen, J.; Thibeault, J.; Trasatti, H.; Xia, K. Biological roles of protein kinetic stability. *Biochemistry* **2017,** *56* (47), 6179-6186.

97.	Guo, H. H.; Choe, J.; Loeb, L. A. Protein tolerance to random amino acid change. *Proc. Natl. Acad. Sci. U. S. A.* **2004,** *101* (25), 9205-9210.

98.	Humphris-Narayanan, E.; Akiva, E.; Varela, R.; Conchúir, S. Ó.; Kortemme, T. Prediction of mutational tolerance in HIV-1 protease and reverse transcriptase using flexible backbone protein design. *PLoS Comp. Biol.* **2012,** *8* (8), e1002639.

99.	Thyagarajan, B.; Bloom, J. D. The inherent mutational tolerance and antigenic evolvability of influenza hemagglutinin. *eLife* **2014,** *3*, e03300.

100.	Schrag, S. J.; Rota, P. A.; Bellini, W. J. Spontaneous mutation rate of measles virus: direct estimation based on mutations conferring monoclonal antibody resistance. *J. Virol.* **1999,** *73* (1), 51-54.

101.	Yewdell, J.; Webster, R.; Gerhard, W. Antigenic variation in three distinct determinants of an influenza type A haemagglutinin molecule. *Nature* **1979,** *279* (5710), 246-248.

102.	Birrer, M. J.; Udem, S.; Nathenson, S.; Bloom, B. R. Antigenic variants of measles virus. *Nature* **1981,** *293* (5827), 67-69.

103.	Heaton, N. S.; Sachs, D.; Chen, C.-J.; Hai, R.; Palese, P. Genome-wide mutagenesis of influenza virus reveals unique plasticity of the hemagglutinin and NS1 proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2013,** *110* (50), 20248-20253.

104.	Doud, M. B.; Hensley, S. E.; Bloom, J. D. Complete mapping of viral escape from neutralizing antibodies. *PLoS Pathog.* **2017,** *13* (3), e1006271.

105.   Fulton, B. O.; Sachs, D.; Beaty, S. M.; Won, S. T.; Lee, B.; Palese, P.; Heaton, N. S. Mutational analysis of measles virus suggests constraints on antigenic variation of the glycoproteins. *Cell Rep.* **2015,** *11* (9), 1331-1338.

106.   Baczko, K.; Pardowitz, I.; Rima, B. K.; Ter Meulen, V. Constant and variable regions of measles virus proteins encoded by the nucleocapsid and phosphoprotein genes derived from lytic and persistent viruses. *Virol. J.* **1992,** *190* (1), 469-474.

107.   Gong, L. I.; Bloom, J. D. Epistatically interacting substitutions are enriched during adaptive protein evolution. *PLoS Genet.* **2014,** *10* (5), e1004328.

108.   Duan, S.; Govorkova, E. A.; Bahl, J.; Zaraket, H.; Baranovich, T.; Seiler, P.; Prevost, K.; Webster, R. G.; Webby, R. J. Epistatic interactions between neuraminidase mutations facilitated the emergence of the oseltamivir-resistant H1N1 influenza viruses. *Nat. Comm.* **2014,** *5* (1), 1-12.

109.   Gong, L. I.; Suchard, M. A.; Bloom, J. D. Stability-mediated epistasis constrains the evolution of an influenza protein. *eLife* **2013,** *2*, e00631.

110.   Wu, N. C.; Dai, L.; Olson, C. A.; Lloyd-Smith, J. O.; Sun, R. Adaptation in protein fitness landscapes is facilitated by indirect paths. *eLife* **2016,** *5*, e16965.

111.   Tomala, K.; Zrebiec, P.; Hartl, D. L. Limits to compensatory mutations: Insights from temperature-sensitive alleles. *Mol. Biol. Evol.* **2019,** *36* (9), 1874-1883.

112.   Sanjuán, R.; Moya, A.; Elena, S. F. The contribution of epistasis to the architecture of fitness in an RNA virus. *Proc. Natl. Acad. Sci. U. S. A.* **2004,** *101* (43), 15376-15379.

113.   Sikosek, T.; Chan, H. S. Biophysics of protein evolution and evolutionary protein biophysics. *J. R. Soc. Interface* **2014,** *11* (100), 20140419.

114.   Alexander, P. A.; He, Y.; Chen, Y.; Orban, J.; Bryan, P. N. A minimal sequence code for switching protein structure and function. *Proc. Natl. Acad. Sci. U. S. A.* **2009,** *106* (50), 21149-21154.

115.   Bloom, J. D.; Gong, L. I.; Baltimore, D. Permissive secondary mutations enable the evolution of influenza oseltamivir resistance. *Science* **2010,** *328* (5983), 1272-1275.

116.   Sebastian, R. M.; Shoulders, M. D. Chemical biology framework to illuminate proteostasis. *Annu. Rev. Biochem.* **2020,** *89*, In press.

117.   Wong, M. Y.; DiChiara, A. S.; Suen, P. H.; Chen, K.; Doan, N.-D.; Shoulders, M. D. Adapting secretory proteostasis and function through the unfolded protein response. In *Coordinating Organismal Physiology Through the Unfolded Protein Response*, Springer: **2017**; pp 1-25.

118.   Moehle, E. A.; Shen, K.; Dillin, A. Mitochondrial proteostasis in the context of cellular and organismal health and aging. *J. Biol. Chem.* **2019,** *294* (14), 5396-5407.

119.   Shpilka, T.; Haynes, C. M. The mitochondrial UPR: mechanisms, physiological functions and implications in ageing. *Nat. Rev. Mol. Cell Biol.* **2018,** *19* (2), 109.

120.   Rutherford, S. L.; Lindquist, S. Hsp90 as a capacitor for morphological evolution. *Nature* **1998,** *396* (6709), 336-342.

121.   Rohner, N.; Jarosz, D. F.; Kowalko, J. E.; Yoshizawa, M.; Jeffery, W. R.; Borowsky, R. L.; Lindquist, S.; Tabin, C. J. Cryptic variation in morphological evolution: HSP90 as a capacitor for loss of eyes in cavefish. *Science* **2013,** *342* (6164), 1372-1375.

122.   Tokuriki, N.; Tawfik, D. S. Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature* **2009,** *459* (7247), 668-673.

123.   Aguilar-Rodríguez, J.; Sabater-Muñoz, B.; Montagud-Martínez, R.; Berlanga, V.; Alvarez-Ponce, D.; Wagner, A.; Fares, M. A. The molecular chaperone DnaK is a source of mutational robustness. *Genome Biol. Evol.* **2016,** *8* (9), 2979-2991.

124.   Alvarez-Ponce, D.; Aguilar-Rodríguez, J.; Fares, M. A. Molecular chaperones accelerate the evolution of their protein clients in yeast. *Genome Biol. Evol.* **2019,** *11* (8), 2360-2375.

125.   Geller, R.; Pechmann, S.; Acevedo, A.; Andino, R.; Frydman, J. Hsp90 shapes protein and RNA evolution to balance trade-offs between protein stability and aggregation. *Nat. Comm.* **2018,** *9* (1), 1-11.

126.   Pechmann, S.; Frydman, J. Interplay between chaperones and protein disorder promotes the evolution of protein networks. *PLoS Comp. Biol.* **2014,** *10* (6).

127.   Wylie, C. S.; Shakhnovich, E. I. Mutation induced extinction in finite populations: lethal mutagenesis and lethal isolation. *PLoS Comp. Biol.* **2012,** *8* (8).

128.   Cobbold, C.; Windsor, M.; Wileman, T. A virally encoded chaperone specialized for folding of the major capsid protein of African swine fever virus. *J. Virol.* **2001,** *75* (16), 7221-7229.

129.   Chabaud, S.; Lambert, H.; Sasseville, A. M.-J.; Lavoie, H.; Guilbault, C.; Massie, B.; Landry, J.; Langelier, Y. The R1 subunit of herpes simplex virus ribonucleotide reductase has chaperone-like activity similar to Hsp27. *FEBS Lett.* **2003,** *545* (2-3), 213-218.

130.   Suzuki, H. A hypothesis about the mechanism of assembly of double-shelled rotavirus particles. In *Viral Gastroenteritis*, Springer: **1996**; pp 79-85.

131.   Latorre, V.; Mattenberger, F.; Geller, R. Chaperoning the mononegavirales: current knowledge and future directions. *Viruses* **2018,** *10* (12), 699.

132.   Hebert, D. N.; Foellmer, B.; Helenius, A. Calnexin and calreticulin promote folding, delay oligomerization and suppress degradation of influenza hemagglutinin in microsomes. *EMBO J.* **1996,** *15* (12), 2961-2968.

133.   Bozzacco, L.; Yi, Z.; Andreo, U.; Conklin, C. R.; Li, M. M.; Rice, C. M.; MacDonald, M. R. Chaperone-assisted protein folding is critical for yellow fever virus NS3/4A cleavage and replication. *J. Virol.* **2016,** *90* (6), 3212-3228.

134.   Doms, R. W.; Lamb, R. A.; Rose, J. K.; Helenius, A. Folding and assembly of viral membrane proteins. *Virol. J.* **1993,** *193* (2), 545-562.

135.   Dubuisson, J.; Rice, C. M. Hepatitis C virus glycoprotein folding: disulfide bond formation and association with calnexin. *J. Virol.* **1996,** *70* (2), 778-786.

136.   Wei, X.; Decker, J. M.; Wang, S.; Hui, H.; Kappes, J. C.; Wu, X.; Salazar-Gonzalez, J. F.; Salazar, M. G.; Kilby, J. M.; Saag, M. S. Antibody neutralization and escape by HIV-1. *Nature* **2003,** *422* (6929), 307-312.

137.   Kosik, I.; Ince, W. L.; Gentles, L. E.; Oler, A. J.; Kosikova, M.; Angel, M.; Magadan, J. G.; Xie, H.; Brooke, C. B.; Yewdell, J. W. Influenza A virus hemagglutinin glycosylation compensates for antibody escape fitness costs. *PLoS Pathog.* **2018,** *14* (1), e1006796.

138.   Hou, S.; Kumar, A.; Xu, Z.; Airo, A. M.; Stryapunina, I.; Wong, C. P.; Branton, W.; Tchesnokov, E.; Götte, M.; Power, C. Zika virus hijacks stress granule proteins and modulates the host stress response. *J. Virol.* **2017,** *91* (16), e00474-17.

139.   Hassan, I. H.; Zhang, M. S.; Powers, L. S.; Shao, J. Q.; Baltrusaitis, J.; Rutkowski, D. T.; Legge, K.; Monick, M. M. Influenza A viral replication is blocked by inhibition of the inositol-requiring enzyme 1 (IRE1) stress pathway. *J. Biol. Chem.* **2012,** *287* (7), 4679-4689.

140.   Fung, T. S.; Huang, M.; Liu, D. X. Coronavirus-induced ER stress response and its involvement in regulation of coronavirus–host interactions. *Virus Res.* **2014,** *194*, 110-123.

141.   Shu, W.; Guo, Z.; Li, L.; Xiong, Z.; Wang, Z.; Yang, Y.; Li, Y.; He, M.; Gong, R.; Gao, B. Regulation of molecular chaperone GRP78 by hepatitis B virus: control of viral replication and cell survival. *Mol. Cell. Biol.* **2020,** *40* (3), e00475-19.

142.   Guerrero, C. A.; Bouyssounade, D.; Zárate, S.; Iša, P.; López, T.; Espinosa, R.; Romero, P.; Méndez, E.; López, S.; Arias, C. F. Heat shock cognate protein 70 is involved in rotavirus cell entry. *J. Virol.* **2002,** *76* (8), 4096-4102.

143.   Schelhaas, M.; Malmström, J.; Pelkmans, L.; Haugstetter, J.; Ellgaard, L.; Grünewald, K.; Helenius, A. Simian Virus 40 depends on ER protein folding and quality control factors for entry into host cells. *Cell* **2007,** *131* (3), 516-529.

144.   Agostini, I.; Popov, S.; Li, J.; Dubrovsky, L.; Hao, T.; Bukrinsky, M. Heat-shock protein 70 can replace viral protein R of HIV-1 during nuclear import of the viral preintegration complex. *Exp. Cell Res.* **2000,** *259* (2), 398-403.
145.   Momose, F.; Naito, T.; Yano, K.; Sugimoto, S.; Morikawa, Y.; Nagata, K. Identification of Hsp90 as a stimulatory host factor involved in influenza virus RNA synthesis. *J. Biol. Chem.* **2002,** *277* (47), 45306-45314.
146.   Kumar, M.; Mitra, D. Heat shock protein 40 is necessary for human immunodeficiency virus-1 Nef-mediated enhancement of viral gene expression and replication. *J. Biol. Chem.* **2005,** *280* (48), 40041-40050.
147.   Chromy, L. R.; Pipas, J. M.; Garcea, R. L. Chaperone-mediated in vitro assembly of polyomavirus capsids. *Proc. Natl. Acad. Sci. U. S. A.* **2003,** *100* (18), 10477-10482.
148.   Streblow, D. N.; Kitabwalla, M.; Malkovsky, M.; Pauza, C. D. Cyclophilin A modulates processing of human immunodeficiency virus type 1 p55Gag: mechanism for antiviral effects of cyclosporin A. *Virol. J.* **1998,** *245* (2), 197-202.
149.   Mayer, M. Recruitment of Hsp70 chaperones: a crucial part of viral survival strategies. In *Rev. Phys. Bioch. Pharm.*, Springer: **2005**; pp 1-46.
150.   Agozzino, L.; Dill, K. A. Protein evolution speed depends on its stability and abundance and on chaperone concentrations. *Proc. Natl. Acad. Sci. U. S. A.* **2018,** *115* (37), 9092-9097.
151.   Williams, T. A.; Fares, M. A. The effect of chaperonin buffering on protein evolution. *Genome Biol. Evol.* **2010,** *2*, 609-619.
152.   Lachowiec, J.; Lemus, T.; Thomas, J. H.; Murphy, P. J.; Nemhauser, J. L.; Queitsch, C. The protein chaperone HSP90 can facilitate the divergence of gene duplicates. *Genetics* **2013,** *193* (4), 1269-1277.
153.   Moore, C. L.; Dewal, M. B.; Nekongo, E. E.; Santiago, S.; Lu, N. B.; Levine, S. S.; Shoulders, M. D. Transportable, chemical genetic methodology for the small molecule-mediated inhibition of heat shock factor 1. *ACS Chem. Biol.* **2016,** *11* (1), 200-210.
154.   Shoulders, M. D.; Ryno, L. M.; Cooley, C. B.; Kelly, J. W.; Wiseman, R. L. Broadly applicable methodology for the rapid and dosable small molecule-mediated regulation of transcription factors in human cells. *J. Am. Chem. Soc.* **2013,** *135* (22), 8129-8132.
155.   Fowler, D. M.; Fields, S. Deep mutational scanning: a new style of protein science. *Nat. Methods* **2014,** *11* (8), 801.
156.   Phillips, A. M.; Doud, M. B.; Gonzalez, L. O.; Butty, V. L.; Lin, Y.-S.; Bloom, J. D.; Shoulders, M. D. Enhanced ER proteostasis and temperature differentially impact the mutational tolerance of influenza hemagglutinin. *eLife* **2018,** *7*, e38795.
157.   Ashenberg, O.; Padmakumar, J.; Doud, M. B.; Bloom, J. D. Deep mutational scanning identifies sites in influenza nucleoprotein that affect viral inhibition by MxA. *PLoS Pathog.* **2017,** *13* (3), e1006288.
158.   Sangster, T. A.; Salathia, N.; Undurraga, S.; Milo, R.; Schellenberg, K.; Lindquist, S.; Queitsch, C. HSP90 affects the expression of genetic variation and developmental stability in quantitative traits. *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105* (8), 2963-2968.

Chapter 2: Host chaperones address the biophysical cost of influenza innate immune escape

2.1 Author contributions

A.M.P. and M.D.S. conceived the project. A.I.P., A.M.P., K.C., and M.D.S. planned research, designed experiments, and analyzed data. A.I.P., A.M.P., and K.C. performed experiments. O.L.A., C.L.M., and J.D.B. provided reagents and designed experiments. V.L.B. and C.W. analyzed sequencing results. J.M., S.M.M., and Y.-S.L. designed, performed, and analyzed molecular dynamics simulations. J.D.B., Y.-S.L., and M.D.S. supervised research.

2.2. Introduction

RNA viruses are exceptionally efficient pathogens that leverage host machineries to replicate their genetic material, synthesize their proteins, and assemble new virions. Perhaps their most remarkable feature, however, is the capacity to rapidly evolve in the face of environmental and immune system pressures. Such rapid evolution is largely mediated by a high mutation rate[1].

Despite its adaptive benefits, rapid genetic change comes at a significant cost for evolving proteins. The majority of amino acid substitutions (especially functionally relevant substitutions that alter or create protein activities) are biophysically deleterious, negatively affecting either protein folding or stability [2-6].

A striking illustration of this phenomenon in RNA viruses is influenza nucleoprotein. Nucleoprotein is a globular protein that oligomerizes to encapsulate influenza genomic material and mediate its import into the host nucleus, a process that is required for transcription and replication of the viral genome [7]. Nucleoprotein is strongly conserved relative to the highly variable influenza surface proteins targeted by antibodies [8]. However, nucleoprotein does experience significant selection pressure from the host immune system, including from innate immune restriction factors [9-10]. In particular, the human restriction factor MxA can prevent influenza ribonucleoprotein import [11-13], cutting short the viral replication cycle.

Adaptive mutations in nucleoprotein that allow escape from human MxA are critical for the efficient replication of new influenza strains in humans following zoonotic transmission. In just the last century, non-human influenza nucleoprotein was introduced into circulating human influenza strains in 2009 and probably in 1918 [14-15], leading to the acquisition of MxA resistance and ultimately to global pandemics [16-17]. While nucleoprotein evolution is driven by immune escape, it is nonetheless clear that a delicate balance exists between immune system resistance and nucleoprotein stability and folding. Several nucleoprotein substitutions known to engender immune escape are destabilizing [9-10], impairing viral growth in the absence of immune pressure [16]. Apparently, the evolving virus must balance the costs of a nucleoprotein folding defect with the benefits of escaping host immunity.

In theory, any mechanism that allows influenza (or other viruses) to uncouple protein folding versus immune escape selection pressures would have tremendous benefits for the pathogen. Recent work by us and others suggests that the host's Hsp90 chaperone can modulate the evolutionary paths traversed by viruses [18-19]. Neither the mechanism of Hsp90's

effect on viral evolution nor its relevance to actual viral strains is clear. A provocative possibility, not yet experimentally explored, is that hijacked host chaperones, whether Hsp90 or any of the other dozens of chaperones beyond Hsp90 that interact extensively with influenza [20-21], potentiate viral evolution directly by assisting the folding of biophysically defective nucleoprotein variants that would otherwise be insufficiently fit to persist in the population. If this hypothesis is correct, it would suggest that subversion of host protein folding chaperones by viruses can make otherwise inaccessible mutational trajectories leading to immune system escape possible, specifically by promoting the folding of escape variants. Such a mechanism for host adaptation would have broad implications for the evolution of not only influenza nucleoprotein, but also other influenza proteins and other viruses.

Rigorously testing this hypothesis requires a method to systematically and quantitatively evaluate whether and how host proteostasis modulates viral protein mutational tolerance. Here, we achieve that objective using deep mutational scanning [22] of influenza nucleoprotein. We apply deep mutational scanning in the context of chemical genetic inhibition of the host's heat shock factor 1 (HSF1 [23]) to create biophysically challenging, chaperone-depleted cellular protein folding environments. This high-throughput approach revealed a number of amino acid positions in influenza nucleoprotein whose mutational tolerance is strongly reduced in chaperone-depleted host cells. We confirmed the strong effects of host chaperones on nucleoprotein mutational tolerance at a number of these sites using head-to-head competition experiments. Most strikingly, the strongly conserved proline (Pro) residue at site 283 in nucleoprotein is rendered highly unfit by HSF1 inhibition at febrile temperatures. Pro283 in nucleoprotein is known to facilitate escape from the human innate immune system restriction factor MxA [9, 16], a feature that critically enhanced the pathogenicity and fitness of the 1918 pandemic influenza strain. We further show that Pro283 disrupts a key structural element in nucleoprotein, rendering it unstable and aggregation-prone. Host chaperones resolve this folding problem, allowing Pro283 to persist in the viral population and thereby promoting MxA escape.

Collectively, our data demonstrate that viral hijacking of host chaperones addresses critical biophysical defects that would otherwise sensitize the virus to host restriction factors. This phenomenon thus has tremendous impact on the capacity of viruses to adapt to their environments, emphasizing the central importance of a hitherto under-appreciated element of the host–pathogen interaction and potentially providing new targets for antiviral intervention.

2.3   Results

2.3.1   Quantifying nucleoprotein variant fitness in distinctive host cell folding environments

We used a deep mutational scanning strategy to systematically and experimentally quantify the fitness of nearly all viable single amino acid substitutions in influenza nucleoprotein in both basal and biophysically challenging host cell environments. To this end, we employed previously reported duplicate nucleoprotein mutant libraries based on the human-adapted A/Aichi/2/1968 (H3N2) influenza strain [24], and competed each viral library in Madin Darby canine kidney (MDCK) cells. The MxA orthologs in MDCK cells are inactive against all influenza strains tested to date [25], and permit robust influenza growth. To create a chaperone-depleted host cell environment for viral propagation, we employed a highly specific, chemically inducible dominant negative form of HSF1 [23], which is the master regulator of cytosolic and nuclear chaperone levels [26]. At both permissive (37 ⁰C) and biophysically restrictive, fever-like (39 ⁰C) temperatures, chemical induction of the HSF1 inhibitor reduces cytosolic chaperone transcript and protein levels, in both the absence (**Figures 2.1** and **2.2A**) and presence (**Figure 2.2A–B**) of influenza. Notably, inhibition of HSF1 (termed the HSF1i condition henceforth) did not significantly alter the replication of wild-type influenza or host cell metabolic fitness over the course of our experiment (**Figure 2.2C–D**). We anticipated that the fever-like temperature would prove moderately more challenging for nucleoprotein folding, potentially restricting the accessible mutational landscape, while the depletion of cytosolic chaperones at each temperature would assess the potential function of host chaperones in regulating nucleoprotein variant fitness.

To maintain library diversity (approximately 10,000 single amino acid substitutions) whilst minimizing co-infection, we infected approximately 10^7 cells in each host environment (**Figure 2.3A**) with about 10^6 infectious virions from our biological duplicate viral nucleoprotein libraries. In addition to performing deep mutational scanning in biological duplicate, we also performed technical duplicates with one of the replicate libraries (**Figure 2.4A**). Following a 48 h infection, we utilized a previously reported barcoded sub-amplicon sequencing strategy [27] to accurately quantify the abundance of nucleoprotein variants after replication at both 37 ⁰C and 39 ⁰C with and without host chaperone depletion (**Figures 2.4B–D**). The change in variant frequency upon selection was then normalized to that of the wild-type residue, such that the resulting differential selection [28] value provides a quantitative measure of the relative fitness of

56

each nucleoprotein variant in the conditions tested. The results of this analysis can be visualized on sequence logo plots (**Figures 2.5–2.8**).
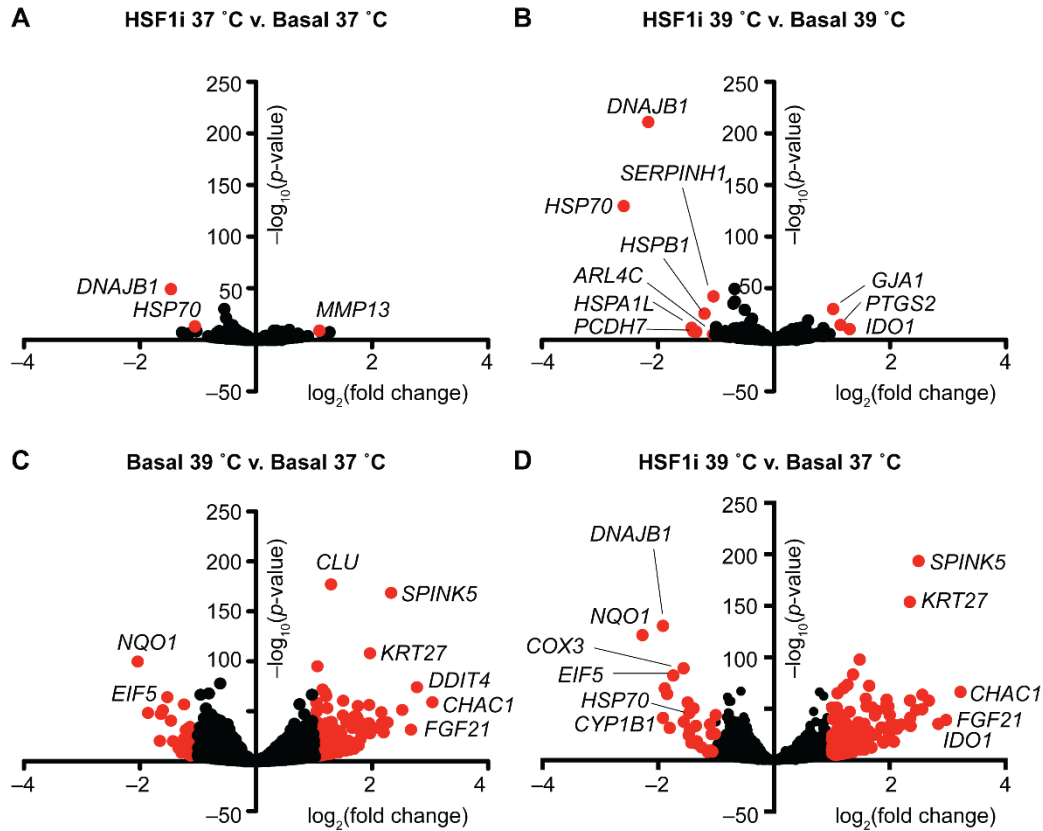
Figure 2.1. Transcriptional profiles of modulated host environments.
(A) Volcano plot of RNA-seq data for HSF1 inhibition at 37 °C. **(B)** Volcano plot of RNA-seq data for HSF1 inhibition at 39 °C. **(C)** Volcano plot of RNA-seq data for 39 °C relative to 37 °C in basal environment. **(D)** Volcano plot of RNA-seq data for HSF1-inhibited environment at 39 °C relative to a basal environment at 37 °C. For A–D, transcripts with > 2-fold change and p-values < 10^{-5} are shown in red, with outliers labeled.

A
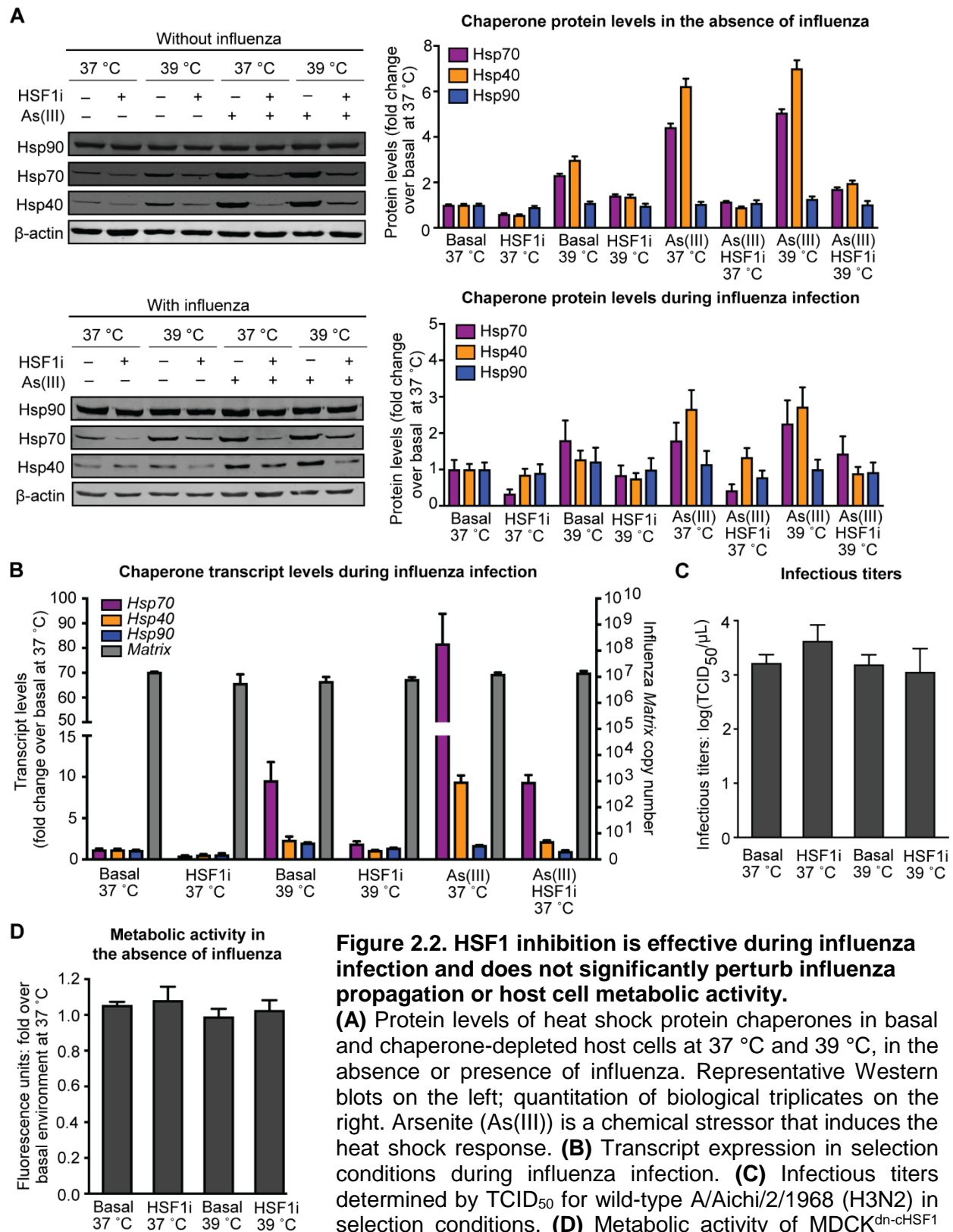
Figure 2.2. HSF1 inhibition is effective during influenza infection and does not significantly perturb influenza propagation or host cell metabolic activity.
(A) Protein levels of heat shock protein chaperones in basal and chaperone-depleted host cells at 37 °C and 39 °C, in the absence or presence of influenza. Representative Western blots on the left; quantitation of biological triplicates on the right. Arsenite (As(III)) is a chemical stressor that induces the heat shock response. **(B)** Transcript expression in selection conditions during influenza infection. **(C)** Infectious titers determined by $TCID_{50}$ for wild-type A/Aichi/2/1968 (H3N2) in selection conditions. **(D)** Metabolic activity of MDCK[dn-cHSF1] cells in each selection condition, as characterized by resazurin assay at 48 hours post-treatment.
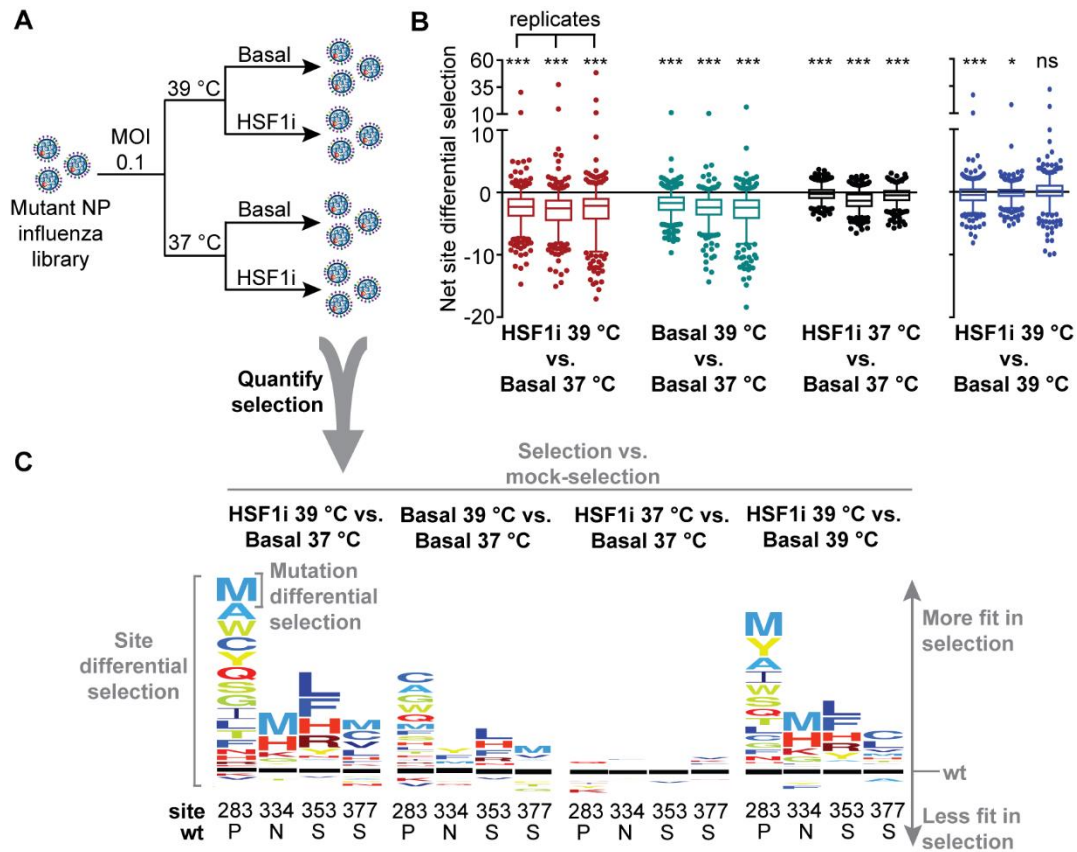
Figure 2.3. Deep mutational scanning reveals positively selected sites upon chaperone depletion at a biophysically restrictive temperature.
(A) Deep mutational scanning selection scheme: each mutant nucleoprotein (NP) influenza library was subjected to selection at 37 °C and 39 °C in both basal and chaperone-depleted (HSF1i) cells at a multiplicity of infection (MOI) of 0.1 virions/cell. Libraries were deep sequenced following replication to quantify the selection on each mutant in the library. Mutation differential selection corresponds to the differential selection of a specific NP mutant, relative to wild-type NP. Site differential selection corresponds to the sum of the mutation differential selection values for all variants at a given NP site. **(B)** Box plots showing the distributions of net site differential selection in both basal and chaperone-depleted (HSF1i) cells at 37 °C and 39 °C. Each selection was performed in biological triplicate; box plots are shown for each replicate. The whiskers show 5th and 95th percentiles, boxes show 25th and 75th percentiles, line shows median value, and points show outliers. Significance of deviation of the mean from zero (no selection) was evaluated by a one-sample *t*-test; * and *** indicate false discovery rates (FDRs) < 0.05 and < 0.0001, respectively. Nested ANOVA accounting for replicates and treatment conditions was performed for all selection conditions normalized to Basal 37 °C ($p = 0.0031$). For pairwise statistical comparisons, *p*-values are provided in the statistics section of the Methods. **(C)** Representative sequence logo plots showing amino acid variants at select sites differentially selected upon relative to wild-type. Size of the amino acid letters corresponds to the magnitude of the mutational differential selection, which is on the same scale for each selection. Amino acids above the black line are more fit in the selection condition compared to the mock-selection condition, amino acids below are less fit, and the black line represents behavior of the wild-type amino acid in the selection condition.
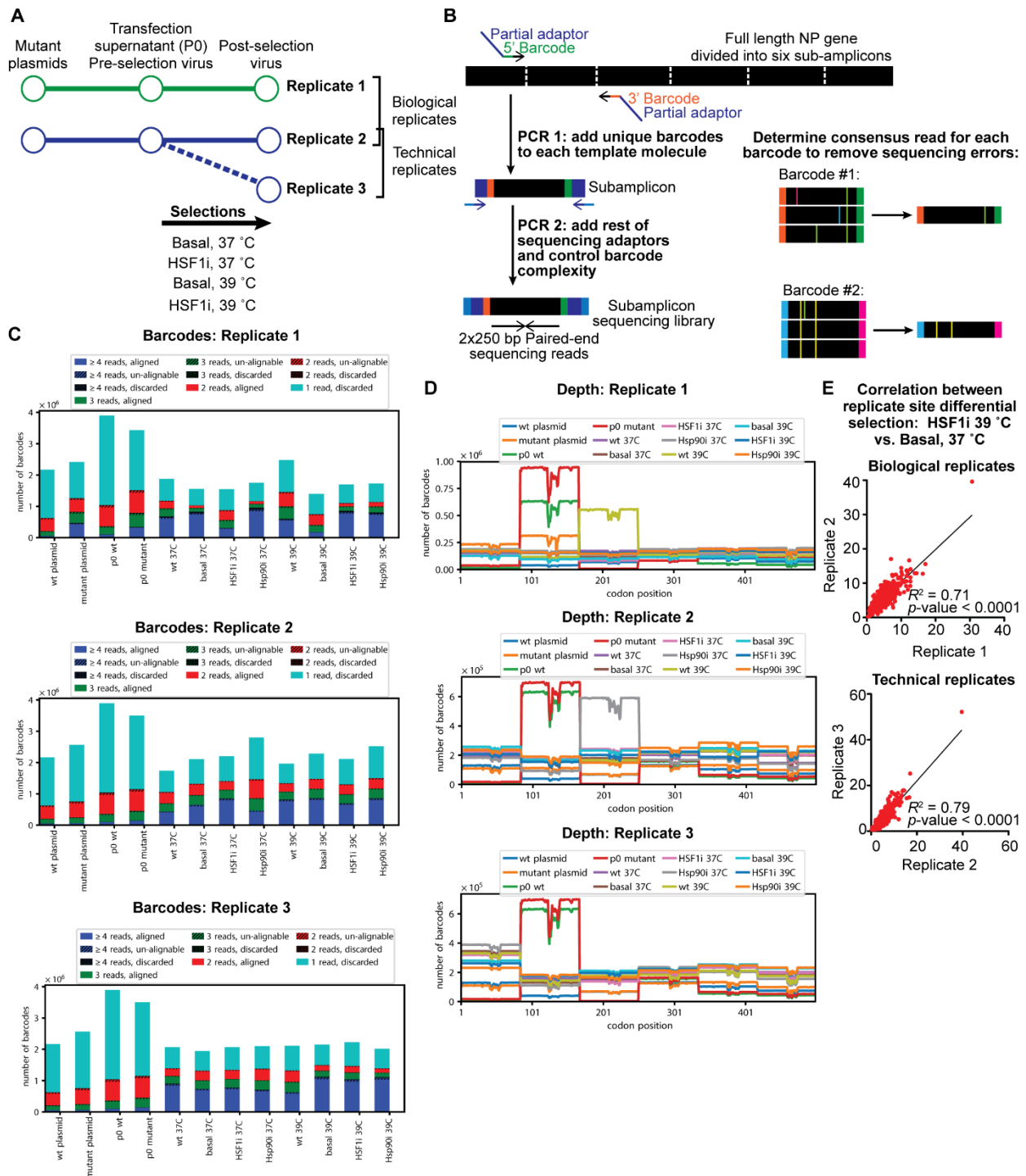
Figure 2.4. Sub-amplicon sequencing enables quantification of variant frequency.
(A) Schematic of replicate structure. **(B)** Sub-amplicon sequencing strategy workflow[27]. **(C)** Number of reads per barcode. **(D)** Number of barcodes per sub-amplicon. **(E)** Correlation plots of the absolute site differential selection[28] between biological and technical replicates for HSF1 inhibition at 39 °C compared to a basal environment at 37 °C. Best-fit line is plotted, with correlation coefficient and the *p*-value for significance of the slope deviating from zero shown on each plot.

60
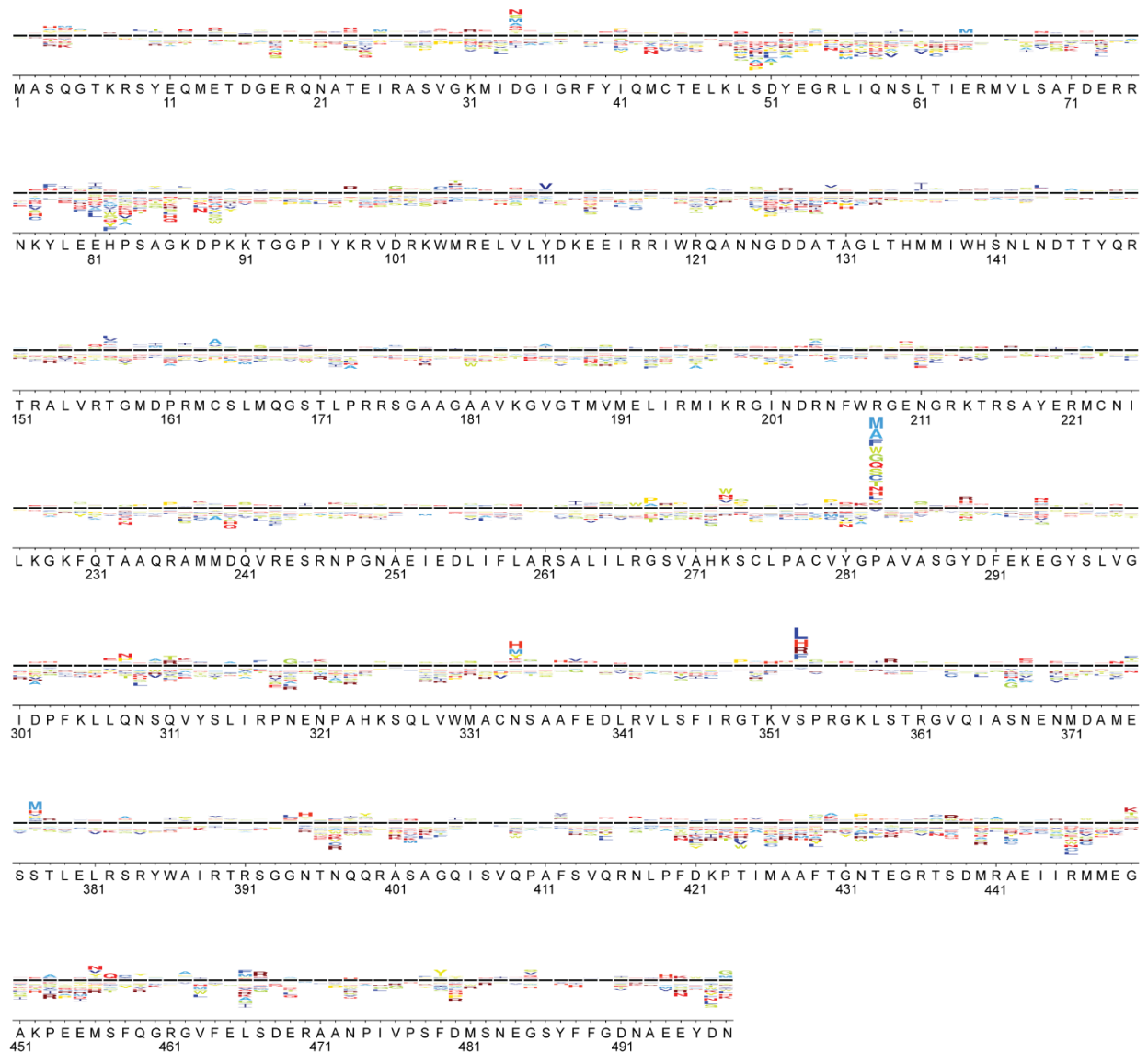
Figure 2.5. Representative full sequence logo plot for nucleoprotein: HSF1-inhibited environment at 39 °C relative to a basal environment at 37 °C.
Wild-type influenza A/Aichi/2/1968 nucleoprotein sequence and residue numbers are shown below logo plot. Size of the amino acid letters corresponds to the magnitude of the mutational differential selection, which is on the same scale for Figures 2.5–8 and Figure 2.15. Amino acids above the black line are more fit in the HSF1-inhibited environment at 39 °C compared to a basal environment at 37 °C, amino acids below are less fit, and the black line represents behavior of the wild-type amino acid in the selection condition.

MASQGTKRSYEQMETDGERQNATEIRASVGKMIDGIGRFYIQMCTELKLSDYEGRLIQNSLTIERMVLSAFDERR
1   11   21   31   41   51   61   71

NKYLEEHPSAGKDPKKTGGPIYKRVDRKWMRELVLYDKEEIRRIWRQANNGDDATAGLTHMMIWHSNLNDTTYQR
        81   91   101   111   121   131   141

TRALVRTGMDPRMCSLMQGSTLPRRSGAAGAAVKGVGTMVMELIRMIKRGINDRNFWRGENGRKTRSAYERMCNI
151   161   171   181   191   201   211   221

LKGKFQTAAQRAMMDQVRESRNPGNAEIEDLIFLARSALILRGSVAHKSCLPACVYGPAVASGYDFEKEGYSLVG
        231   241   251   261   271   281   291

IDPFKLLQNSQVYSLIRPNENPAHKSQLVWMACNSAAFEDLRVLSFIRGTKVSPRGKLSTRGVQIASNENMDAME
301   311   321   331   341   351   361   371

SSTLELRSRYWAIRTRSGGNTNQQRASAGQISVQPAFSVQRNLPFDKPTIMAAFTGNTEGRTSDMRAEIIRMMEG
        381   391   401   411   421   431   441

AKPEEMSFQGRGVFELSDERAANPIVPSFDMSNEGSYFFGDNAEEYDN
451   461   471   481   491

Figure 2.6. Representative full sequence logo plot for nucleoprotein: HSF1-inhibited environment at 39 °C relative to a basal environment at 39 °C.
Wild-type influenza A/Aichi/2/1968 nucleoprotein sequence and residue numbers are shown below logo plot. Size of the amino acid letters corresponds to the magnitude of the mutational differential selection, which is on the same scale for Figures 2.5–8 and Figure 2.15. Amino acids above the black line are more fit in an HSF1-inhibited environment at 39 °C compared to a basal environment at 39 °C, amino acids below are less fit, and the black line represents behavior of the wild-type amino acid in the selection condition.
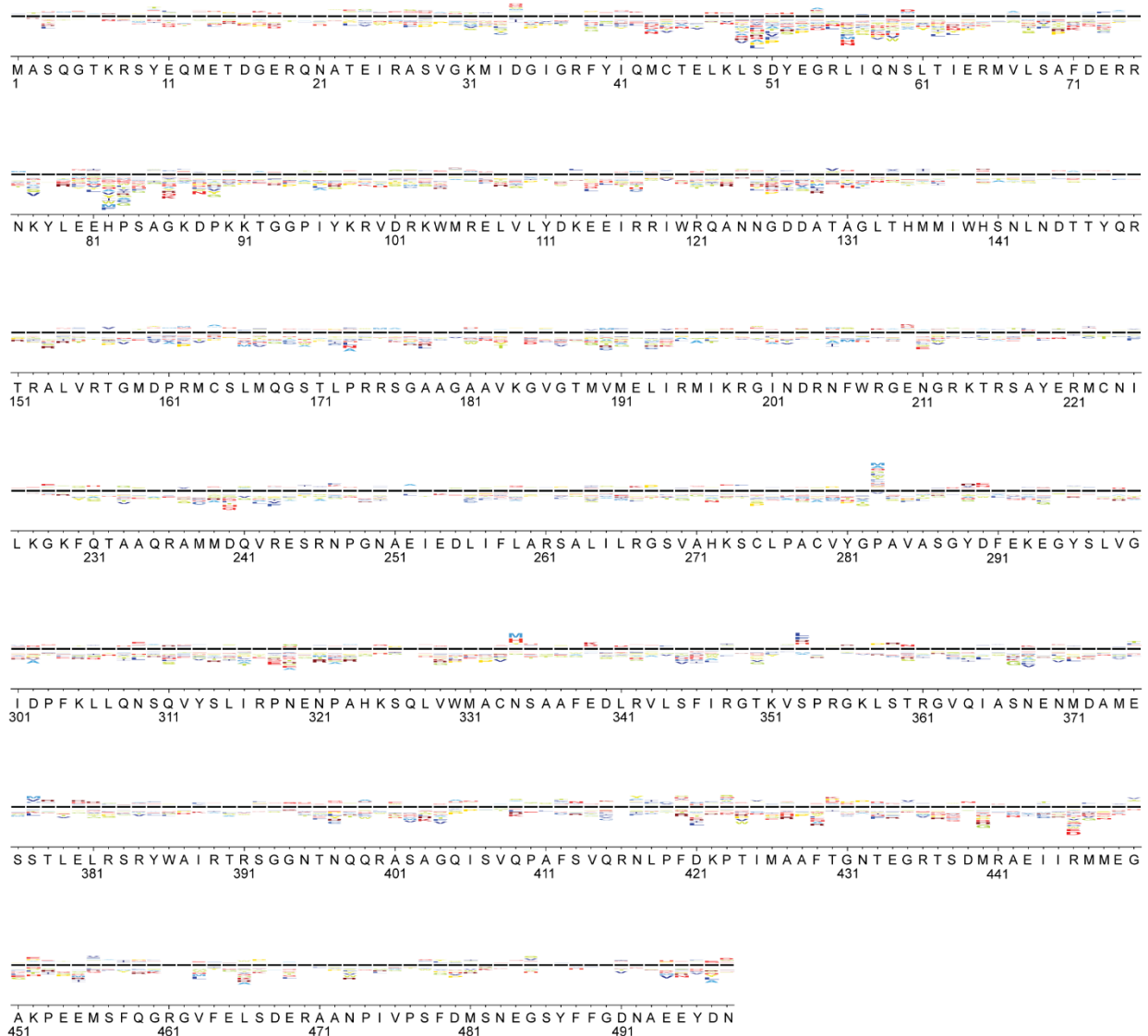
Figure 2.7. Representative full sequence logo plot for nucleoprotein: 39 °C relative to 37 °C in basal environment.

Wild-type influenza A/Aichi/2/1968 nucleoprotein sequence and residue numbers are shown below logo plot. Size of the amino acid letters corresponds to the magnitude of the mutational differential selection, which is on the same scale for Figures 2.5–8 and Figure 2.15. Amino acids above the black line are more fit at 39 °C compared to 37 °C, amino acids below are less fit, and the black line represents behavior of the wild-type amino acid in the selection condition.

63

M A S Q G T K R S Y E Q M E T D G E R Q N A T E I R A S V G K M I D G I G R F Y I Q M C T E L K L S D Y E G R L I Q N S L T I E R M V L S A F D E R R
1   11   21   31   41   51   61   71

N K Y L E E H P S A G K D P K K T G G P I Y K R V D R K W M R E L V L Y D K E E I R R I W R Q A N N G D D A T A G L T H M M I W H S N L N D T T Y Q R
        81   91   101   111   121   131   141

T R A L V R T G M D P R M C S L M Q G S T L P R R S G A A G A A V K G V G T M V M E L I R M I K R G I N D R N F W R G E N G R K T R S A Y E R M C N I
151   161   171   181   191   201   211   221

L K G K F Q T A A Q R A M M D Q V R E S R N P G N A E I E D L I F L A R S A L I L R G S V A H K S C L P A C V Y G P A V A S G Y D F E K E G Y S L V G
        231   241   251   261   271   281   291

I D P F K L L Q N S Q V Y S L I R P N E N P A H K S Q L V W M A C N S A A F E D L R V L S F I R G T K V S P R G K L S T R G V Q I A S N E N M D A M E
301   311   321   331   341   351   361   371

S S T L E L R S R Y W A I R T R S G G N T N Q Q R A S A G Q I S V Q P A F S V Q R N L P F D K P T I M A A F T G N T E G R T S D M R A E I I R M M E G
        381   391   401   411   421   431   441

A K P E E M S F Q G R G V F E L S D E R A A N P I V P S F D M S N E G S Y F F G D N A E E Y D N
451   461   471   481   491

Figure 2.8. Representative full sequence logo plot for nucleoprotein: HSF1-inhibited environment at 37 °C relative to a basal environment at 37 °C.
Wild-type influenza A/Aichi/2/1968 nucleoprotein sequence and residue numbers are shown below logo plot. Size of the amino acid letters corresponds to the magnitude of the mutational differential selection, which is on the same scale for Figures 2.5–8 and Figure 2.15. Amino acids above the black line are more fit in an HSF1-inhibited environment at 37 °C compared to a basal environment at 37 °C, amino acids below are less fit, and the black line represents behavior of the wild-type amino acid in the selection condition.
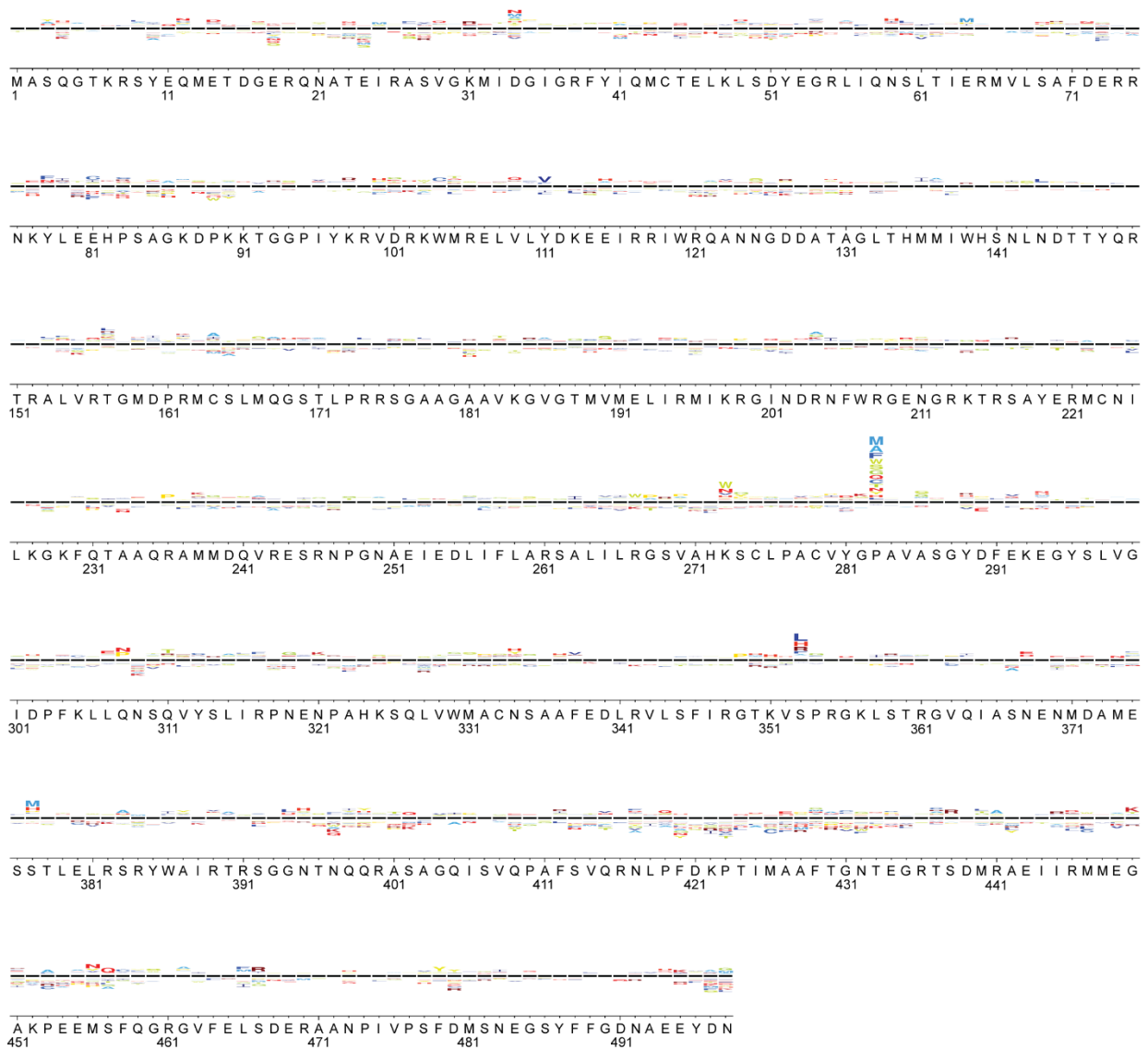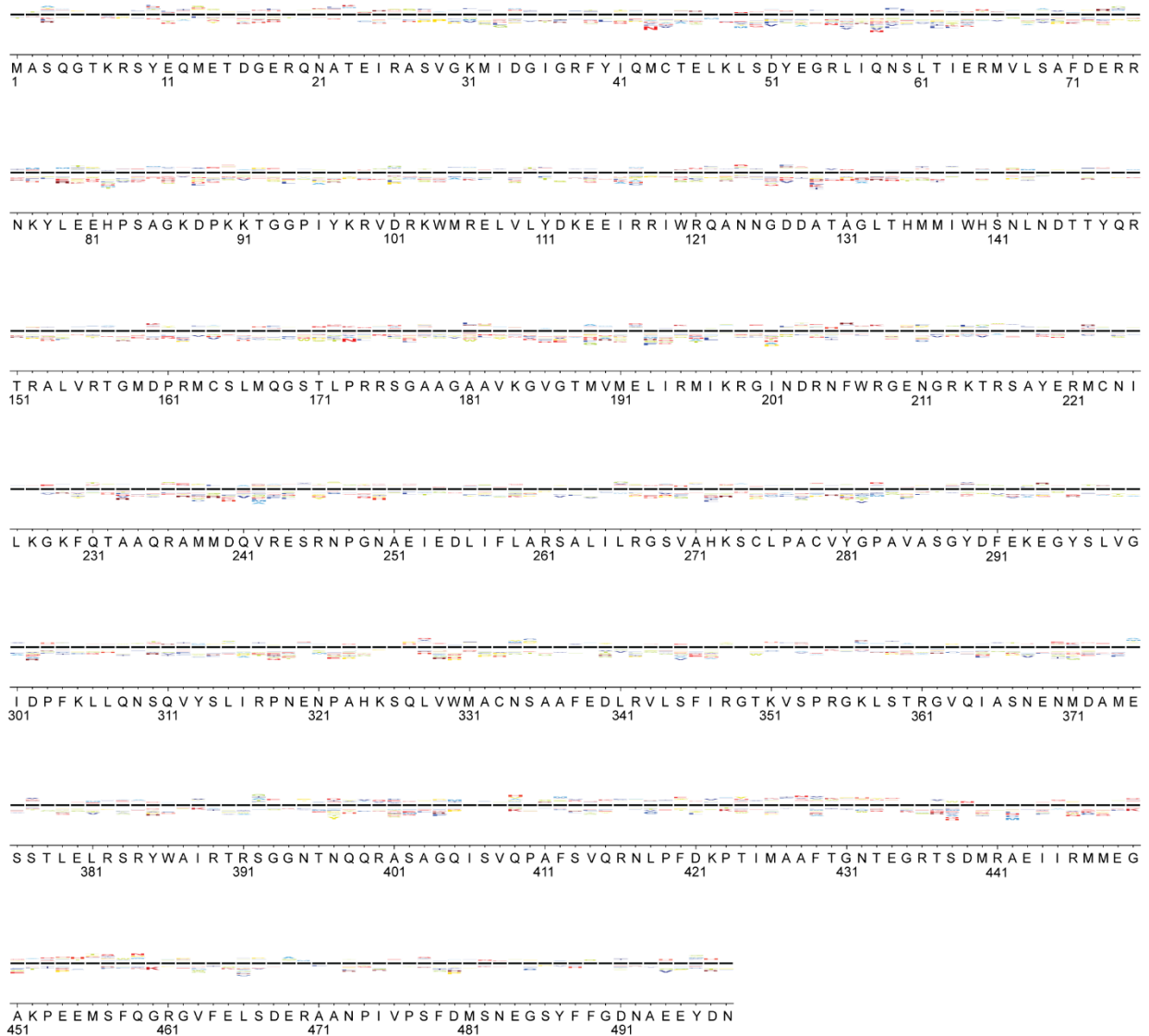
As expected, we found that the fever-like temperature is restrictive, generally reducing the fitness of nucleoprotein variants relative to the wild-type sequence (see net site differential selection plots in **Figure 2.3B** and also the logo plot in **Figure 2.7** — Basal 39 ºC vs. Basal 37 ºC). Host chaperone depletion caused by inhibition of HSF1 modestly reduces nucleoprotein variant fitness on average (**Figures 2.3B, 2.6** and **2.8** — HSF1i 39 ºC vs. Basal 39 ºC and HSF1i 37 ºC vs. Basal 37 ºC), whereas chaperone depletion at the restrictive, fever-like temperature substantially reduces the fitness of nucleoprotein variants (**Figures 2.3B** and 2.**5** — HSF1i 39 ºC vs. Basal 37 ºC; statistical analyses in Methods). Most strikingly, chaperone depletion at elevated temperature results in very high levels of differential selection at several specific sites in nucleoprotein, including sites 283, 334, 353, and 377 (**Figures 2.3C** and **2.5** — HSF1i 39 ºC vs. Basal 37 ºC). At each of these sites, multiple amino acids confer strongly enhanced fitness relative to the wild-type residue. This phenotype is specifically revealed upon host chaperone depletion at 39 ºC, as we do not observe significant positive differential selection at these sites upon depleting chaperones at 37 ºC (**Figures 2.3C** and **2.8** — HSF1i 37 ºC vs. Basal 37 ºC) and only modest positive differential selection upon increasing the temperature in an environment with basal levels of chaperones (**Figures 2.3C** and **2.7** — Basal 39 ºC vs. Basal 37 ºC).

Competitions between thousands of variants are often inherently noisy, largely owing to differences in variant composition between replicate libraries [27]. However, we observed strong correlation for selection on nucleoprotein sites between biological ($R^2 = 0.71$) and technical ($R^2 = 0.79$) replicates of our deep mutational scanning experiments (**Figure 2.4E**). Thus, the selection strength imparted by host chaperone depletion at a restrictive, fever-like temperature substantially exceeds the experimental noise. Moreover, these deep mutational scanning results were recapitulated in all pairwise competitions we performed between individual highly selected variants and virus carrying the wild-type nucleoprotein (**Figure 2.9**), and no significant fitness change was observed for a synonymous variant used as a control (**Figure 2.9**, false discovery rate [FDR] = 0.4). Cumulatively, these findings confirm the validity of the deep mutational scanning data and indicate that depleting chaperones at a moderately elevated temperature creates a stringent host environment that strongly selects against wild-type residues in nucleoprotein at certain sequence positions.
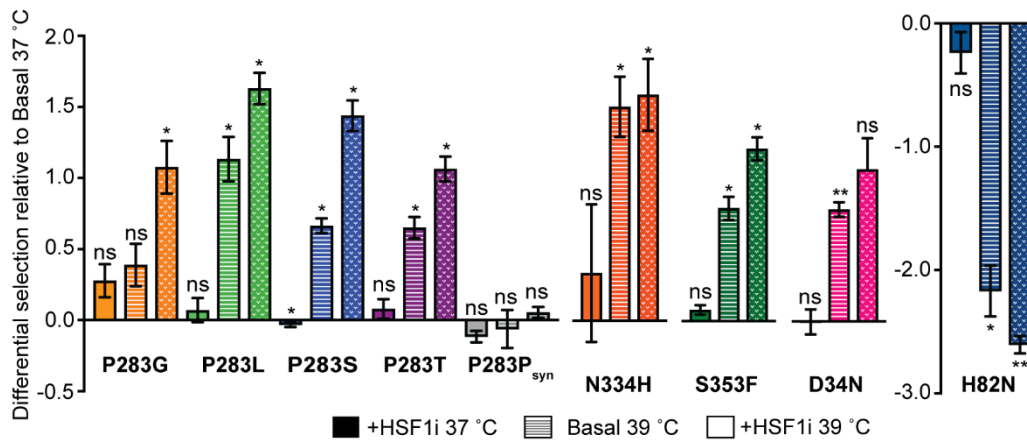
Figure 2.9. Pairwise competitions recapitulate deep mutational scanning batch competition.
Error bars indicate standard error from biological replicates ($N = 3$). Significance of deviation from zero (no selection) was evaluated by a one-sample t-test followed by Benjamini-Hochberg adjustment for multiple-comparison; * and ** indicate FDR < 0.05 and < 0.01, respectively.

2.3.2 Biophysical costs of individual nucleoprotein variants

The strong differential selection we observed in chaperone-depleted host cells at a restrictive temperature suggested to us that specific wild-type nucleoprotein residues in the Aichi influenza strain, the strain our libraries are based on, may entail a biophysical cost that is nonetheless tolerated under permissive conditions, perhaps owing to competing selection forces. This hypothesis is particularly compelling for the previously characterized stabilizing nucleoprotein variants [10] N334H (**Figures 2.3C** and 2.**5**) and M136I (**Figure 2.5**), which were positively selected in biophysically challenging conditions. Interestingly, these variants exhibited modestly enhanced fitness upon increased temperature and significantly enhanced fitness upon chaperone depletion at increased temperature. The unmasking of the deleterious fitness effects of the wild-type sequence at these positions upon host chaperone depletion supports the hypothesis that chaperones can indeed rescue biophysically deleterious NP variants.

We observed a similar and even more striking trend at site 283, where numerous amino acid substitutions, including serine, leucine, threonine, and glycine, were validated in our pairwise competition experiments to be significantly more fit than the wild-type proline residue when host chaperones were depleted at a restrictive temperature (**Figures 2.3C** and **2.9B** — HSF1i 39 ºC vs. Basal 37 ºC). Intriguingly, the identity of the amino acid at nucleoprotein site 283 is known to critically modulate influenza sensitivity to the human antiviral restriction factor

66

MxA, with proline at that position contributing greatly to MxA escape [9, 16]. Although a structure of the MxA:nucleoprotein complex is not currently available, site 283 is likely located at the MxA–nucleoprotein interface [29]. Moreover, Pro283 is nearly universally conserved in human influenza nucleoprotein, but rarely occurs in avian influenza nucleoprotein. This characteristic difference between human and avian influenza strains is attributed to the necessity of Pro283 to escape human MxA, whereas the avian MxA ortholog lacks known antiviral activity [30]. Indeed, the Leu283Pro substitution enabled the 1918 pandemic influenza strain to escape MxA following zoonotic transmission [16]. Pro283 in nucleoprotein has thus greatly impacted the fitness of modern human influenza strains.

The observation that an adaptive amino acid substitution as important as Pro283 can be rendered unfit by biophysically challenging, host chaperone-depleted conditions motivated us to elucidate the underlying molecular basis of this fitness effect. Our hypothesis was that the depletion of host chaperones exacerbates a biophysical defect in nucleoprotein folding that is caused by installation of a proline at position 283. Prior work has shown that nucleoprotein is engaged by numerous cytosolic chaperones, including chaperones like Hsp40 and Hsp70 that are depleted in our HSF1-inhibited host cell environment [31-35] (**Figures 2.1** and **2.2**). For example, Hsp40 (*DNAJB1*) directly interacts with nucleoprotein and facilitates its nuclear import [34], while Hsp70 is implicated in modulating nucleoprotein nuclear export [35]. The heat shock proteins can also regulate antiviral responses indirectly through their interactions with nucleoprotein [31-33]. Thus, the disruption of critical nucleoprotein–host chaperone interactions by HSF1 inhibition-mediated chaperone depletion may indeed be the source of significant differential selection at site 283, especially at a biophysically restrictive temperature.

This possibility raises the question of whether the Pro283 nucleoprotein variant is, in fact, biophysically defective relative to other variants at site 283. Although there is currently no high-resolution structure of a nucleoprotein variant containing Pro283, crystal structures of avian influenza nucleoprotein variants with serine or leucine at position 283 are available [36-37]. Strikingly, in these structures site 283 is in the middle of an α-helix (**Figure 2.10A**). Proline is classically regarded as a "helix-breaker," owing in part to its inability to form an $i + 4$ hydrogen bond important for α-helix stability [38]. Therefore, it seemed reasonable to anticipate that the replacement of serine (or leucine) with a proline at position 283, as is observed in human influenza strains, would indeed be biophysically problematic.

To assess this hypothesis, we first performed molecular dynamics simulations in explicit water to evaluate whether Pro283 affects either the overall structure of nucleoprotein or the

structure of the α-helix centered at position 283. Although the overall nucleoprotein structure was not grossly perturbed by Pro283, as would be expected given that this variant is still capable of supporting influenza replication, our simulations revealed that a Pro283 nucleoprotein variant is significantly less α-helical at residues 282–284 than is a Ser283-containing variant (**Figure 2.10B**).
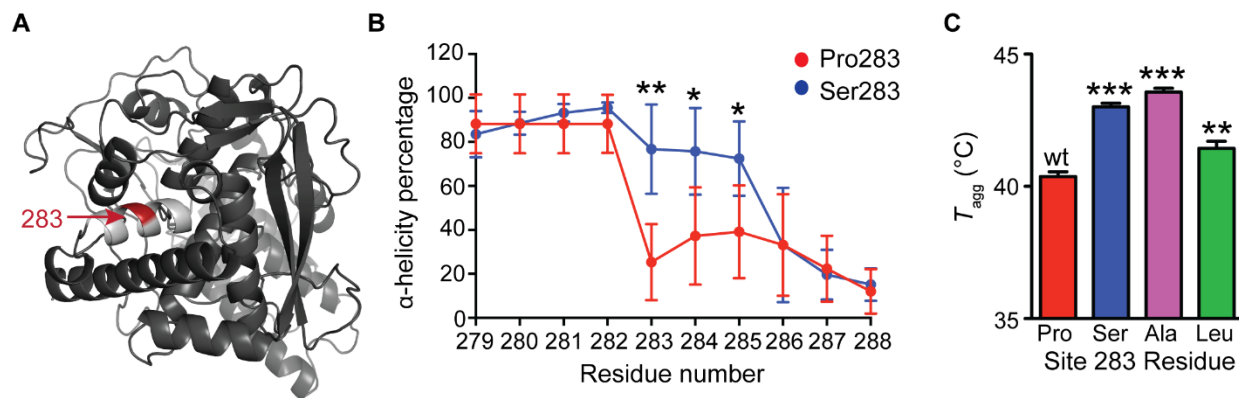
Figure 2.10. Pro283 disrupts nucleoprotein α-helical content and is destabilized relative to other variants at site 283.
(A) Nucleoprotein crystal structure reveals site 283 is in the center of an α-helix (PDB ID: 2IQH[36]).
(B) Molecular dynamics simulations show that Pro283 disrupts nucleoprotein α-helical content compared to Ser283. A Student's t-test indicates significant differences at sites 282–284. **(C)** Apparent (irreversible) melting temperatures (T_{agg}) of wild-type and site 283 nucleoprotein variants evaluated by circular dichroism. Student's t-test indicates significance of deviation from wild-type apparent melting temperature.

The significant structural consequences of Pro283 observed in these simulations prompted us to experimentally investigate whether Pro283 affects nucleoprotein stability. Following a previously reported protocol [10], we recombinantly expressed and purified a monomeric form of nucleoprotein with either proline, serine, leucine, or alanine at position 283 (**Figure 2.11**). Circular dichroism spectra indicated that all four variants had grossly similar secondary structures (**Figure 3.21C**), consistent with our simulations. Thermal denaturation of nucleoprotein is irreversible, leading to rapid aggregation and precipitation of the protein. Nonetheless, fitting these irreversible thermal melts to a two-state model revealed that Pro283 nucleoprotein does indeed precipitate at a significantly lower temperature than all three of the non-Pro variants studied (**Figures 2.10C** and **2.11C**), and is therefore less stable and more aggregation-prone.

The observation that Pro283 nucleoprotein has a higher propensity to aggregate than other variants is consistent with either a kinetic or thermodynamic defect caused by substitution with Pro283. Given that Gly is, like Pro, known as an α-helix breaker and that Gly283 is

positively selected relative to Pro283 (see **Figure 2.3C**– HSF1i 39 ⁰C vs. Basal 37 ⁰C) upon chaperone depletion at febrile temperatures, we favor a substantial contribution from a kinetic defect that may be associated with the propensity of Pro to form both cis and trans amide bonds[38]. The substantive biophysical defect endowed by Pro283 on nucleoprotein likely explains the enhanced dependence of this variant on host chaperones, and explains why other variants are significantly more fit in the absence of those key chaperones. Moreover, these results may help to explain why Pro283 is not observed in avian influenza strains [16]. Birds typically have body temperatures ranging from 39–43 ⁰C [39], the upper end of which may be too extreme to permit chaperone-mediated rescue of the biophysically defective Pro283 nucleoprotein variant.
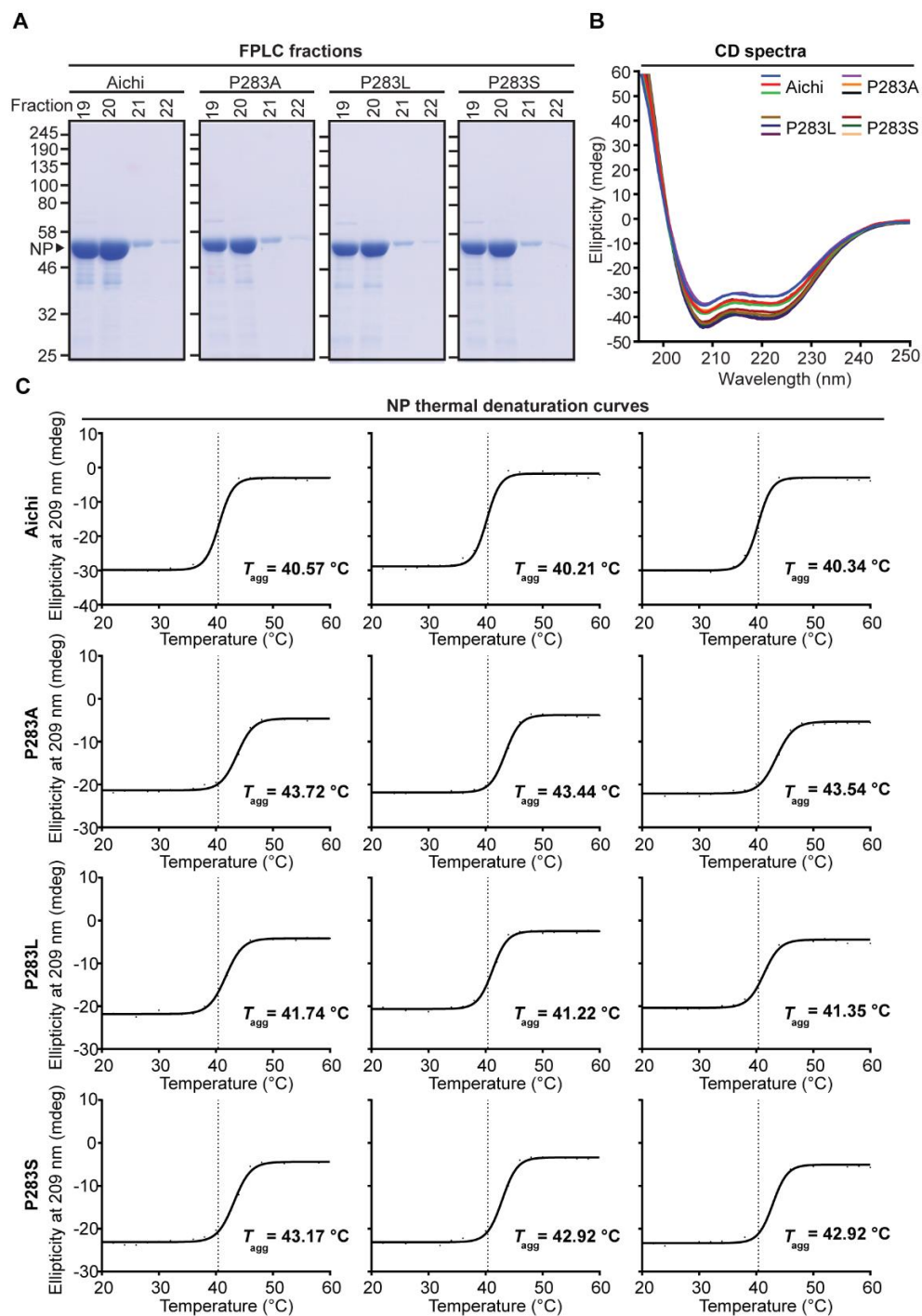
Figure 2.11. Purification and thermal denaturation of recombinant nucleoprotein variants.
(A) Purified nucleoprotein (56 kDa) by Ni-NTA column chromatography and size exclusion chromatography. **(B)** Circular dichroism wavelength scans at 20 °C for nucleoprotein variants. All variants exhibited similar circular dichroism spectra characteristic of an α-helical protein. Scans for each variant were performed in triplicate. **(C)** Thermal denaturation curves for nucleoprotein. Apparent (irreversible) melting temperatures (T_{agg}) were obtained from sigmoidal fits over 20–60 °C.

2.3.3 Evaluating competing selection pressures of MxA escape and host chaperone depletion

Cumulatively, these results suggest that nucleoprotein variants critical for innate immune system escape, most especially Pro283, can be folding-defective and display compromised fitness in biophysically challenging host environments (**Figure 2.12A**, left). This finding motivated us to evaluate whether Pro283 fitness remains compromised under biophysically challenging conditions even upon the addition of an MxA selection pressure that normally selects strongly in favor of the Pro283 variant (**Figure 2.12A**, right). To this end, we performed pairwise viral competitions between the biophysically stable Ser283 variant and the MxA-resistant Pro283 variant in each of our host environments in the presence of either active or inactive MxA (**Figures 2.12B** and **2.13**). In permissive folding environments (Basal and HSF1i at
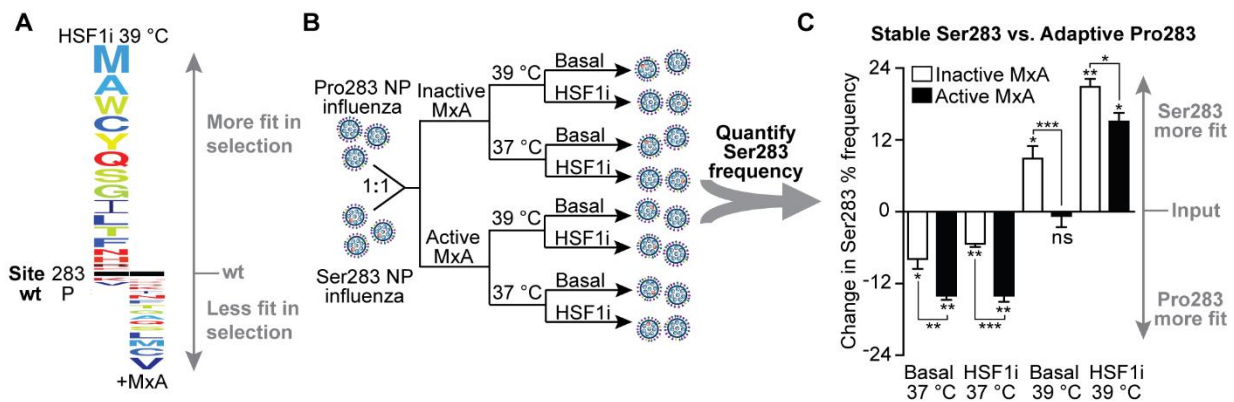
Figure 2.12. Host chaperones define the immune escape capacity of the Pro283 nucleoprotein variant.

(A) Deep mutational scanning reveals opposing selection forces from chaperone depletion versus MxA-mediated immune selection. Left: chaperone-depletion at 39 °C shown relative to basal environment at 37 °C, as in Figure 3.3C. Right: MxA selection from previously published deep mutational scanning data[9]. Size of the amino acid letters corresponds to the magnitude of the mutational differential selection, which is on the same scale for each selection. Amino acids above the black line are more fit than wild-type Pro283 in the selection condition, and amino acids below are less fit. **(B)** Pairwise viral competition scheme: wild-type nucleoprotein (NP) influenza was mixed with mutant NP influenza in a 1:1 ratio. The mixture was then subjected to selection at 37 °C and 39 °C in both basal and chaperone-depleted (HSF1i) cells at an MOI of 0.1 virions/cell, in the presence of either active or inactive MxA. **(C)** Change in mutant frequency upon selection is plotted on the y-axis, with the origin representing the mutant frequency in the input (inoculum). Error bars indicate standard error from biological replicates ($N = 3$). Significance of deviation from input (no selection) was evaluated for each selection condition by a one-sample t-test; * and ** above individual bars indicate FDR < 0.05 and < 0.01, respectively. Nested ANOVA accounting for replicates and treatment conditions was performed for all selections ($p = 1.16 \times 10^{-9}$). The significance of the difference in mutant frequency in the presence of active versus inactive MxA was evaluated for each selection condition by post-hoc pairwise comparisons; *, **, and *** between the bars indicate p-values < 0.05, < 0.01, and < 0.001, respectively.

37 ºC), Pro283 was enriched compared to Ser283, thereby enabling MxA escape (**Figure 2.12C**). In contrast, in biophysically challenging environments (Basal and HSF1i at 39 ºC), the stability defects of Pro283 were exacerbated, and Ser283 was enriched compared to Pro283. Strikingly, Ser283 was enriched even in the presence of MxA selection pressure when chaperones are depleted at 39 ºC, thereby hindering immune escape.

Altogether, these data reveal that HSF1-regulated chaperones can define the fitness of biophysically destabilized immune escape variants in influenza. This observation suggests a model in which enhanced fitness conferred by immune escape is often counterbalanced over the course of influenza evolution by biophysical defects that have a substantive fitness cost (**Figure 2.14, left**). Remarkably, at least in the case of influenza nucleoprotein, our data show that the virus is able to hijack host chaperones to resolve these biophysical defects (**Figure 2.14, right**). By this mechanism, the virus manages to uncouple protein folding fitness costs from the advantageous consequences of immune escape, expanding the accessible mutational landscape to access essential viral protein variants capable of both folding and immune escape.

Protein levels of FLAG-MxA and FLAG-MxA (T103A)
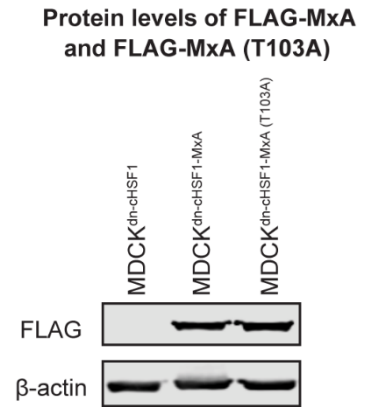
Figure 2.13. Protein levels of FLAG-MxA and FLAG-MxA(T103A) in MDCKdn-cHSF1-MxA and MDCKdn-cHSF1-MxA(T103A) cells, respectively. MDCK^dn-cHSF1 is shown as a negative control. Representative blot is shown (N = 2).
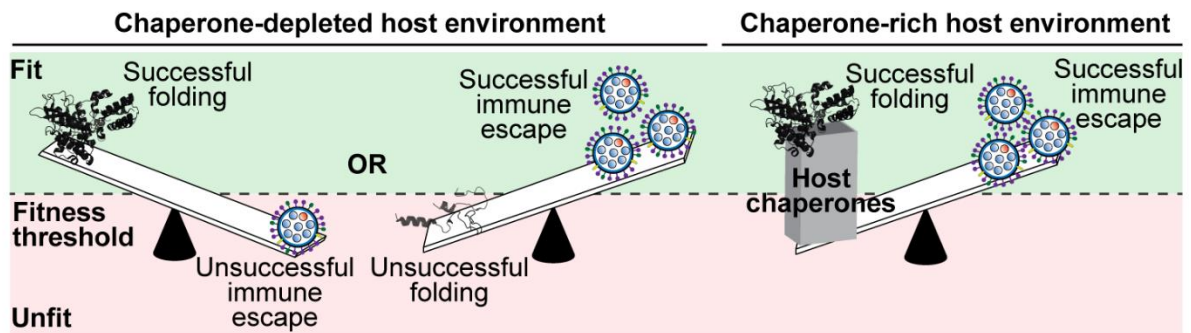
Figure 2.14. Host chaperones mediate the accessibility of biophysically destabilized adaptive mutations.
(Left) Model for a chaperone-depleted environment, where immune escape and acceptable protein biophysical properties present strongly competing selection pressures. **(Right)** Model for a chaperone-rich host environment in which, at least for some viral protein variants, these pressures can be uncoupled as chaperones enable viral proteins to maintain acceptable folding and structural properties while still endowing escape from the immune system.

2.4   Discussion

This work provides, to the best of our knowledge, the first direct experimental evidence that host chaperones mediate the accessibility of biophysically deleterious, adaptive viral protein variants. This feature of the host–pathogen interaction is apparent in multiple sites across the nucleoprotein gene. Particularly noteworthy, we find that the destabilized Pro283 nucleoprotein variant is not tolerated in a chaperone-depleted host environment at a restrictive temperature, as nucleoprotein is unable to engage host chaperones to address the Pro283-induced biophysical defect. Remarkably, we observe that even in the presence of selection pressure imposed by MxA that strongly favors Pro283 [9], the fitness of Pro283 nucleoprotein is still contingent on the host's chaperone levels and biophysical environment.

Based on these results, we expect that host chaperones can impact the accessibility of adaptive viral protein variants far beyond nucleoprotein and influenza, as amino acid substitutions are largely destabilizing [6] and many viral proteins are known to engage host chaperones [34, 40-41]. Moreover, previous work has revealed (1) that viral evolution is fundamentally constrained by protein stability [4, 10] and (2) the role of the Hsp90 chaperone in viral replication across numerous viral families [41-44]. Thus, the evolutionary trajectories of diverse viral proteins are likely to be influenced by numerous host chaperones [18-19].

Since our work suggests that host chaperones preferentially rescue biophysically defective viral protein variants, as more data accumulate in this field we may eventually be able to predict how chaperones will impact fitness in a rational manner based on protein variant biophysical properties. Further, our infections in the presence of the MxA restriction factor demonstrate that host chaperones can mediate the accessibility of escape variants irrespective of competing selection pressures. These data raise the possibility of antiviral therapeutic adjuvants targeting host chaperones that inhibit the development of antiviral resistance by constraining the accessible mutational landscape. We further observe that temperature critically influences the fitness of viral variants, with most variants suffering fitness costs at elevated temperatures that mimic fever conditions and/or the body temperatures of birds and small mammals [39]. Thus, fever and host-switching events may impose selection on viral variants that hampers adaptation. Based on our findings here, the nature of such selection is likely to be strongly influenced by host-specific differences in chaperone network compositions.

We anticipate that these phenomena extend far beyond the host–pathogen interface and apply to protein evolution more broadly. Previous work by Lindquist and others suggested that the Hsp90 chaperone can potentiate and buffer genetic variation in endogenous proteins [45-49].

Here, the impact of the host chaperone environment on nucleoprotein variant fitness is *not* driven predominantly by Hsp90 (**Figure 2.15**), even though NP does engage this chaperone [50], but instead by inhibition of HSF1 modulating the composition of a complex network of multiple protein folding and quality control factors. Moreover, our work shows experimentally that chaperones have the largest effect on the fitness of biophysically defective protein variants, a result that may help to explain extensive prior work with Hsp90 in which the potential biophysical mechanism of effects on protein evolution have not been experimentally evaluated.

HSF1-regulated host proteostasis network components may modulate nucleoprotein evolution directly, for example by impacting nucleoprotein–chaperone interactions, or indirectly, perhaps by perturbing levels of endogenous chaperone client proteins with antiviral properties. Our data support the former case, as we find that destabilized nucleoprotein variants are particularly sensitive to chaperone depletion. Nonetheless, we would not rule out possible contributions from secondary effects. Deciphering between primary and secondary effects will first require identifying the individual components of the intricate protein folding network that are primarily responsible for modulating nucleoprotein fitness, followed by systematic elimination of potential downstream effectors. Whether primary or secondary, the evolutionary implications of the protein folding network clearly extend well beyond Hsp90 and also play critical roles in evolution at the host–pathogen interface.

Finally, this work provides experimental evidence for the longstanding hypothesis that chaperones buffer the fitness cost of biophysically destabilized protein variants [45, 51-52]. Experimental validation of this concept in metazoan cells for the first time, to the best of our knowledge, has significant consequences for understanding the constraints on protein evolution, which have so far focused on inherent biophysical properties of proteins [2, 10]. For specific destabilized adaptive variants, fitness has been attributed to compensatory stabilizing mutations elsewhere in the protein structure [10, 53]. Cases of this idiosyncratic epistasis mediating pathogen adaptation have motivated efforts to determine the pervasiveness of compensatory mutations [10]. Our data reveal that the constraints on protein evolution are still more complex, establishing that protein variant fitness is constrained not just by inherent stability but also by the cellular environment in which the protein folds.

Figure 2.15. Representative full sequence logo plot for nucleoprotein: Hsp90-inhibited environment at 39 °C relative to basal environment at 39 °C.
Wild-type influenza A/Aichi/2/1968 nucleoprotein sequence and residue numbers are shown below logo plot. Deep mutational scanning was also performed in an Hsp90-inhibited environment to determine whether Hsp90 caused the fitness effects observed in the HSF1i environment, see discussion. Size of the amino acid letters corresponds to the magnitude of the mutational differential selection, which is on the same scale for Figures 2.5–3.8 and Figure 2.15. Amino acids above the black line are more fit upon Hsp90 inhibition at 39 °C compared to a basal environment at 39 °C, amino acids below are less fit, and the black line represents behavior of the wild-type amino acid in the selection condition.

2.5   Materials and Methods

Plasmids. The following plasmids were used to generate A/Aichi/2/1968 influenza virus: pHWAichi68-NP [54], pHWNan95-PB2 [54], pHWNan95-PB1 [54], pHWNan95-PA [54], pHW184-HA [55], pHW186-NA [55], pHW187-M [55], and pHW188-NS [55]. For nucleoprotein recombinant expression and biophysical studies, a pET28b(+) expression vector encoding monomeric A/Aichi/2/1968 nucleoprotein with an R416A mutation (which prevents RNA binding), deletion of residues 2–7, and a C-terminal 6×-His tag was used [10]. A lentiviral vector containing a FLAG-tagged human MxA or inactive MxA(T103A) sequence under a CMV promoter with a GSG linker (DYKDDDDKGSG) at the C-terminus was used for generation of the MxA-expressing MDCK cell line [9]. Downstream of MxA, the plasmid contained an internal ribosome entry site (IRES) followed by an mCherry reporter gene to assist the selection of stable single colony cell lines.

Antibodies. Antibodies used were as follows: rat monoclonal anti-FLAG (Agilent; 200474), mouse monoclonal anti-β-actin (Sigma; A1978), rabbit polyclonal anti-Hsp90, rat monoclonal anti-Hsp70, and rabbit monoclonal anti-Hsp40 were obtained from (Cell Signaling; 4877, 4873, and 4871, respectively). IRDyes 800CW goat anti-rat, 680LT goat anti-mouse, and 800CW goat anti-rabbit secondary antibodies were obtained from LI-COR (926-32219, 926-68020, and 926-32211, respectively).

Cell lines. MDCK$^{dn-cHSF1}$ cell line construction: MDCK cells were originally purchased from the American Type Culture Collection (Manassas, VA) and validated as MDCKs by STR profiling (Science Exchange). Cells were cultured at 37 °C in a 5% CO_2(g) atmosphere in DMEM (CellGro) supplemented with 10% fetal bovine serum (FBS; CellGro) and 1% penicillin/streptomycin/glutamine (CellGro). The parental MDCK cells were first transduced with lentivirus encoding a blasticidin-resistant tetracycline repressor, and then with lentivirus encoding a zeocin-resistant, tetracycline-inducible dn-cHSF1 construct [23]. Heterostable cells expressing the tetracycline repressor and the dn-cHSF1 construct were then selected using 8 µg/mL zeocin and 4 µg/mL blasticidin. Single colonies were generated by serial dilution in 96-well plates, expanded, and then selected based on functional testing of HSF1 inhibition [23] using qPCR (**Figure 2.2B**), as previously described.

MDCK$^{dn-cHSF1-MxA}$ and MDCK$^{dn-cHSF1-MxA(T103A)}$ cell line construction: MDCK$^{dn-cHSF1}$ cells were engineered to constitutively express human MxA or the inactive MxA-T103A mutant [9] by transducing with lentivirus encoding FLAG-tagged MxA variants. 72 h post-transduction, single

colonies were generated by serial dilution in 96-well plates. Wells with clonal transduced cells were identified as single clusters of cells expressing mCherry, expanded, and then characterized as described below.

Characterization of cellular environments for influenza competitions

Quantitative RT-PCR. MDCK$^{dn-cHSF1}$ cells were plated in 12-well plates at a density of 100,000 cells/well and pre-treated with 0.1% DMSO or doxycycline at a final concentration of 1 µg/mL upon seeding. After 18 h, the cells were infected with wild-type influenza A/Aichi/2/1968 at an MOI of 1 virion/cell. Infectious media was replaced with fresh WSN media (OptiMEM-I from Thermo Fisher Scientific supplemented with 0.5% heat-inactivated FBS, 0.3% BSA from Invitrogen, 1% of penicillin/streptomycin from Bio Whittaker, and 100 µg/mL of $CaCl_2$ from Sigma) supplemented with 0.1% DMSO or 1 µg/mL doxycycline 2 h post-infection. Infected cells were treated with 100 µM sodium arsenite (Alfa Aesar) for 90 min prior to harvesting for a heat shock activation control and harvested 8 h post-infection. Cellular RNA was extracted using Total RNA Kit I with Homogenizer columns (Omega). 1 µg RNA was reverse transcribed into cDNA using the Applied Biosystems High-Capacity Reverse Transcription kit. The reverse transcription reaction (20 µL) was diluted to 80 µL with molecular biology grade water and 2 µL of each sample were used for qPCR with the 2✕ Sybr Green Reaction Mix (Roche) and primers for *canis RPLP2* (housekeeping gene), *HSP90AA1* (Hsp90), *HSPA1A* (Hsp70), *DNAJB1* (Hsp40), and influenza matrix protein (primer sequences are provided in **Table 2.1**). Transcript levels of heat shock proteins were normalized to *RPLP2*, and normalized transcript levels for each treated condition were quantified relative to the basal environment at 37 °C (**Figure 2.2B**). For a positive control of productive infection, matrix protein transcript levels were compared with a prepared standard curve using the pHW187-M plasmid.

Table 2.1. RT-PCR primer sequences.

Primer name	Primer sequence
C. lupus familiaris DNAJB1 (fwd)	ACATGAACTTTGGCCGTTCC
C. lupus familiaris DNAJB1 (rev)	CTCTTTCCATCGGGGTTCAG
C. lupus familiaris HSP70 (fwd)	GGGGAGGACTTCGACAACAG
C. lupus familiaris HSP70 (rev)	GGACGACAAGGTCCTCTTGG
C. lupus familiaris HSPAA1 (fwd)	TGGGTTACATGGCAGCAAAG
C. lupus familiaris HSPAA1 (rev)	AGACTGAAGCCGGAGGACAG
C. lupus familiaris RPLP2 (fwd)	GCTACGTCGCCTCCTACCTG
C. lupus familiaris RPLP2 (rev)	GCTCGCTGATGACCTTGTTG
MATRIX_fwd	AGATGAGTCTTCTAACCGAGGTCG
MATRIX_rev	TGCAAAAACATCTTCAAGTCTCTG

RNA extraction and sequencing. MDCK[dn-cHSF1] cells were seeded at 100,000 cells/well in a 12-well plate and treated with 0.1% DMSO or 1 µg/mL doxycycline for 48 h. Each treatment was performed in biological triplicate. Cellular RNA was harvested using the RNeasy Plus Mini Kit with QIAshredder homogenization columns (Qiagen). RNA-Seq libraries were prepared using the Kapa mRNA HyperPrep RNA-Seq library construction kit (Kapa/Roche), with fragmentation times of 7 and 6 min at 94 ºC, for the 37 ºC and 39 ºC samples, respectively, and final amplifications of 15 and 12 PCR cycles, respectively. The resulting libraries were sequenced on an Illumina HiSeq using 40-bp single-end reads. RNA-seq quality control and differential expression analysis was performed as previously described [18]. Briefly, reads were aligned against the *Canis familiaris* genome assembly canFam3 with an ensembl annotation using STAR v. 2.5.3a. Gene expression was quantified using RSEM v. 1.3.0. Differential expression analysis to compare conditions (**Figure 2.1**) was performed using DESeq2 version 1.10.1 running under R version 3.2.3, complete dataset available online.

Resazurin cell growth assays. MDCK[dn-cHSF1] cells were seeded at 100,000 cells/well in 12-well plates and pre-treated with 0.1% DMSO or 1 µg/mL doxycycline. To mimic the infection conditions, we replaced the cellular growth media with WSN media, and supplemented with 0.1% DMSO or 1 µg/mL doxycycline 18 h after seeding. 48 h after the mock infection, the media was replaced with WSN media containing 50 µM resazurin sodium salt (Sigma). After 4 h of incubation, 100 µL of media was used to quantify resorufin fluorescence (excitation 530 nm; emission 590 nm) using a Take-3 plate reader (BioTeK) (**Figure 2.2D**).

Immunoblot analyses. MDCK[dn-cHSF1] cells were seeded at 200,000 cells/well in 6-well plates and treated with 1 µg/mL doxycycline or left untreated for 48 h. 8 h prior to harvesting, cells were treated with 100 µM sodium arsenite for 2 h and then allowed to recover in fresh media for 6 h. For protein levels examined during influenza infection, MDCK[dn-cHSF1] cells were seeded at 200,000 cells/well in 6-well plates, pre-treated with 0.1% DMSO or 1 µg/mL doxycycline for 24 h, and then infected with wild-type influenza A/Aichi/2/1968 at an MOI of 1 for 8 h. Cells were then harvested, washed 3× with PBS, and lysed (1% Triton, 200 mM NaCl, 50 mM Tris at pH 7.5, 1 mM EDTA, 1.5 mM $MgCl_2$, and protease inhibitor tablet from Thermo Fisher Scientific). After clearing the lysate by centrifugation at 21,100 × g for 15 min at 4 °C, 100 µg of total protein lysate was separated on a 12% SDS-PAGE gel and transferred to a nitrocellulose membrane using the Trans-Blot Turbo system (Bio-Rad). Membranes were then probed for Hsp90, Hsp70, Hsp40 and β-actin, as indicated, and imaged using an Odyssey Infrared Imaging System (Li-

Cor) (**Figure 2.2A**). To verify that MDCK[dn-cHSF1-MxA] and MDCK[dn-cHSF1-MxA(T103A)] cell lines expressed FLAG-MxA and FLAG-MxA(T103A), respectively, lysate from biological duplicate samples was separated on a 12% SDS-PAGE gel as described above, and immunoblotted for the FLAG epitope and β-actin (**Figure 2.13**; representative blot shown).

Deep mutational scanning. Two independent plasmid mutant libraries previously generated [24] from the pHWAichi68-NP template plasmid were used to create mutant viral libraries by transfecting a co-culture of 25,000 MDCK-SIAT1 and 300,000 HEK 293T cells, as previously described [24]. For each library, co-cultures in eight 6-well plates were transfected to maintain library diversity, and transfection supernatants were combined to generate the input mutant viral libraries. Viruses were generated and grown in WSN media. Infectious titers of viral libraries were determined by a tissue culture infectious dose ($TCID_{50}$) assay. Briefly, 10-fold serial dilutions of viruses (in technical triplicates) were prepared in 96-well plates and incubated with 5×10^3 MDCK-SIAT1 cells/well for 72 h at 37 °C. The wells were then scored for cytopathic effects and viral titers were calculated using a Reed-Muench Calculator, available at: https://github.com/jbloomlab/reedmuenchcalculator.

Two plasmid libraries were used to generate two biological replicate viral libraries, one of which was used twice to perform two technical replicates of the deep mutational scanning (**Figure 2.4**). MDCK[dn-cHSF1] cells were plated in 15 cm plates at a density of 6×10^6 cells/dish, and treated with 0.1% DMSO or 1 µg/ml doxycycline for 24 h at either 37 °C or 39 °C. Deep mutational scanning was also performed in cells treated with an Hsp90 inhibitor (100 nM 17-AAG; 90 min. pre-treatment). 5×10^6 infectious virions (as determined by a $TCID_{50}$ assay) from each viral library were used to infect four 15 cm plates from each condition at an MOI of 0.1 virions/cell (the cells expanded to ~12.5×10^6 cells per plate by the time of infection). In addition, one 15 cm plate at both 37 °C and 39 °C was either mock- infected (negative control) or infected with wild-type virus. For infection, the cellular growth media was replaced with WSN media containing a mutant virus library, wild-type virus, or no virus for mock infection. After 2 h, the inoculum was replaced with fresh WSN media containing 0.1% DMSO, 1 µg/mL doxycycline, or 100 nM 17-AAG. 48 h post-infection, the viral supernatant was harvested, centrifuged at 1000 × g for 5 min to remove cell debris, and stored at –80 °C. Viral RNA was extracted from the infectious supernatant using a Viral RNA Mini Kit (Qiagen) and reverse transcribed using the AccuScript High Fidelity 1st Strand cDNA Synthesis Kit (Agilent) using 5'-BsmBI-Aichi68-NP and 3'-BsmBI-Aichi68-NP primers (primer sequences in **Table 2.2**). At least

10^6 NP molecules were PCR-amplified for preparation of the sequencing libraries, as previously described [27] (**Figure 2.4B**). The resulting amplicons were sequenced on an Illumina HiSeq 2500 in rapid run mode with 2× 250 bp paired-end reads (**Figure 2.4C–D**). dms_tools (http://jbloomlab.github.io/dms_tools/) [56] was used to align reads to the Aichi NP reference sequence, count amino acid variants across NP, and calculate the differential selection for each variant between two selection conditions, as previously described [9, 28].

Table 2.2. Sequencing primer sequences.

Primer name	Primer sequence
5'-BsmBI-Aichi68-NP	CATGATCGTCTCAGGGAGCAAAAGCAGGGTAGATAATCACTCACAG
3'-BsmBI-Aichi68-NP	CATGATCGTCTCGTATTAGTAGAAACAAGGGTATTTTTCTTTA
Subamplicon_1_fwd	CTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNTCACTCACAGAGTGACATCGAAATC
Subamplicon_1_rev	GGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNGTTTTCTTAGGATCCTTCCCCGC
Subamplicon_2_fwd	CTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNGGAATAAATATCTGGAAGAACATCCCAGC
Subamplicon_2_rev	GGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNCCTAGGGAGAGTCGAACCCTG
Subamplicon_3_fwd	CTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNCCCAGGATGTGCTCTCTGATG
Subamplicon_3_rev	GGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNCAGAAAGATGAGATCTTCGATCTCAGC
Subamplicon_4_fwd	CTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNGAGAAAGTCGGAACCCAGGAAAT
Subamplicon_4_rev	GGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNCTTAATACTCTTAGATCTTCAAATGCAGCAGA
Subamplicon_5_fwd	CTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNTGGTGTGGATGGCATGCAAT
Subamplicon_5_rev	GGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNTGATGGTTGGTTTGTCAAATGGGAG
Subamplicon_6_fwd	CTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNACCTGCATTTTCTGTGCAAAGAAAC
Subamplicon_6_rev	GGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNCGGGTTATTAGTAGAAACAAGGGTATTTTTCT
Universal_fwd	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
Rnd2_rev_1	CAAGCAGAAGACGGCATACGAGATTAGTTGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Rnd2_rev_2	CAAGCAGAAGACGGCATACGAGATCCGGTGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Rnd2_rev_3	CAAGCAGAAGACGGCATACGAGATCCGGTGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Rnd2_rev_4	CAAGCAGAAGACGGCATACGAGATGTAGCCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Rnd2_rev_5	CAAGCAGAAGACGGCATACGAGATGGACTGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Rnd2_rev_6	CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT

Rnd2_rev_7	CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTCAGACGT GTGCTCTTCCGATCT
Rnd2_rev_8	CAAGCAGAAGACGGCATACGAGATGATCTGGTGACTGGAGTTCAGACGT GTGCTCTTCCGATCT
Rnd2_rev_9	CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTCAGACGT GTGCTCTTCCGATCT
Rnd2_rev_10	CAAGCAGAAGACGGCATACGAGATCACTGTGTGACTGGAGTTCAGACGT GTGCTCTTCCGATCT
Rnd2_rev_11	CAAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAGTTCAGACGT GTGCTCTTCCGATCT
Rnd2_rev_12	CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTTCAGACGT GTGCTCTTCCGATCT
Rnd2_rev_13	CAAGCAGAAGACGGCATACGAGATTCAAGTGTGACTGGAGTTCAGACGT GTGCTCTTCCGATCT
Rnd2_rev_14	CAAGCAGAAGACGGCATACGAGATCTGATCGTGACTGGAGTTCAGACGT GTGCTCTTCCGATCT
Rnd2_rev_15	CAAGCAGAAGACGGCATACGAGATAAGCTAGTGACTGGAGTTCAGACGT GTGCTCTTCCGATCT
Rnd2_rev_16	CAAGCAGAAGACGGCATACGAGATTACAAGGTGACTGGAGTTCAGACGT GTGCTCTTCCGATCT
Rnd2_rev_17	CAAGCAGAAGACGGCATACGAGATTGTTGACTGTGACTGGAGTTCAGAC GTGTGCTCTTCCGATCT
Rnd2_rev_18	CAAGCAGAAGACGGCATACGAGATACGGAACTGTGACTGGAGTTCAGAC GTGTGCTCTTCCGATCT
Rnd2_rev_19	CAAGCAGAAGACGGCATACGAGATTCTGACATGTGACTGGAGTTCAGAC GTGTGCTCTTCCGATCT
Rnd2_rev_20	CAAGCAGAAGACGGCATACGAGATGTGCGGACGTGACTGGAGTTCAGAC GTGTGCTCTTCCGATCT
Rnd2_rev_21	CAAGCAGAAGACGGCATACGAGATCGTTTCACGTGACTGGAGTTCAGAC GTGTGCTCTTCCGATCT
Rnd2_rev_22	CAAGCAGAAGACGGCATACGAGATAAGGCCACGTGACTGGAGTTCAGAC GTGTGCTCTTCCGATCT
Rnd2_rev_23	CAAGCAGAAGACGGCATACGAGATTCCGAAACGTGACTGGAGTTCAGAC GTGTGCTCTTCCGATCT
Rnd2_rev_24	CAAGCAGAAGACGGCATACGAGATCGTACGGTGACTGGAGTTCAGACGT GTGCTCTTCCGATCT
Rnd2_rev_25	CAAGCAGAAGACGGCATACGAGATATCCACTCGTGACTGGAGTTCAGAC GTGTGCTCTTCCGATCT
Rnd2_rev_26	CAAGCAGAAGACGGCATACGAGATATCAGTGTGACTGGAGTTCAGACGT GTGCTCTTCCGATCT
Rnd2_rev_27	CAAGCAGAAGACGGCATACGAGATAAAGGAATGTGACTGGAGTTCAGAC GTGTGCTCTTCCGATCT

Pairwise viral competitions. Nucleoprotein variants that reproducibly exhibited the most positive or negative differential selection in the deep mutational scan were selected for pairwise competitions (P283M, P283A, P283G, P283L, P283S, P283T, S353L, S353F, N334H, D34N, H82N and a synonymous control P283P). The nucleoprotein mutants were generated by introducing point mutations into the pHWAichi68-NP plasmid using the QuikChange II XL Site-Directed Mutagenesis Kit (Agilent). Technical difficulties with specific site-directed mutagenesis reactions prevented generation of the P283M, P283A, and S353L mutant plasmids. The remaining nine mutant plasmids were generated successfully, and used to produce the

corresponding mutant viruses by transfecting a co-culture of 25,000 MDCK-SIAT1 and 300,000 HEK 293T cells, as previously described [55]. The resultant viruses were titered using a TCID$_{50}$ assay. For each competition, 100,000 cells/well of MDCK$^{dn-cHSF1}$, MDCK$^{dn-cHSF1-FLAG-MxA}$, or MDCK$^{dn-cHSF1-FLAG-MxA(T103A)}$ cells were plated in 12-well dishes and treated with 0.1% DMSO or 1 µg/mL doxycycline for 18 h at either 37 °C or 39 °C. Cells were infected with a 1:1 mixture of wild-type and mutant viruses at an MOI of 0.1 virions/cell in triplicate under conditions identical to that of the deep mutational scanning experiment. After 2 h, the inoculum was replaced with fresh WSN media containing either 0.1% DMSO or 1 µg/mL doxycycline. 48 h post-infection, infectious supernatants were harvested, centrifuged at 1000 × g for 5 min to remove cell debris, and stored at –80 °C. Viral RNA was extracted from the infectious supernatant using the QIAamp Viral RNA Mini Kit and at least 10^6 NP molecules were reverse transcribed using the SuperScript III Reverse Transcriptase (Thermo Fisher Scientific) with 5'-BsmBI-Aichi68-NP and 3'-BsmBI-Aichi68-NP primers (**Table 2.2**). The amplicons were visualized on a 1% analytical agarose gel to verify amplification of the NP gene (1.5 kb). The dsDNA was purified using 1.5× AMPure XP beads (Beckman Coulter) and quantified using a Quant-iT PicoGreen Assay (Life Technologies). Illumina NexteraXT sequencing libraries were prepared using a Mosquito HTS Liquid Handler (TTP Labtech) and sequenced on an Illumina MiSeqv2 in two runs of either 40-bp single-end or 150-bp paired-end reads. To call sequence variants, reads were aligned to the Aichi NP reference sequence using bwa mem (v. 0.7.12-r1039) (arXiv:1303.3997v2) with flag –t 16, and sorted and indexed bam files were generated using samtools (v 1.3) [57]. These bam files were processed using samtools mpileup with flags –excl-flags 2052, -d 30000000 and the same Aichi NP reference sequence used for mapping [58]. For pairwise competitions in the absence of MxA, mutant allele frequencies were normalized to wild-type allele frequencies for each sample, and the resulting values were used to calculate the differential selection [28] (**Figure 2.9**). For pairwise competitions in the presence of wild-type or inactive MxA, the change in mutant allele frequencies is reported (**Figure 2.12C**).

Molecular dynamics (MD) simulations. Two sets of simulations were performed for nucleoproteins with the sequence of the human H3N2 variant (**Figures 2.10A–B**). In one set, nucleoprotein residue 283 was proline (this system is termed Pro283 hereafter). In the other set, residue 283 was serine (this system is termed Ser283 hereafter). The initial structures of both systems were prepared using the comparative modeling software RosettaCM [59], with the structures of H1N1 influenza A virus nucleoprotein (PDB ID: 2IQH [36]) and H5N1 nucleoprotein (PDB ID: 2Q06 [37]) used as templates with equal weights. The first 20 amino acids in the N-

terminal region, whose 3D coordinates are missing in the template structures, were considered flexible and also of minimal impact to the region near residue 283. Therefore, they were removed in the following simulation.

In the threading procedure, the target nucleoprotein sequence was aligned with the templates [60] and assigned coordinates from the template PDB structures (2IQH and 2Q06). The helix formed by residues 278–286 of 2IQH and 2Q06 was removed to allow RosettaCM to construct this region without influence from the templates. This region (residues 278–286), along with all the other regions missing in the template PDBs, were patched in the hybridization step. During hybridization, RosettaCM generated hybridized structures that contained pieces from each of the threaded structures, providing more accurate comparative models that were energetically favorable. Additionally, RosettaCM used fragments and minor *ab initio* folding to fill-in residues not previously aligned with any template sequences during the threading process. One thousand models for each nucleoprotein system were created, and the best-scoring model without a disulfide bond was used as the initial structure for further MD simulations.

Five runs of MD simulations for each nucleoprotein system were carried out using GROMACS with the oplsaa/tip4p force field [61-62]. The N-terminus of the initial structure of the MD simulation was capped with an acetyl group, while the C-terminus was free (ending with COO^-). This structure was energy minimized in a vacuum, and then immersed in the center of a cubic box containing pre-equilibrated water molecules with an edge of 12 nm. The system was electrostatically neutralized by adding 11 Cl^- ions. The solvated system was further energy minimized to remove any bad contacts. The solvated nucleoprotein then underwent two stages of equilibrations. The first stage of equilibration consisted of a 50 ps isochoric–isothermal (NVT) simulation at 300 K and a subsequent 50 ps isobaric–isothermal (NPT) simulation at 300 K and 1 bar. During the first stage of equilibration, the nucleoprotein heavy atoms were restrained by a harmonic potential with a force constant of 1,000 kJ $mol^{-1}nm^{-2}$ to equilibrate the solvent molecules and adjust the density. The second stage of equilibration consisted of an additional 100 ps NVT simulation at 300 K without any restraints to equilibrate the whole system, followed by a 100 ns NPT production simulation at 300 K and 1 bar. The V-rescale thermostat was coupled to both the nucleoprotein and solvent separately, with coupling time constants of 0.1 ps. The pressure was maintained using the Parrinello-Rahman barostat with a coupling time constant of 2.0 ps and isothermal compressibility of 4.5×10^{-5} bar^{-1}. The leapfrog algorithm with a time step of 2 fs was used for dynamics evolution. All bonds involving hydrogen were constrained using the LINCS algorithm. All neighbor searching, electrostatic interactions and

van der Waals interactions were truncated at 1.0 nm. Electrostatics were treated using the particle mesh Ewald (PME) summation with a Fourier spacing of 0.12 nm and an order of 4. A long-range dispersion correction for energy and pressure was applied to account for the 1.0 nm cut-off of Lennard-Jones interactions. Five 100 ns trajectories were produced for the Pro283 and Ser283 systems, respectively. The trajectories between 50 ns and 100 ns were used for analysis.

Recombinant expression and biophysical characterization of nucleoprotein variants

The P283S, P283A, and P283L amino acid substitutions were introduced into the wild-type influenza A/Aichi/2/1968 nucleoprotein in a pET28b(+) expression vector using the QuikChange II XL Site-Directed Mutagenesis Kit (Agilent). Notably, this nucleoprotein construct contained an R416A mutation and deletion of residues 2–7 to obtain non-aggregated, RNA-free nucleoprotein in a CD-compatible buffer, as previously described [10]. Mutagenized plasmid DNA was isolated using the E.Z.N.A. Plasmid Mini Kit I (Omega). For bacterial expression, BL21(DE3) chemically competent cells were transformed with 1 µL of purified plasmid and incubated overnight on LB-kanamycin agar plates. Colonies were used to inoculate 50 mL LB-kanamycin cultures overnight. 10 mL of starter cultures were then used to inoculate 1 L LB-kanamycin cultures, which were shaken at 37 °C until reaching an OD_{600} of 0.3–0.6. Cultures were then chilled on ice, and induced with 500 µM IPTG (Sigma) overnight at 20 °C. Cells were then pelleted, dounce-homogenized, and lysed by sonication in 50 mL of lysis buffer (50 mM sodium phosphate at pH 8.0, 500 mM NaCl, 0.5% Triton X-100, 10 mM imidazole, 1 mM PMSF, and 0.1 mg/mL $MgCl_2$). Cells were sonicated for 2 min (30% amplitude, 10 sec on, 10 sec off; Branson Digital Sonifier). Lysates were then clarified for 30 min at 10,000 × g at 4 °C. Clarified lysates were passed through 0.45 µm filters. His-Tagged nucleoprotein variants were then incubated on Ni-NTA (Millipore) columns for 60 min at 4 °C and washed with Ni-NTA Wash Buffer (50 mM sodium phosphate at pH 8.0, 300 mM NaCl, and 20 mM imidazole). Proteins were eluted using Ni-NTA Elution Buffer (50 mM sodium phosphate at pH 8.0, 300 mM NaCl, and 250 mM imidazole). Eluates were dialyzed overnight into analysis buffer (20 mM sodium phosphate at pH 7.0 with 300 mM NaF) using 3.5 kDa molecular weight cutoff SnakeSkin dialysis tubing (Fisher Thermo Scientific). Dialyzed proteins were concentrated using Amicon Ultra 3K MWCO filters (Millipore) and further purified over a size exclusion column (Bio-Rad ENrich SEC 650) (**Figure 2.11**).

For circular dichroism analysis (**Figure 2.11B**), proteins were diluted to 5 μM in analysis buffer (quantified by A280 with a BioTek-Take3 micro-volume plate using a molar extinction coefficient of 56,600 M^{-1} cm^{-1}). Thermal melts (**Figure 2.11C**) were performed at a scan rate of 2 °C per min, maintaining each temperature for 5 min before measurement of ellipticity at 209 nm. T_{agg} values were obtained, as the thermal denaturation of nucleoprotein was irreversible and resulted in aggregation and precipitation (**Figures 2.10C and 2.11C**). All circular dichroism analyses were performed on a Jasco J-1500 Circular Dichroism Spectrophotometer with a 1 mm QS quartz cuvette (Hellma).

Statistical analyses. Deep mutational scanning was performed in biological duplicate with two technical replicates of one of the biological replicates (**Figure 2.4A**). MxA and MxA(T103A) protein expression in MDCK[dn-cHSF1-FLAG-MxA] and MDCK[dn-cHSF1-FLAG-MxA(T103A)] cells, respectively, were evaluated in biological duplicate (**Figure 2.13**). All other experiments were performed in biological triplicate with replicates being independent experimental entireties (i.e., from plating the cells to acquiring the data). Correlation between deep mutational scanning replicates was determined by linear regression using Graph Pad Prism software, reporting R^2 (**Figure 2.4E**). Site differential selection values from deep mutational scanning (**Figure 2.3B**) were tested for significance of deviation from zero (wild-type behavior) using a one-sample t-test in GraphPad Prism. The raw p-values were adjusted for multiple-comparison using the Benjamini-Hochberg procedure [63], setting an acceptable false-discovery rate (FDR) at 0.05 (p.adjust function in R). Analysis of Variance (ANOVA) was performed on all differential selection values across selections normalized to the Basal 37 °C condition using a nested ANOVA framework (accounting for replicates), modeling treatment/temperature as a fixed effect and the replicate as a random effect (lme function/RMLE in the R statistical environment, followed by ANOVA computation anova.lme with sequential adjustment). Post-hoc pairwise comparisons were performed by general linear hypotheses testing using Tukey's method (as implemented in glht, in the multcomp R package) comparing all the means with single-step adjustment for multiple comparison. The p-values for pairwise comparisons against HSF1i 37 °C vs. Basal 37 °C were: 6.148×10^{-6} for Basal 39 °C vs. Basal 37 °C and 9.196×10^{-8} for HSF1i 39 °C vs. Basal 37 °C and 0.71 for comparison between HSF1i 39 °C vs. Basal 37 °C and Basal 39 °C vs. Basal 37 °C (**Figure 2.3B**). Differential selection values from pairwise competitions (**Figure 2.9B**) were tested for significance of deviation from zero (wild-type behavior), using a one-sample t-test in GraphPad Prism with false-discovery rate correction. For pairwise competitions in the presence of MxA, significance of deviation from the input Ser283 frequency was determined using a one-

sample *t*-test in GraphPad Prism followed by false-discovery rate correction; Basal 37 °C Inactive MxA (t = 5.916, df = 2); Basal 37 °C Active MxA (t = 29.01, df = 2); HSF1i 37 °C Inactive MxA (t = 18.24, df = 2); HSF1i 37 °C Active MxA (t = 17.91, df = 2); Basal 39 °C Inactive MxA (t = 4.924, df = 2); Basal 39 °C Active MxA (t = 0.5877, df = 2); HSF1i 39 °C Inactive MxA (t = 18.98, df = 2); HSF1i 39 °C Active MxA (t = 12.29, df = 2). ANOVA was performed for competitions in presence of MxA as described above. For RNA-seq, log2 fold-changes, *p*-values, and Benjamini-Hochberg-adjusted *p*-values (ADP) are reported for all expressed protein-coding genes and can be found online. For MD simulations, each of the five 100 ns simulations was considered an independent replicate, and the %-time spent in an α-helical conformation was transformed using $\ln(P/(1-P))$ before *t*-tests were conducted to satisfy the prerequisite assumptions of normality (**Figure 2.10B**); 278 (t = 0.95655, df = 7); 279 (t = 0.52331, df = 6); 280 (t = −0.46542, df = 6); 281 (t = −1.21825, df = 6); 282 (t = −3.64906, df = 7); 283 (t = −2.83812, df = 8); 284 (t = −2.83606, df = 8); 285 (t = 0.17240, df = 7); 286 (t = 0.34125, df = 8). For apparent melting temperatures determined by circular dichroism, melt curves were performed in biological triplicate, the average and SEM are reported, and significance of deviation from wild-type was evaluated by a Student's *t*-test; Ser (t = 19.59, df = 4); Ala (t = 23.94, df = 4); Leu (t = 5.644, df = 4).

2.6   Acknowledgements

2.7   References

1.   Lauring, A. S.; Frydman, J.; Andino, R. The role of mutational robustness in RNA virus evolution. *Nat. Rev. Microbiol.* **2013,** *11* (5), 327-36.
2.   DePristo, M. A.; Weinreich, D. M.; Hartl, D. L. Missense meanderings in sequence space: a biophysical view of protein evolution. *Nat. Rev. Genet.* **2005,** *6* (9), 678-687.
3.   Weinreich, D. M.; Delaney, N. F.; Depristo, M. A.; Hartl, D. L. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **2006,** *312* (5770), 111-4.
4.   Wylie, C. S.; Shakhnovich, E. I. A biophysical protein folding model accounts for most mutational fitness effects in viruses. *Proc. Natl. Acad. Sci. U. S. A.* **2011,** *108* (24), 9916-9921.
5.   Bloom, J. D.; Labthavikul, S. T.; Otey, C. R.; Arnold, F. H. Protein stability promotes evolvability. *Proc. Natl. Acad. Sci. U. S. A.* **2006,** *103* (15), 5869-5874.
6.   Tokuriki, N.; Tawfik, D. S. Stability effects of mutations and protein evolvability. *Curr. Opini. Struct. Biol.* **2009,** *19* (5), 596-604.
7.   Arranz, R.; Coloma, R.; Chichón, F. J.; Conesa, J. J.; Carrascosa, J. L.; Valpuesta, J. M.; Ortín, J.; Martín-Benito, J. The structure of native influenza virion ribonucleoproteins. *Science* **2012,** *338*, 1634-1637.
8.   Bhatt, S.; Holmes, E. C.; Pybus, O. G. The genomic rate of molecular adaptation of the human influenza A virus. *Mol. Biol. Evol.* **2011,** *28* (9), 2443-2451.
9.   Ashenberg, O.; Padmakumar, J.; Doud, M. B.; Bloom, J. D. Deep mutational scanning identifies sites in influenza nucleoprotein that affect viral inhibition by MxA. *PLoS Pathog.* **2017,** *13* (3), e1006288.
10.   Gong, L. I.; Suchard, M. A.; Bloom, J. D. Stability-mediated epistasis constrains the evolution of an influenza protein. *eLife* **2013,** *2*, e00631.
11.   Zimmermann, P.; Mänz, B.; Haller, O.; Schwemmle, M.; Kochs, G. The viral nucleoprotein determines Mx sensitivity of influenza A viruses. *J. Virol.* **2011,** *85* (16), 8133-8140.
12.   Turan, K.; Mibayashi, M.; Sugiyama, K.; Saito, S.; Numajiri, A.; Nagata, K. Nuclear MxA proteins form a complex with influenza virus NP and inhibit the transcription of the engineered influenza virus genome. *Nucleic Acids Res.* **2004,** *32* (2), 643-652.
13.   Götz, V.; Magar, L.; Dornfeld, D.; Giese, S.; Pohlmann, A.; Höper, D.; Kong, B.-W.; Jans, D. A.; Beer, M.; Haller, O. Influenza A viruses escape from MxA restriction at the expense of efficient nuclear vRNP import. *Sci. Rep.* **2016,** *6*, 23138.
14.   Rabadan, R.; Levine, A. J.; Robins, H. Comparison of avian and human influenza A viruses reveals a mutational bias on the viral genomes. *J. Virol.* **2006,** *80* (23), 11887-11891.
15.   Smith, G. J.; Bahl, J.; Vijaykrishna, D.; Zhang, J.; Poon, L. L.; Chen, H.; Webster, R. G.; Peiris, J. M.; Guan, Y. Dating the emergence of pandemic influenza viruses. *Proc. Natl. Acad. Sci. U. S. A.* **2009,** *106* (28), 11709-11712.
16.   Mänz, B.; Dornfeld, D.; Götz, V.; Zell, R.; Zimmermann, P.; Haller, O.; Kochs, G.; Schwemmle, M. Pandemic influenza A viruses escape from restriction by human MxA through adaptive mutations in the nucleoprotein. *PLoS Pathog.* **2013,** *9* (3), e1003279.
17.   Morens, D. M.; Taubenberger, J. K.; Fauci, A. S. The persistent legacy of the 1918 influenza virus. *N. Engl. J. Med.* **2009,** *361* (3), 225-229.
18.   Phillips, A. M.; Gonzalez, L. O.; Nekongo, E. E.; Ponomarenko, A. I.; McHugh, S. M.; Butty, V. L.; Levine, S. S.; Lin, Y.-S.; Mirny, L. A.; Shoulders, M. D. Host proteostasis modulates influenza evolution. eL*ife* **2017,** *6*, e28652.
19.   Geller, R.; Pechmann, S.; Acevedo, A.; Andino, R.; Frydman, J. Hsp90 shapes protein and RNA evolution to balance trade-offs between protein stability and aggregation. *Nat. Comm.* **2018,** *9*, 1781.
20.   Watanabe, T.; Watanabe, S.; Kawaoka, Y. Cellular networks involved in the influenza virus life cycle. *Cell Host Microbe* **2010,** *7* (6), 427-439.

21.	York, A.; Hutchinson, E. C.; Fodor, E. Interactome analysis of the influenza A virus transcription/replication machinery identifies protein phosphatase 6 as a cellular factor required for efficient virus replication. *J. Virol.* **2014,** *88* (22), 13284-13299.

22.	Fowler, D. M.; Fields, S. Deep mutational scanning: a new style of protein science. *Nat. Methods* **2014,** *11* (8), 801-807.

23.	Moore, C. L.; Dewal, M. B.; Nekongo, E. E.; Santiago, S.; Lu, N. B.; Levine, S. S.; Shoulders, M. D. Transportable, chemical genetic methodology for the small molecule-mediated inhibition of heat shock factor 1. *ACS Chem. Biol.* **2015,** *11* (1), 200-210.

24.	Doud, M. B.; Ashenberg, O.; Bloom, J. D. Site-Specific Amino Acid Preferences Are Mostly Conserved in Two Closely Related Protein Homologs. *Mol. Biol. Evol.* **2015,** *32* (11), 2944-60.

25.	Seitz, C.; Frensing, T.; Höper, D.; Kochs, G.; Reichl, U. High yields of influenza A virus in Madin–Darby canine kidney cells are promoted by an insufficient interferon-induced antiviral state. *J. Gen. Virol.* **2010,** *91* (7), 1754-1763.

26.	Trinklein, N. D.; Murray, J. I.; Hartman, S. J.; Botstein, D.; Myers, R. M. The role of heat shock transcription factor 1 in the genome-wide regulation of the mammalian heat shock response. *Mol. Biol. Cell* **2004,** *15* (3), 1254-1261.

27.	Doud, M. B.; Bloom, J. D. Accurate measurement of the effects of all amino-acid mutations on influenza hemagglutinin. *Viruses* **2016,** *8* (6), 155.

28.	Doud, M. B.; Hensley, S. E.; Bloom, J. D. Complete mapping of viral escape from neutralizing antibodies. *PLoS Pathog.* **2017,** *13* (3), e1006271.

29.	Sun, Y. The antiviral mechanism of MxA revealed by a molecular docking-based structural model of the complex of MxA and influenza A virus NP/RNPs. *bioRxiv* **2017**.

30.	Bernasconi, D.; Schultz, U.; Staeheli, P. The interferon-induced Mx protein of chickens lacks antiviral activity. *J. Interferon Cytokine Res.* **1995,** *15* (1), 47-53.

31.	Goodman, A. G.; Smith, J. A.; Balachandran, S.; Perwitasari, O.; Proll, S. C.; Thomas, M. J.; Korth, M. J.; Barber, G. N.; Schiff, L. A.; Katze, M. G. The cellular protein P58IPK regulates influenza virus mRNA translation and replication through a PKR-mediated mechanism. *J. Virol.* **2007,** *81* (5), 2221-30.

32.	Melville, M. W.; Tan, S.-L.; Wambach, M.; Song, J.; Morimoto, R. I.; Katze, M. G. The cellular inhibitor of the PKR protein kinase, P58IPK, is an influenza virus-activated co-chaperone that modulates heat shock protein 70 activity. *J. Biol. Chem.* **1999,** *274* (6), 3797-3803.

33.	Sharma, K.; Tripathi, S.; Ranjan, P.; Kumar, P.; Garten, R.; Deyde, V.; Katz, J. M.; Cox, N. J.; Lal, R. B.; Sambhara, S. Influenza A virus nucleoprotein exploits Hsp40 to inhibit PKR activation. *PLoS One* **2011,** *6* (6), e20215.

34.	Batra, J.; Tripathi, S.; Kumar, A.; Katz, J. M.; Cox, N. J.; Lal, R. B.; Sambhara, S.; Lal, S. K. Human Heat shock protein 40 (Hsp40/DnaJB1) promotes influenza A virus replication by assisting nuclear import of viral ribonucleoproteins. *Sci. Rep.* **2016,** *6*, 19063.

35.	Hirayama, E.; Atagi, H.; Hiraki, A.; Kim, J. Heat shock protein 70 is related to thermal inhibition of nuclear export of the influenza virus ribonucleoprotein complex. *J. Virol.* **2004,** *78* (3), 1263-1270.

36.	Ye, Q.; Krug, R. M.; Tao, Y. J. The mechanism by which influenza A virus nucleoprotein forms oligomers and binds RNA. *Nature* **2006,** *444* (7122), 1078.

37.	Ng, A. K.; Zhang, H.; Tan, K.; Li, Z.; Liu, J. H.; Chan, P. K.; Li, S. M.; Chan, W. Y.; Au, S. W.; Joachimiak, A.; Walz, T.; Wang, J. H.; Shaw, P. C. Structure of the influenza virus A H5N1 nucleoprotein: implications for RNA binding, oligomerization, and vaccine design. *FASEB J.* **2008,** *22* (10), 3638-47.

38.	Guzzo, A. V. The Influence of Amino Acid Sequence on Protein Structure. *Biophys. J.* **1965,** *5* (6), 809-822.

39.   White, C. R.; Phillips, N. F.; Seymour, R. S. The scaling and temperature dependence of vertebrate metabolism. *Biology Lett.* **2006,** *2* (1), 125-7.
40.   Geller, R.; Vignuzzi, M.; Andino, R.; Frydman, J. Evolutionary constraints on chaperone-mediated folding provide an antiviral approach refractory to development of drug resistance. *Genes Dev.* **2007,** *21* (2), 195-205.
41.   Geller, R.; Taguwa, S.; Frydman, J. Broad action of Hsp90 as a host chaperone required for viral replication. *Biochim. Biophys. Acta, Mol. Cell Res.* **2012,** *1823* (3), 698-706.
42.   Chase, G.; Deng, T.; Fodor, E.; Leung, B. W.; Mayer, D.; Schwemmle, M.; Brownlee, G. Hsp90 inhibitors reduce influenza virus replication in cell culture. *J. Virol.* **2008,** *377* (2), 431-439.
43.   Momose, F.; Naito, T.; Yano, K.; Sugimoto, S.; Morikawa, Y.; Nagata, K. Identification of Hsp90 as a stimulatory host factor involved in influenza virus RNA synthesis. *J. Biol. Chem.* **2002,** *277* (47), 45306-45314.
44.   Hu, J.; Seeger, C. Hsp90 is required for the activity of a hepatitis B virus reverse transcriptase. *Proc. Natl. Acad. Sci. U. S. A* **1996,** *93* (3), 1060-1064.
45.   Queitsch, C.; Sangster, T. A.; Lindquist, S. Hsp90 as a capacitor of phenotypic variation. *Nature* **2002,** *417* (6889), 618-624.
46.   Bogumil, D.; Dagan, T. Cumulative impact of chaperone-mediated folding on genome evolution. *Biochemistry* **2012,** *51* (50), 9941-9953.
47.   Sangster, T. A.; Salathia, N.; Lee, H. N.; Watanabe, E.; Schellenberg, K.; Morneau, K.; Wang, H.; Undurraga, S.; Queitsch, C.; Lindquist, S. HSP90-buffered genetic variation is common in Arabidopsis thaliana. *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105* (8), 2969-2974.
48.   Rohner, N.; Jarosz, D. F.; Kowalko, J. E.; Yoshizawa, M.; Jeffery, W. R.; Borowsky, R. L.; Lindquist, S.; Tabin, C. J. Cryptic variation in morphological evolution: HSP90 as a capacitor for loss of eyes in cavefish. *Science* **2013,** *342* (6164), 1372-1375.
49.   Geiler-Samerotte, K. A.; Zhu, Y. O.; Goulet, B. E.; Hall, D. W.; Siegal, M. L. Selection transforms the landscape of genetic variation interacting with Hsp90. *PLoS Biol.* **2016,** *14* (10), e2000465.
50.   Watanabe, T.; Kawakami, E.; Shoemaker, J. E.; Lopes, T. J.; Matsuoka, Y.; Tomita, Y.; Kozuka-Hata, H.; Gorai, T.; Kuwahara, T.; Takeda, E.; Nagata, A.; Takano, R.; Kiso, M.; Yamashita, M.; Sakai-Tagawa, Y.; Katsura, H.; Nonaka, N.; Fujii, H.; Fujii, K.; Sugita, Y.; Noda, T.; Goto, H.; Fukuyama, S.; Watanabe, S.; Neumann, G.; Oyama, M.; Kitano, H.; Kawaoka, Y. Influenza virus-host interactome screen as a platform for antiviral drug development. *Cell Host Microbe* **2014,** *16* (6), 795-805.
51.   Tokuriki, N.; Tawfik, D. S. Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature* **2009,** *459* (7247), 668.
52.   Wyganowski, K. T.; Kaltenbach, M.; Tokuriki, N. GroEL/ES buffering and compensatory mutations promote protein evolution by stabilizing folding intermediates. *J. Mol. Biol.* **2013,** *425* (18), 3403-14.
53.   Ortlund, E. A.; Bridgham, J. T.; Redinbo, M. R.; Thornton, J. W. Crystal structure of an ancient protein: evolution by conformational epistasis. *Science* **2007,** *317* (5844), 1544-1548.
54.   Bloom, J. D. An experimentally determined evolutionary model dramatically improves phylogenetic fit. *Mol. Biol. Evol.* **2014,** *31* (8), 1956-1978.
55.   Hoffmann, E.; Neumann, G.; Kawaoka, Y.; Hobom, G.; Webster, R. G. A DNA transfection system for generation of influenza A virus from eight plasmids. *Proc. Natl. Acad. Sci. U. S. A.* **2000,** *97* (11), 6108-6113.
56.   Bloom, J. D. Software for the analysis and visualization of deep mutational scanning data. *BMC Bioinform.* **2015,** *16*, 168.
57.   Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009,** *25* (16), 2078-9.

58.   Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **2011,** *27* (21), 2987-93.

59.   Song, Y.; DiMaio, F.; Wang, R. Y.-R.; Kim, D.; Miles, C.; Brunette, T.; Thompson, J.; Baker, D. High-resolution comparative modeling with RosettaCM. *Structure* **2013,** *21* (10), 1735-1742.

60.   Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **2011,** *7* (1), 539.

61.   Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008,** *4* (3), 435-447.

62.   Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996,** *118* (45), 11225-11236.

63.   Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **1995**, 289-300.

Chapter 3: HSF1 activation can restrict HIV replication

3.1   Author Contributions

E.E.N. and M.D.S. conceived the project. E.E.N., A.I.P., M.B.D., E.P.B., and M.D.S. designed the experiments. E.E.N., A.I.P., and M.B.D. performed the experiments. E.E.N., A.I.P., V.L.B., E.P.B., and M.D.S. analyzed the data. M.D.S. supervised the research.

3.2   Introduction

Human immunodeficiency virus-1 (HIV-1) remains a serious global health threat, with approximately 37 million people currently living with HIV/AIDS.[1] While the number of HIV-related deaths continues to decline, owing to advances in treatment and prevention strategies in the past decades,[2] the epidemic still claims nearly one million lives annually. The problems of latent infection and drug resistance remain, as does the continued failure to develop an effective HIV vaccine.

With respect to the development of new therapeutic modalities impervious to antiviral resistance mechanisms, not just for HIV but also for other RNA viruses, the alternative strategy of targeting host systems instead of the rapidly mutating virus itself has gained increasing traction.[3-4] As a minimalistic pathogen, HIV-1 requires complex interactions with host systems for replication.[5-6] A clear understanding of the intimate interplay between the host and the virus is essential to provide an effective roadmap for viable, host-targeted antiviral therapeutics.[4, 7]

Stress responses evolved to defend cells against damaging internal and external stimuli. In some cases, stress responses can provide defenses against invading pathogens. However, numerous viral pathogens have also developed strategies to take advantage of these same host stress signaling pathways. A prominent example of the latter is the cellular heat shock response (HSR), which is responsible for maintaining proteostasis in the cytosol and nucleus.[8] The HSR is controlled by its master regulator, the heat shock factor 1 (HSF1) transcription factor. High levels of HSF1 activity can be triggered by a variety of stressors, including protein misfolding in the cytosol. HSF1-mediated upregulation of numerous heat shock protein (HSP) chaperones and quality control proteins serves to restore proteostasis, after which HSF1 activity is reduced to basal levels.[9] Many host chaperones, including Hsp70 and Hsp90, are hijacked by diverse viruses to assist viral protein folding and thereby enable virion production.[10-13] Inhibition of these same chaperones can suppress viral replication.[14-18] Moreover, chaperones can potentiate the evolution of viral proteins. Changes in cellular proteostasis capacity can modulate viral evolutionary trajectories,[19-21] and even define the accessibility of destabilized viral protein variants that can enable innate immune system escape.[22]

Hence, host HSF1 activity and the functions of HSF1-regulated host chaperones are often beneficial for viruses.[10-13, 19-21] However, this conclusion derives largely from studies on just a few viruses, including influenza, Dengue, Zika, and polio – with limited studies on retroviruses. Similar phenomena might be expected for retroviruses, which also have high mutation rates and a need to fold their proteins. On the other hand, the requirement for host genome integration in

94

particular adds an additional step that could be differentially influenced by HSF1 and other HSPs.

Prior work has suggested an intimate role for the host cell's HSR in multiple steps of the HIV-1 lifecycle. The complexity of HSF1 engagement during HIV-1 replication is perhaps best illustrated by HSF1's apparent ability to either assist[12, 23-26] or restrict[27-28] HIV-1 propagation depending on the method used to perturb HSF1 activity. For example, heat stress stimulates HIV-1 gene transcription[23] and viral replication.[25-26] In other work, transient overexpression of wild-type HSF1 assisted HIV-1 generation[24] and reactivation from latency,[12] while HSF1 knockdown proved deleterious for HIV-1 production. Alternatively, transient overexpression of a constitutively active variant of HSF1 suppressed long terminal repeat (LTR)-driven viral transcription[27] and downregulated HIV-induced inflammation.[28] Similarly, the reported roles of individual HSF1-controlled chaperones in HIV-1 replication extensively vary between different experimental systems.[29-36] In sum, although the details are still unclear, there is clearly a complicated interplay between the host's HSR and the HIV-1 lifecycle.

Our objective was to isolate the direct effects of HSF1 activation from the indirect effects of the cellular stressors that are traditionally used to activate HSF1, thereby gaining a clear understanding of the consequences of HSF1 activity for HIV-1 replication. The achievement of this goal requires a tool for stress-independent HSF1 activation. Heat induction of HSF1 activity is unsuitable because heat is a pleiotropic stress that causes acute and severe protein misfolding throughout the proteome. Genetic methods are preferred as they avoid HSR activation, however the extent of HSF1 activation is limited by cellular compensation mechanisms. For example, overexpression of wild-type HSF1 increases the protein levels of the transcription factor, but the excess HSF1 protein is subject to chaperone-mediated regulation and is thus kept in an inactive state.[37] Genetic HSF1 knockdown is also inefficient, owing to compensating proteostasis mechanisms.[38] Finally, unregulated overexpression of constitutively active HSF1 variants must be employed with great caution to avoid nonphysiologic levels of HSF1 induction and consequent pleiotropic remodeling of the transcriptome.[39] Chemical methods for directly regulating HSF1 activity are preferred.[40-42]
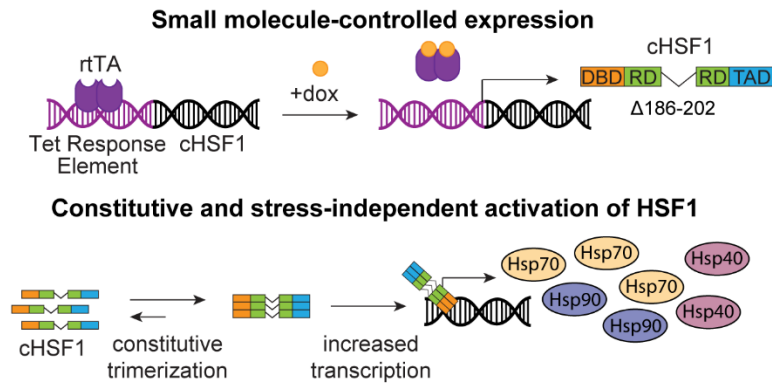
3.3   Results

We first sought to generate a system in which stress-independent, small molecule-mediated induction of HSF1 activity was possible. We engineered a stable, single-colony human T lymphocyte (CEM) cell line in which the expression of a constitutively active variant of HSF1 (cHSF1)[39, 43] was placed under control of the doxycycline (dox)-dependent tetracycline (tet) repressor.[39] Treatment of the resulting CEM^cHSF1 cell line (**Figure 3.1A**) with dox resulted in the expression of HSF1 target genes, as demonstrated by the increased transcript levels of HSP90, HSP70, and HSP40 (**Figure 3.1B**). The single colony cell line was carefully chosen to ensure that the upregulation of these downstream chaperones was similar in magnitude to that caused by HSF1 activation by the prototypical chemical stressor As(III)[44] (**Figure 3.1B**), ensuring that HSF1 activity was not induced beyond physiologically accessible levels.[39] We also generated a fluorescent control cell line (CEM^Ctrl) in which cyan fluorescent protein (CFP) expression was similarly placed under control of the tet repressor.

With CEM^cHSF1 and CEM^Ctrl cell lines in hand, we next sought to test whether stress-independent HSF1 activation impacted HIV-1 replication. We began by treating CEM^cHSF1 and CEM^Ctrl cells with dox for 18 h to activate cHSF1 or CFP expression, respectively. Next, we infected these preactivated cells with NL4-3 HIV-1 at a multiplicity of infection (MOI) of 0.04 for 96 h, followed by harvesting the infectious supernatant and titering using a p24 enzyme-linked immunosorbent assay (ELISA).

We observed that cHSF1 activation significantly reduced total p24 viral titers relative to cells with basal HSF1 activity (**Figure 3.1C**). In contrast, dox-induced expression of CFP in the CEM^Ctrl cell line did not alter p24 titers, showing that the result was attributable to cHSF1 activity and not to dox treatment. We further assessed infection kinetics by harvesting the viral supernatant at successive time points and titering using the p24 assay. The relative difference in p24 titers between cHSF1-activated versus vehicle-treated CEM^cHSF1 cells became more pronounced as the infection progressed, with no significant difference observed in dox- versus vehicle-treated CEM^Ctrl cells at any time point (**Figure 3.1D**). Finally, we used the TZM-bl assay[45] to quantify the infectious titers of collected viral supernatants. Successful infection of reporter TZM-bl cells activates the expression of β-galactosidase in an HIV-1 Tat-dependent manner, turning reporter cells blue in the presence of a 5-bromo-4-chloro-3-indolyl-*p*-D-galactopyranoside (X-Gal) chromogenic substrate. The fraction of stained cells is then proportional to the number of infectious viral particles.[45] We observed that, as also occurred with

Figure 3.1. Stress-independent cHSF1 activation decreases HIV-1 replication and the infectivity of produced virions.
(**A**) Chemical genetic tool for stress-independent, small molecule-regulated activation of HSF1. Treatment of CEM^cHSF1 cells with dox induces expression of cHSF1, which constitutively trimerizes and upregulates the expression of HSF1 target genes in the absence of acute proteostatic stress. (**B**) qPCR analysis of Hsp90 (*HSP90AA1*), Hsp70 (*HSPA1A*), and Hsp40 (*DNAJB1*) expression in CEM^cHSF1 after 18 h of treatment with 1 µg/mL dox or 2 h of treatment with 100 µM sodium arsenite as a positive heat shock control. (**C**) Fold-change in p24 titers after 96 h of HIV-1 infection at an MOI of 0.04 in CEM^cHSF1 and CEM^Ctrl cells, treated with 1 µg/mL dox, relative to vehicle-treated cells. (**D**) Schematic of a timecourse infection and total p24 viral titers during different infection time points in CEM^cHSF1 and CEM^Ctrl cells. (**E**) Fold-change in infectious TZM-bl titers after 96 h of HIV-1 propagation in CEM^cHSF1 and CEM^Ctrl cells, treated with 1 µg/mL dox, relative to vehicle-treated cells. *, **, ***, and ns correspond to *p*-values <0.05, <0.001, <0.0001, and not significant, respectively.

the p24 titers, infectious titers were indeed decreased by cHSF1 activation in CEM^cHSF1 cells, whereas they did not change upon CFP activation in CEM^Ctrl cells (**Figure 3.1E**).

      The high mutation rate of HIV-1 often promotes rapid escape from inhibitory pressure. Therefore, we next asked whether continuous propagation of HIV-1 under pressure from cHSF1 activity would result in rapid antiviral escape. We performed three serial passages in cHSF1-activated versus vehicle-treated CEM^cHSF1 cells (**Figure 3.2A**). At each passage, the pre-activated cells were infected at an MOI of 0.04 for 96 h, followed by harvesting the viral supernatant and titering. The infectious (TZM-bl) titers were used to initiate the subsequent passage at the same MOI. Notably, both total and infectious viral titers were still decreased in +cHSF1 cells relative to vehicle-treated cells even after the third serial passage (**Figures 3.2B** and **C**), indicating that the virus cannot readily adapt to cHSF1-mediated replication restriction.
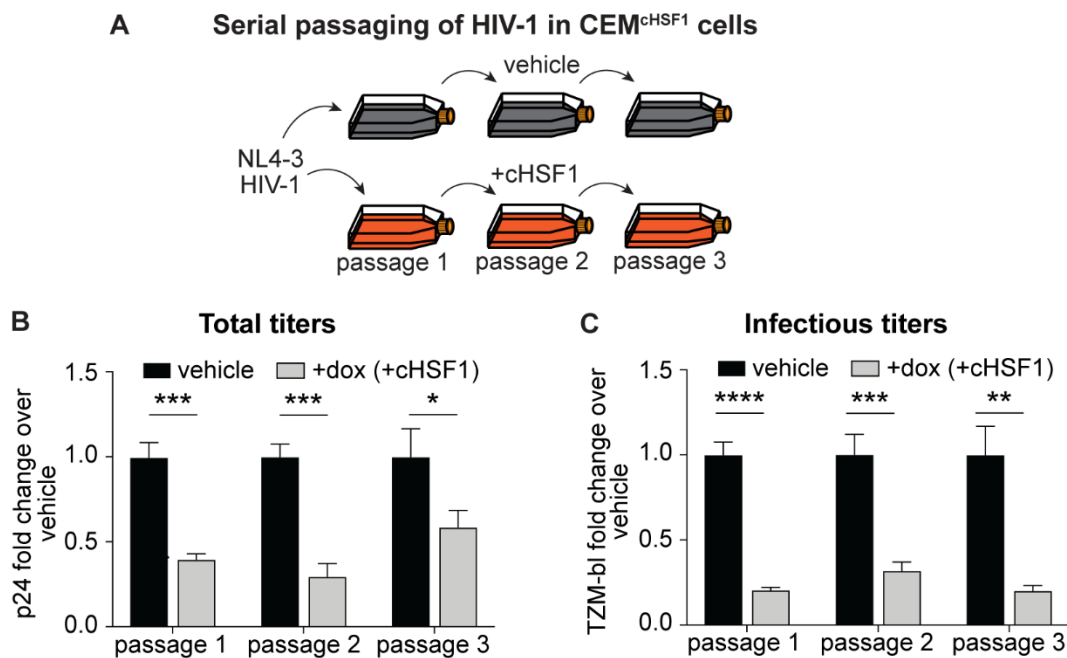
Figure 3.2. HIV-1 does not adapt to escape HSF1 activation over the course of three serial passages.
(**A**) Schematic of NL4-3 HIV-1 serial passaging. CEM^cHSF1 cells were pretreated with 1 µg/mL dox for 18 h prior to infection with NL4-3 HIV-1 for 96 h at an MOI of 0.04. Infectious titers of the viral supernatant were determined using the TZM-bl assay, and then, the supernatant was used to infect the subsequent passage at the same MOI. (**B**) Fold-change in total p24 and (**C**) infectious TZM-bl titers at each passage. *, **, ***, and **** correspond to p-values < 0.05, < 0.01, < 0.001, and < 0.0001, respectively.

One potential trivial explanation for HSF1's effect on HIV-1 replication could be cytotoxicity. We assessed cell health in the conditions of our viral propagation experiments by using a CellTiter-Glo assay. We observed that cellular ATP levels were not significantly altered by stress-independent cHSF1 activation (**Figure 3.3A**).
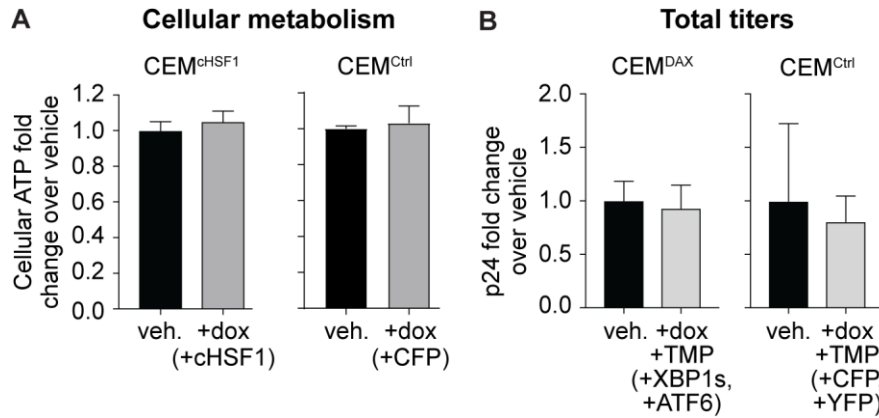
Figure 3.3. Decreased HIV-1 replication is not attributable to HSF1-induced cytotoxicity. (**A**) Cellular ATP levels upon 1 µg/mL dox treatment of CEM[cHSF1] and CEM[Ctrl] cells for 96 h, as assessed using a CellTiter-Glo assay. (**B**) Total p24 viral titers upon 1 µg/mL dox and 10 µM TMP treatment of CEM[DAX] and CEM[Ctrl] cell lines.

We next asked whether the observed inhibition of HIV-1 was specific to the HSR or could be replicated by stress-independent activation of other protein misfolding stress responses. We engineered a stable cell line, termed CEM[DAX], in which the IRE1-XBP1s and ATF6 arms of the unfolded protein response (UPR), which is responsible for maintaining proteostasis in the secretory pathway,[46] could be activated by small molecules in a stress-independent manner. Our approach was to render XBP1s expression dox-inducible by placing the *XBP1s* gene under control of the tetracycline promoter.[47] To control the ATF6 arm of the UPR, we fused the transcriptionally active form of ATF6 to an *Escherichia coli* dihydrofolate reductase (DHFR) destabilized domain.[48] Treatment of CEM[DAX] cells with trimethoprim (TMP) stabilizes the DHFR domain, resulting in ATF6 transcriptional activity. This strategy is well-established for stress-independent control of the IRE1-XBP1s and ATF6 arms of the unfolded protein response.[21, 41, 49-54] We verified the selective, dox-dependent induction of XBP1s target genes and the selective, TMP-dependent induction of ATF6 target genes in CEM[DAX] cells using qPCR (**Figure 3.4**).[49] We also employed a fluorescent control CEM[Ctrl] cell line stably engineered to express dox-inducible CFP and *E. coli* DHFR-fused yellow fluorescent protein (YFP). We then pretreated CEM[DAX] and CEM[Ctrl] cells with dox and TMP for 18 h to activate the
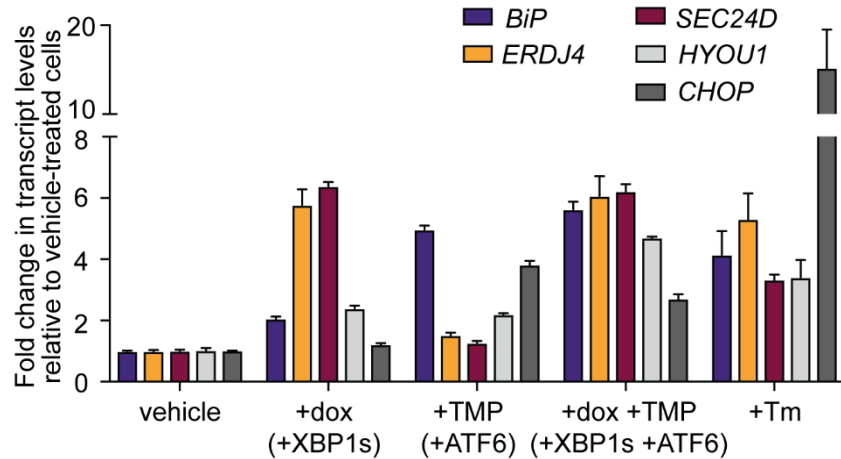
Figure 3.4. Selective induction of XBP1s and/or ATF6 target genes in CEMDAX cells. qPCR analysis of the unfolded protein response genes (*BiP*, *ERDJ4*, *SEC24D*, *HYOU1*, *CHOP*) expression levels upon tet-XBP1s activation with 1 µg/mL doxycycline for 18 h, DHFR-ATF6 activation with 10 µM trimethoprim for 18 h, or both, and 10 µg/mL tunicamycin (Tm) treatment as a positive control for stress-mediated unfolded protein response activation.

corresponding constructs, infected the cells with HIV-1 at an MOI of 0.04, and measured the resulting p24 titers after 96 h. No significant change in p24 titers was observed upon dox and TMP treatment in either the CEM^DAX or the CEM^Ctrl cells (**Figure 3.3B**). Thus, HSF1-mediated abrogation of HIV-1 replication is a specific feature of HSR activation, not a general consequence of inducing protein misfolding stress responses.

Next, we used RNA-Seq to globally assess transcriptome remodeling owing to cHSF1 activation in CEM^cHSF1 cells. In particular, we were interested in whether or not stress-independent cHSF1 induction might elicit an antiviral response in CEM^cHSF1 cells. As expected, given the specificity of our stress-independent, chemical induction of cHSF1, the most prominent upregulated genes were all well-known components of the HSR (**Figure 3.5**).[39] Also, as expected, no significant induction of UPR target genes was observed (**Figure 3.5A**).

We applied gene set enrichment analysis (GSEA)[55] (**Tables 3.1** and **3.2**) to better understand key features of the transcriptome remodeling caused by cHSF1 activation. We observed that known HSR-related gene sets were massively enriched (MSigDB c5 collection; **Figure 3.6A**, **Tables 3.1** and **3.2**). Furthermore, the HSF1-binding motif itself was strongly enriched upstream of genes that were found to be responsive to stress-independent cHSF1 activation (MSigDB c3 collection, **Figure 3.7**). However, we did not observe any significant enrichment of antiviral restriction factors using the MSigDB c5 collection (see **Figure 3.6B** for example enrichment plots, **Tables 3.1** and **3.2**). Similarly, when other functional databases

regrouped in the MSigDB c2 collections were interrogated, viral- and interferon-response pathways tended either not to display any bias or even to be enriched among downregulated gene sets (**Tables 3.3** and **3.4**).

Figure 3.5. Transcriptional profile of HSF1-activated host environment.
(A) Heat map showing the differential expression of select proteostasis genes upon stress-independent cHSF1 activation with 1 µg/mL dox for 18 h in CEM^cHSF1 cells relative to vehicle treatment. **(B)** Volcano plot showing distribution of expressed transcripts upon HSF1 activation with 1 µg/mL doxycycline for 18 h in CEM^cHSF1 cells. Red dots correspond to transcripts displaying ≥ 2-fold expression changes with p-values ≤ 10^{-5}. Genes with *p*-value of 0 were assigned an absolute maximum *p*-value of 10^{-300} for the purpose of display on the volcano plot.

Table 3.1. Gene set enrichment analysis results of upregulated genes for cHSF1 activation using MSigDB c5.
The top 15 upregulated gene groups are shown, complete table is available <u>online</u>.

Gene ontology group	NES	FDR q-val
GO_UNFOLDED_PROTEIN_BINDING	2.53	0.00E+00
GO_REGULATION_OF_CELLULAR_RESPONSE_TO_HEAT	2.48	0.00E+00
GO_CHAPERONE_MEDIATED_PROTEIN_FOLDING	2.45	0.00E+00
GO_CELLULAR_RESPONSE_TO_HEAT	2.45	0.00E+00
GO_PROTEIN_FOLDING	2.43	0.00E+00
GO_RESPONSE_TO_HEAT	2.41	0.00E+00
GO_HEAT_SHOCK_PROTEIN_BINDING	2.40	0.00E+00
GO_PROTEIN_REFOLDING	2.34	0.00E+00
GO_RESPONSE_TO_TEMPERATURE_STIMULUS	2.34	0.00E+00
GO_DE_NOVO_PROTEIN_FOLDING	2.32	0.00E+00
GO_CHAPERONE_COFACTOR_DEPENDENT_PROTEIN_REFOLDING	2.29	0.00E+00
GO_CHAPERONE_COMPLEX	2.24	1.86E-05
GO_ATPASE_REGULATOR_ACTIVITY	2.21	3.40E-05
GO_CHAPERONE_MEDIATED_PROTEIN_COMPLEX_ASSEMBLY	2.21	4.72E-05
GO_CHAPERONE_BINDING	2.18	1.47E-04

Table 3.2. Gene set enrichment analysis results of downregulated genes for cHSF1 activation using MSigDB c5.
The top 15 downregulated gene groups are shown, complete table is available <u>online</u>.

Gene ontology group	NES	FDR q-val
GO_STRUCTURAL_CONSTITUENT_OF_RIBOSOME	-2.91	0.00E+00
GO_RIBOSOMAL_SUBUNIT	-2.82	0.00E+00
GO_COTRANSLATIONAL_PROTEIN_TARGETING_TO_MEMBRANE	-2.81	0.00E+00
GO_ESTABLISHMENT_OF_PROTEIN_LOCALIZATION_TO_ENDOPLASMIC_RETICULUM	-2.78	0.00E+00
GO_CYTOSOLIC_RIBOSOME	-2.72	0.00E+00
GO_CYTOSOLIC_LARGE_RIBOSOMAL_SUBUNIT	-2.71	0.00E+00
GO_NUCLEAR_TRANSCRIBED_MRNA_CATABOLIC_PROCESS_NONSENSE_MEDIATED_DECAY	-2.69	0.00E+00
GO_PROTEIN_LOCALIZATION_TO_ENDOPLASMIC_RETICULUM	-2.68	0.00E+00
GO_LARGE_RIBOSOMAL_SUBUNIT	-2.67	0.00E+00
GO_RIBOSOME	-2.55	0.00E+00
GO_SMALL_RIBOSOMAL_SUBUNIT	-2.35	1.44E-05
GO_ORGANELLAR_RIBOSOME	-2.28	1.32E-05
GO_CYTOSOLIC_SMALL_RIBOSOMAL_SUBUNIT	-2.21	6.07E-05
GO_KILLING_OF_CELLS_OF_OTHER_ORGANISM	-2.18	2.03E-04
GO_MITOCHONDRIAL_TRANSLATIONAL_TERMINATION	-2.16	3.16E-04

GSEA enrichment plots for cHSF1 activation in CEM^cHSF1 cells

A Heat shock response

Chaperone-mediated protein folding
p-value: 0.0000
NES: 2.4520

Unfolded protein binding
p-value: 0.0000
NES: 2.5343

Regulation of cellular response to heat
p-value: 0.0000
NES: 2.4798

B Antiviral response

Cellular response to virus
p-value: 0.1324
NES: -1.2339

Response to type I interferon
p-value: 0.0284
NES: 1.4028

Antigen processing and presentation of peptide antigen via MHC class I
p-value: 0.0135
NES: 1.4142

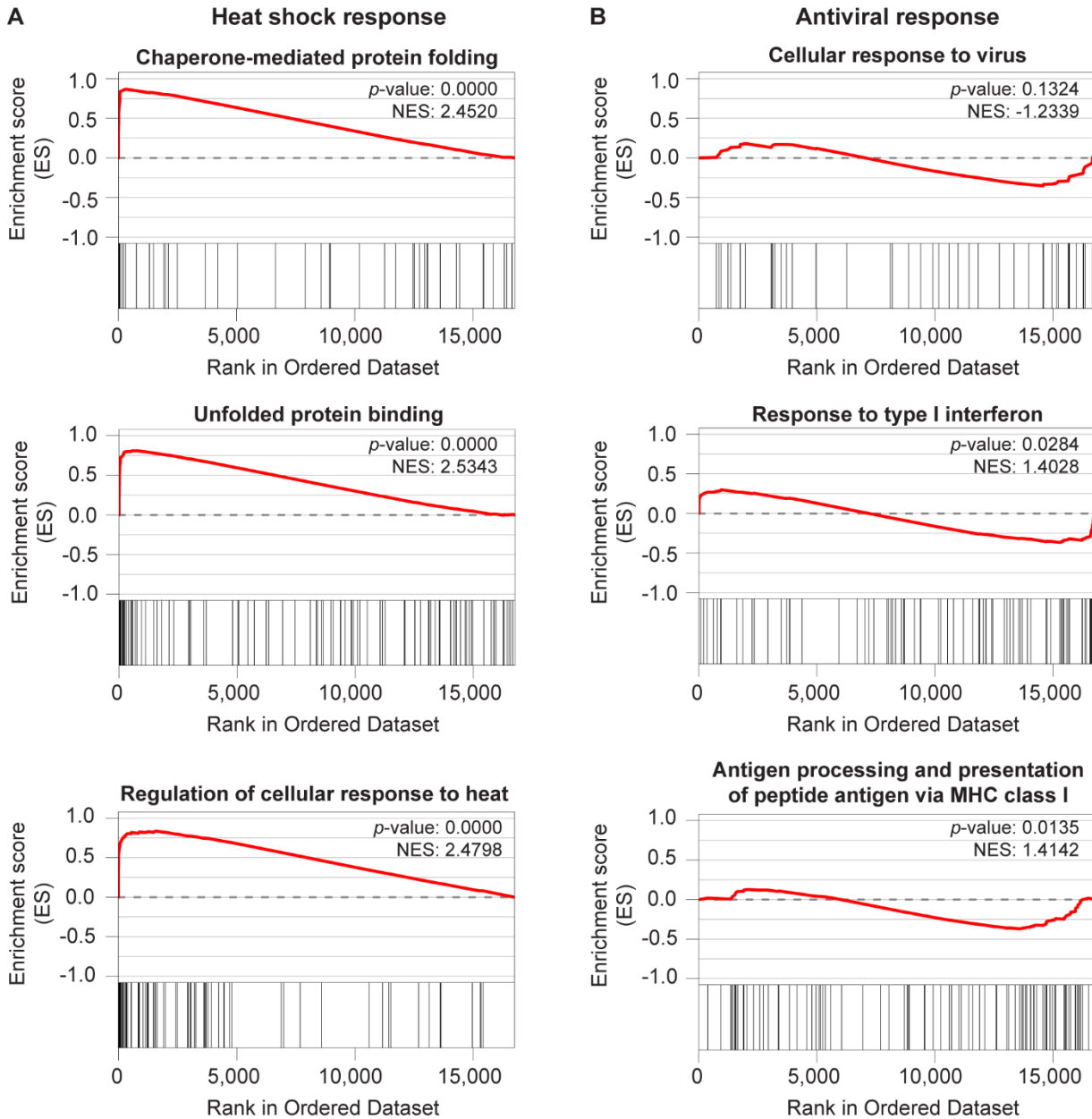**Figure 3.6. Stress-independent HSF1 induction activates heat shock response genes and does not trigger a broad-scale antiviral response.**
(A) Selected gene set enrichment plots for heat shock response-related and (**B**) antiviral response-related gene sets in CEM^cHSF1 cells treated with 1 μg/mL doxycycline for 18 h to induce cHSF1. These enrichment plots are drawn from the MSigDB c5 collection.

Table 3.3. Gene set enrichment analysis results of upregulated genes for cHSF1 activation using MSigDB c2.
The top 15 upregulated gene groups are shown, complete table is available online.

Gene group	NES	FDR q-val
REACTOME_CELLULAR_RESPONSE_TO_HEAT_STRESS	2.52	0.00E+00
REACTOME_HSP90_CHAPERONE_CYCLE_FOR_STEROID_HORMONE_RECEPTORS_SHR	2.46	0.00E+00
REACTOME_REGULATION_OF_HSF1_MEDIATED_HEAT_SHOCK_RESPONSE	2.45	0.00E+00
REACTOME_HSF1_DEPENDENT_TRANSACTIVATION	2.42	0.00E+00
REACTOME_HSF1_ACTIVATION	2.38	0.00E+00
REACTOME_ATTENUATION_PHASE	2.37	0.00E+00
YORDY_RECIPROCAL_REGULATION_BY_ETS1_AND_SP100_DN	2.34	0.00E+00
REACTOME_FORMATION_OF_TUBULIN_FOLDING_INTERMEDIATES_BY_CCT_TRIC	2.20	1.03E-04
GARGALOVIC_RESPONSE_TO_OXIDIZED_PHOSPHOLIPIDS _BLUE_UP	2.20	1.37E-04
KIM_WT1_TARGETS_8HR_UP	2.17	2.68E-04
REACTOME_COOPERATION_OF_PREFOLDIN_AND_TRIC_CCT_IN_ACTIN_AND_TUBULIN_FOLDING	2.15	5.80E-04
CHANDRAN_METASTASIS_TOP50_UP	2.13	7.03E-04
KEGG_ANTIGEN_PROCESSING_AND_PRESENTATION	2.12	1.06E-03
REACTOME_UPTAKE_AND_ACTIONS_OF_BACTERIAL_TOXINS	2.11	1.16E-03
REACTOME_SEMAPHORIN_INTERACTIONS	2.11	1.12E-03

Table 3.4. Gene set enrichment analysis results of downregulated genes for cHSF1 activation using MSigDB c2.
The top 15 downregulated gene groups are shown, complete table is available online.

Gene group	NES	FDR q-val
KEGG_RIBOSOME	-3.05	0.00E+00
REACTOME_SRP_DEPENDENT_COTRANSLATIONAL_PROTEIN_TARGETING_TO_MEMBRANE	-3.02	0.00E+00
REACTOME_SELENOAMINO_ACID_METABOLISM	-2.98	0.00E+00
REACTOME_NONSENSE_MEDIATED_DECAY_NMD_INDEPENDENT_OF_THE_EXON_JUNCTION_COMPLEX_EJC	-2.93	0.00E+00
REACTOME_NONSENSE_MEDIATED_DECAY_NMD	-2.83	0.00E+00
REACTOME_EUKARYOTIC_TRANSLATION_INITIATION	-2.79	0.00E+00
REACTOME_REGULATION_OF_EXPRESSION_OF_SLITS _AND_ROBOS	-2.65	0.00E+00
REACTOME_TRANSLATION	-2.61	0.00E+00
CHNG_MULTIPLE_MYELOMA_HYPERPLOID_UP	-2.54	0.00E+00
ELVIDGE_HIF1A_TARGETS_DN	-2.54	0.00E+00
BILANGES_SERUM_AND_RAPAMYCIN_SENSITIVE_GENES	-2.50	0.00E+00
MENSE_HYPOXIA_UP	-2.42	0.00E+00
MOOTHA_VOXPHOS	-2.35	1.48E-05
REACTOME_RRNA_PROCESSING_IN_THE_NUCLEUS_AND _CYTOSOL	-2.34	1.37E-05
REACTOME_RRNA_PROCESSING	-2.34	2.53E-05

These observations suggest that stress-independent HSF1 activation in CEM[cHSF1] cells does not inhibit HIV-1 replication by inducing a general antiviral response. We next examined individual genes within the broad gene ontology group "Defense Response to Virus" (**Table 3.5**). While typical components of the general antiviral defense response, including many interferon-related genes, were not enriched or even downregulated, we were intrigued to note that the most upregulated gene in the entire gene set was *ZC3HAV1*. *ZC3HAV1* encodes the zinc finger antiviral protein ZAP (also known as PARP13), and was upregulated 3.2-fold in our RNA-Seq

GSEA enrichment plot for HSF1 motif activation

Figure 3.7. Heat shock factor (HSF) motif is enriched upon stress-independent cHSF1 induction.
cHSF1 was induced with 1 µg/mL doxycycline for 18 h in CEM[cHSF1] cells. Gene set enrichment plot of HSF motif was generated using TTCNRGNNNNTTC_HSF_Q6 gene set from MSigDB c3 collection.

experiment upon cHSF1 induction. ZAP is known to restrict the replication of multiple viruses, particularly including HIV-1,[56-57] by targeting viral mRNA in the cytoplasm for degradation.[58] ZAP can also bind HSF1[59] and assist HSF1 binding to DNA prior to heat shock.[60] Indeed, the first intron of ZAP possesses an HSF1-binding motif, located in a putative chromatin regulatory region denoted by a peak of H3K27-acetylated histones, as reported by the Encyclopedia of DNA Elements (ENCODE) consortium in an immortalized B-cell line (chromatin immunoprecipitation (ChIP)-Seq ENCODE track on the UCSC Genome Browser).[61] We used qPCR to confirm that the induction of cHSF1 in CEM[cHSF1] cells indeed triggered upregulation of ZAP mRNA (**Figure 3.8**). On the basis of these observations, ZAP induction is likely to contribute to cHSF1-mediated inhibition of HIV-1 replication.

Table 3.5. Gene list for the "Defense Response to Virus" gene ontology group.
The top 15 genes are shown, complete table is available online.

Gene	Rank	ES	Gene	Rank	ES	Gene	Rank	ES
ZC3HAV1	44	0.03	PML	182	0.15	ILF3	648	0.19
GBP3	69	0.06	AGBL5	227	0.16	ANKRD17	741	0.19
BST2	79	0.09	RNF216	266	0.18	EIF2AK2	824	0.20
MICB	101	0.11	DHX9	488	0.18	PUM2	859	0.20
FLNA	113	0.14	MAVS	622	0.18	STAT2	913	0.21

3.4   Discussion

Although ZAP induction may play a role in the inhibition of HIV-1 replication, the key finding from our RNA-Seq analysis was that cHSF1 activation largely drives a transcriptional remodeling of the cellular chaperone network, with minimal impacts on immune responses and traditional viral restriction factors. A number of these chaperones have been implicated in the HIV lifecycle and play important roles in viral protein folding and assembly.[62-64] Thus, it is possible that the remodeled cytosolic and nuclear proteostasis network, which did not evolve to support HIV-1 replication but rather to ensure cellular proteostasis, might disrupt these steps in the lifecycle by diverting viral proteins from function or the orchestrated virion assembly process. In this regard, it is noteworthy that comparing the total (p24) to the infectious (TZM-bl) viral titers, we observed that the fraction of produced virions that are infection-competent significantly decreased upon cHSF1 activation (**Figure 3.9**). This observation is consistent with cHSF1 activation disrupting steps in the viral lifecycle such as viral protein folding and/or virion assembly that could lead to the production of a larger fraction of defective viral particles. Because host chaperones not only directly modulate viral protein folding and assembly but also participate in earlier steps of the viral replication cycle, such as nuclear import,[34] genome integration,[65] and transcription,[33, 36] we do not exclude the possibility of additional inhibitory roles of the cHSF1-remodeled proteostasis network in these processes.
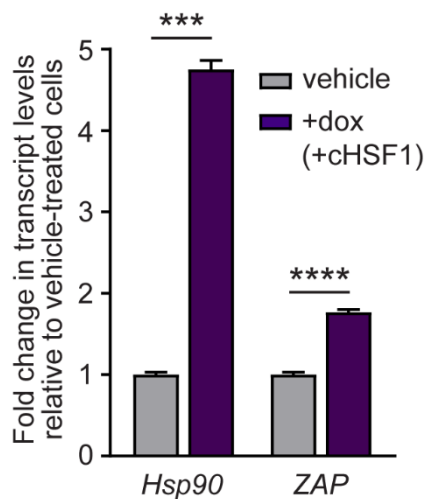
Figure 3.8. Induction of cHSF1 activates ZAP transcription.
PCR analysis of the classical heat shock response gene Hsp90 and ZAP expression levels upon treatment of CEM^cHSF1 cells with 1 µg/mL doxycycline for 18 h. *** and ****correspond to *p*-values <0.001 and <0.0001, respectively.
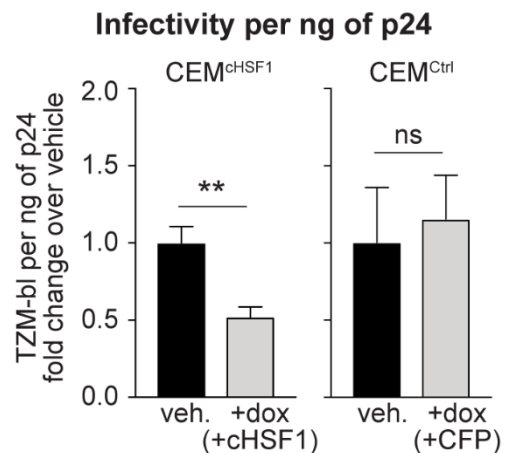
Figure 3.9. HSF1 reduces infectivity of newly produced virions.
Fold-change in infectious TZM-bl titers per ng of p24 after 96 h of HIV-1 infection at an MOI of 0.04 in CEM^cHSF1 and CEM^Ctrl cells treated with 1 µg/mL dox, relative to vehicle-treated cells. ** and ns correspond to *p*-values < 0.01 and not significant, respectively.

In summary, the use of a chemically controlled cHSF1 construct allowed us to investigate the direct consequences of HSF1 activation at physiologically relevant levels, eliminating the requirement for inducing global protein misfolding while also avoiding the off-target consequences of cHSF1 overexpression. We were also able to avoid the complications associated with transient overexpression of HSF1 or cHSF1,[39, 42] including off-target gene induction, which convoluted prior studies. Using this approach, we demonstrated that stress-independent HSF1 activation restricts HIV-1 replication in CEM cells. When cHSF1 was activated, fewer total HIV-1 virions were produced and the proportion of infectious virions was also lowered. Moreover, cHSF1-mediated inhibition of HIV-1 replication persisted through three consecutive serial passages without detectable recovery of viral titers, suggesting that escape mechanisms are not readily available to the virus. The effects of cHSF1 activation were HSR-specific and not attributable to reductions in host cell health, off-target cHSF1 activity, or activation of protein misfolding stress responses in general.

The exact molecular mechanisms of HSF1-mediated restriction of HIV-1 replication remain an important subject for further study and are likely multifactorial. First, viral transcripts are known to be targeted to degradation by the HSF1-controlled host restriction factor ZAP, which has an HSF1-binding promoter and was transcriptionally upregulated in our system despite the absence of a general antiviral response induced by cHSF1. Second, cHSF1 activation reduces the infectivity of newly formed virions. This observation suggests that the remodeled host chaperone network promotes the formation of defective viral particles. While deciphering the origins of HSF1-mediated inhibition of HIV-1 replication and elucidating *in vivo* relevance of these findings requires future investigation, this work clearly implicates HSF1 as a host antiviral restriction factor for HIV-1 and motivates continued consideration of host HSR-targeted therapeutics to address retroviral infections.

3.5 Materials and Methods

Plasmids. The following lentiviral destination vectors were used for stable cell line construction: pLenti CMV/TO zeo DEST with either human cHSF1[41] or CFP inserts (Addgene), pLenti6/V5 DEST Gateway with a tetracycline repressor insert (Invitrogen), pLenti CMV puro DEST (Invitrogen) with a DHFR.YFP fusion insert, and previously described DHFR.ATF6(1–373)- and XBP1s-encoding pLenti vectors.[49]

Cell culture. Human T lymphocytes (CEM) were grown in RPMI-1640 medium (Corning) supplemented with 10% heat-inactivated fetal bovine serum (FBS; CellGro), 1% penicillin/streptomycin/glutamine (CellGro) at 37 °C with 5% CO_2(g). TZM-bl reporter cells were obtained from the NIH AIDS Reagent Program. TZM-bl cells were cultured in DMEM (Corning) supplemented with 10% heat-inactivated fetal bovine serum (Cellgro), 1% penicillin/streptomycin/glutamine (Cellgro) at 37 °C with 5% CO_2(g).

Stable cell line construction. For the CEM[cHSF1] and CEM[Ctrl] cell lines construction, CEM cells were transduced first with lentivirus encoding a blasticidin-resistant tetracycline repressor and then with lentiviruses encoding zeocin-resistant cHSF1 or CFP constructs, respectively. Transduction was accomplished by spinoculation with 2 µg/mL polybrene (Sigma-Aldrich) at 1,240 × *g* for 1–1.5 h. Heterostable cell lines expressing the tetracycline repressor and cHSF1 or CFP were then selected using 10 µg/mL blasticidin (Gibco) and 50 µg/mL zeocin (Invitrogen). Single colonies of CEM[cHSF1] and CEM[Ctrl] cells were generated from the heterostable population by seeding 30–40 cells in a 96-well plate in 100 µl of RPMI media without antibiotics for 10–14 days. Clonal populations were then selected and expanded in 24-well plates in 500 µL of RPMI containing 10 µg/mL blasticidin and 50 µg/mL zeocin. Cells were grown to confluency and then screened based on functional testing of the cHSF1 construct using RT-PCR for CEM[cHSF1] cells (described below) and CFP fluorescence for the CEM[Ctrl] cell line with or without 1 µg/mL doxycycline (dox; Alfa Aesar). CEM[Ctrl] cell lines were also engineered to express DHFR.YFP by transduction with lentivirus encoding DHFR.YFP (selection using puromycin at 8 µg/mL and single colony selection as above upon visual inspection of 10 µM trimethoprim (TMP; Alfa Aesar)-treated cells). For the CEM[DAX] cell line construction, CEM cells were transduced first with lentivirus encoding a blasticidin-resistant tetracycline repressor, then with lentiviruses encoding geneticin-resistant hXBP1s and zeocin-resistant DHFR.ATF6, following the protocol described above. Single colonies were selected and expanded in RPMI with 10 µg/mL blasticidin, 500

µg/mL geneticin sulfate (G418, Enzo Life Sciences) and 50 µg/mL zeocin. The cell lines were characterized by CellTiter-Glo viability assay, RT-PCR, and RNA-Seq, as described below.

Quantitative RT-PCR. CEM^{cHSF1} and CEM^{DAX} cells were seeded at a density of 2×10^5 cells/well in a 6-well plate. The cells were treated with 0.1 % DMSO, 1 µg/mL doxycycline, 10 µM TMP, or both (as indicated) for 18 h, 100 µM sodium arsenite (Alfa Aesar) for 2 h for the heat shock activation control, or 10 µg/mL tunicamycin (Sigma-Aldrich) for 6 h for the UPR activation control. Each treatment was performed in biological triplicate. Cellular RNA was extracted using the E.Z.N.A Total RNA Kit (Omega). 1 µg RNA was reverse transcribed into cDNA using the High-Capacity Reverse Transcription kit (Applied Biosystems). The cDNA reaction was then diluted to 80 µL with molecular biology-grade water and 2 µL of each sample was used for RT-PCR with the 2x Sybr Green Reaction Mix (Roche). To assess heat shock response (HSR) activation, primers for human *RPLP2* (housekeeping gene), *HSP90AA1*, *HSPA1A*, *DNAJB1* genes were used (**Table 3.6**). To assess unfolded protein response (UPR) activation, primers for human *GAPDH* (housekeeping gene), *BIP*, *SEC24D*, *ERDJ4*, *HYOU1*, *CHOP* were used (**Table 3.6**). Transcript levels were first normalized to *RPLP2* levels for HSR activation and to *GAPDH* for UPR activation for every sample, and then normalized for drug-treated versus vehicle-treated cells.

Table 3.6. RT-PCR primer sequences.

Gene	Forward primer sequence	Reverse primer sequence
RPLP2	CGTCGCCTCCTACCTGCT	CCATTCAGCTCACTGATAACCTT
HSP90AA1	GATAAACCCTGACCATTCC	AAGACAGGAGCGCAGTTTCATAAA
HSPA1A	GGAGGCGGAGAAGTACA	GCTGATGATGGGGGTTACA
DNAJB1	TGTGTGGCTGCACAGTGAAC	ACGTTTCTCGGGTGTTTTGG
GAPDH	TGGAAGGACTCATGACCACA	AGGGGTCTACATGGCAACTG
BIP	GCCTGTATTTCTAGACCTGCC	TTCATCTTGCCAGCCAGTTG
ERDJ4	CTGTATGCTGATTGGTAGAGTCAA	AGTAGACAAAGGCATCATTTCCAA
SEC24D	AGCAGACTGTCCTGGGAAGC	TTTGTTTGGGGCTGGAAAAG
HYOU1	GCAGACCTGTTGGCACTGAG	TCACGATCACCGGTGTTTTC
CHOP	GGAGCTGGAAGCCTGGTATG	GCCAGAGAAGCAGGGTCAAG
ZAP	GCACTTGTTAACGATTCTTTATCTG	AGCGGACAACCCTTACACAG

HIV-1 infection. CEM^{cHSF1} and CEM^{Ctrl} cells were seeded in T75 vented tissue culture flasks (Corning) at a density of 2.5×10^7 cells/flask in 15 mL of RPMI media. The cells were pre-treated with 1 µg/mL dox for 18 h and infected with NL4-3 virus at a multiplicity of infection (MOI) of 0.04, based on the infectious (TZM-bl) viral titers. To remove unbound virions from

culture after 12 h of infection, the cells were pelleted at 2,000 × g for 5 min, washed with 25 mL of PBS, and resuspended in 50 mL of RPMI media supplemented with 1 µg/mL dox. After 4 days. the viral supernatant was harvested, clarified at 2,000 × g for 5 min to remove cell debris, and stored at –80 °C. The supernatant was titered using p24 and TZM-bl assays. The infectious TZM-bl titers were used to initiate the subsequent serial passaging infection at an MOI of 0.04. For time-course of infection studies, the cells were plated in 24-well plates at density 1.5×10^5 cells/well in 1.5 mL of RPMI media with and without 1 µg/mL doxycycline. After 18 h, the cells were infected with NL4-3 virus at an MOI of 0.05. The infectious supernatant was harvested at 48, 72 and 96 hours post infection and stored at –80 °C prior titering using the p24 assay. CEM^{DAX} cells were seeded at a density of 5×10^5 cells/well in a 24-well plate and pre-treated with 1 µg/mL dox and 10 µM TMP for 18 h. The cells were infected with NL4-3 virus at an MOI of 0.04 as described above and viral supernatant was titered using the p24 assay.

p24 assay. ELISA plates were prepared by coating 96-well plates with 100 µL/well of 0.02 mg/mL anti-HIV-1 Gag antibody (Aalto Bio Reagents Ltd) in 1 M NaHCO$_3$ (pH 8.5). After 3 h incubation at rt, plates were washed twice with 200 µL of Tris-buffered saline (TBS; VWR) and blocked with 250 µL of 2% milk in TBS overnight at rt. The plates were stored at –20 °C and washed with 1× phosphate-buffered saline (PBS) twice prior to use. Cell-free viral supernatants were assayed for HIV-1 Gag p24 levels by ELISA. Virus was lysed with an effective concentration of 1% Empigen detergent (Sigma) in a 96-well plate for at least 1 h. After 1 h, the lysate was thoroughly mixed and diluted 500–1000 fold with 0.05% Empigen in TBS. p24 standards were prepared by a serial 4-fold dilution of 40 ng/µL recombinant HIV-1 Gag (Aalto Bio Reagents Ltd) in 0.05% Empigen in TBS. The supernatant from uninfected cells was used as a negative control. First, 100 µL of each diluted lysate, p24 standard dilutions, and a negative control were added to ELISA plates in duplicate and incubated for 3 h at rt. Next, the plate was washed twice with TBS, and then 100 µL/well of 1:10,000 diluted mouse monoclonal anti-HIV-1-p24 antibody conjugated to alkaline phosphatase (Aalto Bio Reagents Ltd) was added and incubated for 1 h at rt. The plate was then washed 4× with 200 µL of 0.1% Tween in PBS and twice with Tropix Assay Buffer (Applied Biosystems). Next, the plate was incubated with 50 µL of ([3-(1-chloro-3'-methoxyspiro[adamantane-4,4'-dioxetane]-3'-yl)phenyl] dihydrogen phosphate) (CSPD) substrate with Sapphire II enhancer (Invitrogen) for 30 min. Luminescence was quantified using a Take-3 plate reader (BioTek), monitoring each well for 2 sec.

TZM-bl assay. TZM-bl reporter cells were seeded at a density of 2.5×10^4 cells/well in 48-well plates. After 5 h, the cells were infected with 100 µL of serially diluted infectious supernatant containing 10 µg/ml polybrene. Each sample was used to infect four technical replicates. After 2 d, the viral supernatant was removed, the cells were washed twice with PBS, and then fixed with 4% paraformaldehyde (Thermo Scientific) for 20 min. After fixing, the cells were washed twice with PBS and stained with 4 mM potassium ferrocyanide, 4 mM ferricyanide, and 0.4 mg/mL 5-bromo-4-chloro-3-indolyl-*p*-D-galactopyranoside (X-Gal) in PBS at 37 °C for 50 min. The cells were washed with PBS, blue cells were counted manually under a microscope, and infectious titers were calculated as the number of blue cells per volume of viral inoculum.

CellTiter-Glo viability assay. CEM^{cHSF1} and CEM^{Ctrl} cells were seeded in biological triplicate at a density of 10^3 cells/well in a 96-well plate and pre-treated with 1 µg/mL dox or water (vehicle) for 96 h. The CellTiter-Glo (Promega) assay was then performed according to the manufacturer's instructions. Luminescence was recorded using a Take-3 plate reader (BioTek). The raw fluorescence units for all treatments were normalized to vehicle control.

RNA-Seq. CEM^{HSF1} or CEM^{Ctrl} cells were seeded at a density of 2×10^5 cells/well in a 24-well plate in quadruplicate. The cells were treated with either vehicle or 1 µg/mL dox for 18 h. Cellular RNA was harvested using the RNeasy Plus Mini Kit with QIAshredder homogenization columns (Qiagen). RNA-Seq libraries were prepared using the Kapa mRNA HyperPrep RNA-Seq library construction kit (Kapa/Roche), with 7 min fragmentation at 94 °C and 15 PCR cycles of final amplification and duplex barcoding. Libraries were quantified using the Fragment Analyzer and qPCR before being sequenced on an Illumina HiSeq 2000 using 40-bp single-end reads in High Output mode. Reads were aligned against hg19 (Feb., 2009) using bwa mem v. 0.7.12-r1039 [RRID:SCR_010910] with flags –t 16 –f and mapping rates, fraction of multiply-mapping reads, number of unique 20-mers at the 5´ end of the reads, insert size distributions and fraction of ribosomal RNAs were calculated using bedtools v. 2.25.0 [RRID:SCR_006646].[66] In addition, each resulting bam file was randomly down-sampled to a million reads, which were aligned against hg19 and read density across genomic features were estimated for RNA-Seq-specific quality control metrics. For mapping and quantitation, reads were aligned against GRCh38/ENSEMBL 89 annotation using STAR v. 2.5.3a with the following flags -runThreadN 8 –runMode alignReads –outFilter-Type BySJout –outFilterMultimapNmax 20 –alignSJoverhangMin 8 –alignSJDBoverhangMin 1 –outFilterMismatchNmax 999 –alignIntronMin 10 –alignIntronMax 1000000 –alignMatesGapMax1000000 –outSAMtype BAM

SortedByCoordinate –quantMode TranscriptomeSAM with –genomeDir ointing to a 75nt-junction GRCh38 STAR suffix array.[67] Gene expression was quantitated using RSEM v. 1.3.0 [RRID:SCR_013027] with the following flags for all libraries: rsem-calculate-expression –calc-pme –alignments -p 8 –forward-prob 0 against an annotation matching the STAR SA reference.[68] Posterior mean estimates (pme) of counts and estimated RPKM were retrieved. For differential expression analysis, dox-treated CEM^cHSF1 cells were compared against vehicle-treated CEM^cHSF1 samples. Briefly, differential expression was analyzed in the R statistical environment (R v.3.4.0) using Bioconductor's DESeq2 package on the protein-coding genes only [RRID:SCR_000154].[69] Dataset parameters were estimated using the estimateSizeFactors(), and estimateDispersions() functions; read counts across conditions were modeled based on a negative binomial distribution, and a Wald test was used to test for differential expression (nbinomWaldtest(), all packaged into the DESeq() function), using the treatment type as a contrast. Shrunken log2 fold-changes were calculated using the lfcShrink function. Fold-changes and p-values were reported for each protein-coding gene. Gene ontology analyses were performed using the online DAVID server.[70] Heat maps for select genes were generated in GraphPad Prism version 7.0b (**Figure 3.5A**).

Gene set enrichment analysis (GSEA). Differential expression results from DESeq2 were retrieved, and the "stat" column was used to pre-rank genes for GSEA analysis. Briefly, these "stat" values reflect the Wald's test performed on read counts as modeled by DESeq2 using the negative binomial distribution. Genes that were not expressed were excluded from the analysis. Alternatively, gene loadings from PCA analysis were used as ranking metrics. GSEA (desktop version, v3.0)[55, 71] was run in the pre-ranked mode against MSigDB 7.0 Hallmark, C2 (curated gene sets), C5 (Gene Ontology), C6 (oncogenic signatures) and C7 (immunologic signatures) sets, using the official gene symbol as the key, with a weighted scoring scheme, normalizing by meandiv, with gene sets between 5 and 2000 genes (5379 gene sets retained for C2, 830 for C3, 9373 for C5, 189 for C6 and 4872 for C7), and 5000 permutations were run for *p*-value estimation. GSEA enrichment plots were replotted using a modified version of the ReplotGSEA.R script (https://github.com/PeeperLab/Rtoolbox/blob/master/R/ReplotGSEA.R).

Statistical analyses. All experiments were performed in biological triplicate with the data presented as mean ± standard deviation. Statistical significance was assessed using unpaired *t*-tests with Welch's correction in GraphPad Prism version 7.0b.

3.6   Acknowledgements

3.7 References

1.   World Health Organization (accessed June 3, 2019) Summary of the global HIV epidemic. **2017**, https://www.who.int/hiv/data/2017_summary-global-hiv-epidemic.png?ua=1

2.   Cihlar, T.; Fordyce, M. Current status and prospects of HIV treatment. *Curr. Opin. Virol.* **2016,** *18*, 50-56.

3.   Jean, M. J.; Fiches, G.; Hayashi, T.; Zhu, J. Current strategies for elimination of HIV-1 latent reservoirs using chemical compounds targeting host and viral factors. *AIDS Res. Hum. Retroviruses.* **2019,** *35* (1), 1-24.

4.   Park, R. J.; Wang, T.; Koundakjian, D.; Hultquist, J. F.; Lamothe-Molina, P.; Monel, B.; Schumann, K.; Yu, H.; Krupzcak, K. M.; Garcia-Beltran, W. et al. A genome-wide CRISPR screen identifies a restricted set of HIV host dependency factors. *Nat. Genet.* **2017,** *49* (2), 193-203.

5.   Brass, A. L.; Dykxhoorn, D. M.; Benita, Y.; Yan, N.; Engelman, A.; Xavier, R. J.; Lieberman, J.; Elledge, S. J. Identification of host proteins required for HIV infection through a functional genomic screen. *Science* **2008,** *319* (5865), 921-926.

6.   Jäger, S.; Cimermancic, P.; Gulbahce, N.; Johnson, J. R.; McGovern, K. E.; Clarke, S. C.; Shales, M.; Mercenne, G.; Pache, L.; Li, K.et al. Global landscape of HIV–human protein complexes. *Nature* **2012,** *481* (7381), 365.

7.   Arhel, N.; Kirchhoff, F. Host proteins involved in HIV infection: new therapeutic targets. *Biochim Biophys. Acta, Mol. Basis Dis.* **2010,** *1802* (3), 313-321.

8.   Aviner, R.; Frydman, J. Proteostasis in viral infection: unfolding the complex virus–chaperone interplay. *Cold Spring Harbor Perspect. Biol.* **2020**, a034090.

9.   Richter, K.; Haslbeck, M.; Buchner, J. The heat shock response: life on the verge of death. *Mol. Cell* **2010,** *40* (2), 253-266.

10.   Filone, C. M.; Caballero, I. S.; Dower, K.; Mendillo, M. L.; Cowley, G. S.; Santagata, S.; Rozelle, D. K.; Yen, J.; Rubins, K. H.; Hacohen, N. et al. The master regulator of the cellular stress response (HSF1) is critical for orthopoxvirus infection. *PLoS Pathog.* **2014,** *10* (2), e1003904.

11.   Tsai, T.-T.; Chen, C.-L.; Tsai, C.-C.; Lin, C.-F. Targeting heat shock factor 1 as an antiviral strategy against dengue virus replication in vitro and in vivo. *Antiviral Res.* **2017,** *145*, 44-53.

12.   Pan, X.-Y.; Zhao, W.; Zeng, X.-Y.; Lin, J.; Li, M.-M.; Shen, X.-T.; Liu, S.-W. Heat shock factor 1 mediates latent HIV reactivation. *Sci. Rep.* **2016,** *6*, 26294.

13.   Wang, F.-W.; Wu, X.-R.; Liu, W.-J.; Liao, Y.-J.; Lin, S.; Zong, Y.-S.; Zeng, M.-S.; Zeng, Y.-X.; Mai, S.-J.; Xie, D., Heat shock factor 1 upregulates transcription of Epstein–Barr Virus nuclear antigen 1 by binding to a heat shock element within the BamHI-Q promoter. *Virology* **2011,** *421* (2), 184-191.

14.   Geller, R.; Taguwa, S.; Frydman, J. Broad action of Hsp90 as a host chaperone required for viral replication. *Biochim. Biophys. Acta, Mol. Cell Res.* **2012,** *1823* (3), 698-706.

15.   Joshi, P.; Sloan, B.; Torbett, B. E.; Stoddart, C. A. Heat shock protein 90AB1 and hyperthermia rescue infectivity of HIV with defective cores. *Virology* **2013,** *436* (1), 162-172.

16.   Heaton, N. S.; Moshkina, N.; Fenouil, R.; Gardner, T. J.; Aguirre, S.; Shah, P. S.; Zhao, N.; Manganaro, L.; Hultquist, J. F.; Noel, J. et al. Targeting viral proteostasis limits influenza virus, HIV, and dengue virus infection. *Immunity* **2016,** *44* (1), 46-58.

17.   Taguwa, S.; Maringer, K.; Li, X.; Bernal-Rubio, D.; Rauch, J. N.; Gestwicki, J. E.; Andino, R.; Fernandez-Sesma, A.; Frydman, J. Defining Hsp70 subnetworks in dengue virus replication reveals key vulnerability in flavivirus infection. *Cell* **2015,** *163* (5), 1108-1123.

18.   Taguwa, S.; Yeh, M.-T.; Rainbolt, T. K.; Nayak, A.; Shao, H.; Gestwicki, J. E.; Andino, R.; Frydman, J. Zika virus dependence on host Hsp70 provides a protective strategy against infection and disease. *Cell Rep.* **2019,** *26* (4), 906-920. e3.

19.   Phillips, A. M.; Gonzalez, L. O.; Nekongo, E. E.; Ponomarenko, A. I.; McHugh, S. M.; Butty, V. L.; Levine, S. S.; Lin, Y.-S.; Mirny, L. A.; Shoulders, M. D. Host proteostasis modulates influenza evolution. *eLife* **2017,** *6,* e28652.

20.   Geller, R.; Pechmann, S.; Acevedo, A.; Andino, R.; Frydman, J. Hsp90 shapes protein and RNA evolution to balance trade-offs between protein stability and aggregation. *Nat. Commun.* **2018,** *9* (1), 1781.

21.   Phillips, A. M.; Doud, M. B.; Gonzalez, L. O.; Butty, V. L.; Lin, Y.-S.; Bloom, J. D.; Shoulders, M. D. Enhanced ER proteostasis and temperature differentially impact the mutational tolerance of influenza hemagglutinin. *eLife* **2018,** *7,* e38795.

22.   Phillips, A. M.; Ponomarenko, A. I.; Chen, K.; Ashenberg, O.; Miao, J.; McHugh, S. M.; Butty, V. L.; Whittaker, C. A.; Moore, C. L.; Bloom, J. D. et al. Destabilized adaptive influenza variants critical for innate immune system escape are potentiated by host chaperones. *PLoS Biol.* **2018,** *16* (9), e3000008.

23.   Kretz-Remy, C.; Munsch, B.; Arrigo, A.-P. NFκB-dependent transcriptional activation during heat shock recovery. Thermolability of the NFκB.IκB complex. *J. Biol. Chem.* **2001,** *276* (47), 43723-43733.

24.   Rawat, P.; Mitra, D. Cellular heat shock factor 1 positively regulates human immunodeficiency virus-1 gene expression and replication by two distinct pathways. *Nucleic Acids Res.* **2011,** *39* (14), 5879-5892.

25.   Roesch, F.; Meziane, O.; Kula, A.; Nisole, S.; Porrot, F.; Anderson, I.; Mammano, F.; Fassati, A.; Marcello, A.; Benkirane, M. et al. Hyperthermia stimulates HIV-1 replication. *PLoS Pathog.* **2012,** *8* (7), e1002792.

26.   Hashimoto, K.; Baba, M.; Gohnai, K.; Sato, M.; Shigeta, S. Heat shock induces HIV-1 replication in chronically infected promyelocyte cell line OM10.1. *Arch. Virol.* **1996,** *141* (3-4), 439-447.

27.   Ignatenko, N. A.; Gerner, E. W. Regulation of the HIV1 long terminal repeat by mutant heat shock factor. *Exp. Cell Res.* **2003,** *288* (1), 1-8.

28.   Pan, X.; Lin, J.; Zeng, X.; Li, W.; Wu, W.; Lu, W. Z.; Liu, J.; Liu, S. Heat shock factor 1 suppresses the HIV-induced inflammatory response by inhibiting nuclear factor-κB. *Cell. Immunol.* **2018,** *327*, 26-35.

29.   Vozzolo, L.; Loh, B.; Gane, P. J.; Tribak, M.; Zhou, L.; Anderson, I.; Nyakatura, E.; Jenner, R. G.; Selwood, D.; Fassati, A. Gyrase B inhibitor impairs HIV-1 replication by targeting HSP90 and the capsid protein. *J. Biol. Chem.* **2010**, jbc. M110. 155275.

30.   Anderson, I.; Low, J. S.; Weston, S.; Weinberger, M.; Zhyvoloup, A.; Labokha, A. A.; Corazza, G.; Kitson, R. A.; Moody, C. J.; Marcello, A. et al. Heat shock protein 90 controls HIV-1 reactivation from latency. *Proc. Natl. Acad. Sci. U. S. A.* **2014,** *111* (15), E1528-E1537.

31.   Urano, E.; Morikawa, Y.; Komano, J., Novel role of HSP40/DNAJ in the regulation of HIV-1 replication. *JAIDS J. Acquired Immune Defic. Syndr.* **2013,** *64* (2), 154-162.

32.   Kumar, M.; Mitra, D. Heat shock protein 40 is necessary for human immunodeficiency virus-1 Nef-mediated enhancement of viral gene expression and replication. *J. Biol. Chem.* **2005**.

33.   Kumar, M.; Rawat, P.; Khan, S. Z.; Dhamija, N.; Chaudhary, P.; Ravi, D. S.; Mitra, D. Reciprocal regulation of human immunodeficiency virus-1 gene expression and replication by heat shock proteins 40 and 70. *J. Mol. Biol.* **2011,** *410* (5), 944-958.

34.   Agostini, I.; Popov, S.; Li, J.; Dubrovsky, L.; Hao, T.; Bukrinsky, M. Heat-shock protein 70 can replace viral protein R of HIV-1 during nuclear import of the viral preintegration complex. *Exp. Cell Res.* **2000,** *259* (2), 398-403.

35.   Iordanskiy, S.; Zhao, Y.; DiMarzio, P.; Agostini, I.; Dubrovsky, L.; Bukrinsky, M. Heat-shock protein 70 exerts opposing effects on Vpr-dependent and Vpr-independent HIV-1 replication in macrophages. *Blood* **2004,** *104* (6), 1867-1872.

36.   Chaudhary, P.; Khan, S. Z.; Rawat, P.; Augustine, T.; Raynes, D. A.; Guerriero, V.; Mitra, D. Hsp70 binding protein 1 (HspBP1) suppresses HIV-1 replication by inhibiting NF-κB mediated activation of viral gene expression. *Nucleic Acids Res.* **2016,** *44* (4), 1613-1629.

37.   Zuo, J.; Baler, R.; Dahl, G.; Voellmy, R. Activation of the DNA-binding ability of human heat shock transcription factor 1 may involve the transition from an intramolecular to an intermolecular triple-stranded coiled-coil structure. *Mol. Cell. Biol.* **1994,** *14* (11), 7557-7568.

38.   Mahat, D. B.; Salamanca, H. H.; Duarte, F. M.; Danko, C. G.; Lis, J. T. Mammalian heat shock response and mechanisms underlying its genome-wide transcriptional regulation. *Mol. Cell* **2016,** *62* (1), 63-78.

39.   Ryno, L. M.; Genereux, J. C.; Naito, T.; Morimoto, R. I.; Powers, E. T.; Shoulders, M. D.; Wiseman, R. L. Characterizing the altered cellular proteome induced by the stress-independent activation of heat shock factor 1. *ACS Chem. Biol.* **2014,** *9* (6), 1273-1283.

40.   Moore, C. L.; Dewal, M. B.; Nekongo, E. E.; Santiago, S.; Lu, N. B.; Levine, S. S.; Shoulders, M. D. Transportable, chemical genetic methodology for the small molecule-mediated inhibition of heat shock factor 1. *ACS Chem. Biol.* **2016,** *11* (1), 200-10.

41.   Shoulders, M. D.; Ryno, L. M.; Cooley, C. B.; Kelly, J. W.; Wiseman, R. L. Broadly applicable methodology for the rapid and dosable small molecule-mediated regulation of transcription factors in human cells. *J. Am. Chem. Soc.* **2013,** *135* (22), 8129-32.

42.   Sebastian, R. M.; Shoulders, M. D. Chemical biology framework to illuminate proteostasis. *Annu. Rev. Biochem.* **2020,** *89*, in press.

43.   Voellmy, R. Dominant-positive and dominant-negative heat shock factors. *Methods* **2005,** *35* (2), 199-207.

44.   Del Razo, L. M.; Quintanilla-Vega, B.; Brambila-Colombres, E.; Calderón-Aranda, E. S.; Manno, M.; Albores, A., Stress proteins induced by arsenic. *Toxicol. Appl. Pharmacol.* **2001,** *177* (2), 132-148.

45.   Kimpton, J.; Emerman, M. Detection of replication-competent and pseudotyped human immunodeficiency virus with a sensitive cell line on the basis of activation of an integrated beta-galactosidase gene. *J. Virol.* **1992,** *66* (4), 2232-2239.

46.   Wong, M. Y.; DiChiara, A. S.; Suen, P. H.; Chen, K.; Doan, N.-D.; Shoulders, M. D. Adapting secretory proteostasis and function through the unfolded protein response. *Curr. Top. Microbiol. Immunol.* **2017,** *414*, 1-25.

47.   Lee, A.-H.; Iwakoshi, N. N.; Glimcher, L. H. XBP-1 regulates a subset of endoplasmic reticulum resident chaperone genes in the unfolded protein response. *Mol. Cell. Biol.* **2003,** *23* (21), 7448-59.

48.   Iwamoto, M.; Bjorklund, T.; Lundberg, C.; Kirik, D.; Wandless, T. J. A general chemical method to regulate protein stability in the mammalian central nervous system. *Chem. Biol.* **2010,** *17* (9), 981-8.

49.   Shoulders, M. D.; Ryno, L. M.; Genereux, J. C.; Moresco, J. J.; Tu, P. G.; Wu, C.; Yates, J. R.; Su, A. I.; Kelly, J. W.; Wiseman, R. L. Stress-independent activation of XBP1s and/or ATF6 reveals three functionally diverse ER proteostasis environments. *Cell Rep.* **2013,** *3* (4), 1279-1292.

50.   Chen, J. J.; Genereux, J. C.; Qu, S.; Hulleman, J. D.; Shoulders, M. D.; Wiseman, R. L. ATF6 activation reduces the secretion and extracellular aggregation of destabilized variants of an amyloidogenic protein. *Chem. Biol.* **2014,** *21* (11), 1564-74.

51.   Dewal, M. B.; DiChiara, A. S.; Antonopoulos, A.; Taylor, R. J.; Harmon, C. J.; Haslam, S. M.; Dell, A.; Shoulders, M. D. XBP1s links the unfolded protein response to the molecular architecture of mature N-glycans. *Chem. Biol.* **2015,** *22* (10), 1301-12.

52.   Wong, M. Y.; Chen, K.; Antonopoulos, A.; Kasper, B. T.; Dewal, M. B.; Taylor, R. J.; Whittaker, C. A.; Hein, P. P.; Dell, A.; Genereux, J. C.; Haslam, S. M.; Mahal, L. K.; Shoulders, M. D. XBP1s activation can globally remodel N-glycan structure distribution patterns. *Proc. Natl. Acad. Sci. U. S. A.* **2018,** *115* (43), E10089-E10098.

53.   Wires, E. S.; Henderson, M. J.; Yan, X.; Bäck, S.; Trychta, K. A.; Lutrey, M. H.; Harvey, B. K. Longitudinal monitoring of Gaussia and Nano luciferase activities to concurrently assess ER calcium homeostasis and ER stress in vivo. *PLoS One* **2017,** *12* (4), e0175481.

54.   Kroeger, H.; Grimsey, N.; Paxman, R.; Chiang, W.-C.; Plate, L.; Jones, Y.; Shaw, P. X.; Trejo, J.; Tsang, S. H.; Powers, E.; Kelly, J. W.; Wiseman, R. L.; Lin, J. H. The unfolded protein response regulator ATF6 promotes mesodermal differentiation. *Sci. Signaling* **2018,** *11* (517).

55.   Subramanian, A.; Tamayo, P.; Mootha, V. K.; Mukherjee, S.; Ebert, B. L.; Gillette, M. A.; Paulovich, A.; Pomeroy, S. L.; Golub, T. R.; Lander, E. S. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **2005,** *102* (43), 15545-15550.

56.   Todorova, T.; Bock, F. J.; Chang, P. Poly (ADP-ribose) polymerase-13 and RNA regulation in immunity and cancer. *Trends Mol. Med.* **2015,** *21* (6), 373-384.

57.   Ghimire, D.; Rai, M.; Gaur, R., Novel host restriction factors implicated in HIV-1 replication. *J. Gen. Virol.* **2018,** *99* (4), 435-446.

58.   Zhu, Y.; Chen, G.; Lv, F.; Wang, X.; Ji, X.; Xu, Y.; Sun, J.; Wu, L.; Zheng, Y.-T.; Gao, G. Zinc-finger antiviral protein inhibits HIV-1 infection by selectively targeting multiply spliced viral mRNAs for degradation. *Proc. Natl. Acad. Sci. U. S. A.* **2011,** *108* (38), 15834-15839.

59.   Fujimoto, M.; Takii, R.; Takaki, E.; Katiyar, A.; Nakato, R.; Shirahige, K.; Nakai, A. The HSF1–PARP13–PARP1 complex facilitates DNA repair and promotes mammary tumorigenesis. *Nat. Commun.* **2017,** *8* (1), 1638.

60.   Fujimoto, M.; Takii, R.; Katiyar, A.; Srivastava, P.; Nakai, A. Poly (ADP-ribose) polymerase 1 promotes the human heat shock response by facilitating heat shock transcription factor 1 binding to DNA. *Mol. Cell. Biol.* **2018,** *38* (13), e00051-18.

61.   The ENCODE Project Consortium. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **2007,** *447* (7146), 799.

62.   Gurer, C.; Cimarelli, A.; Luban, J. Specific incorporation of heat shock protein 70 family members into primate lentiviral virions. *J. Virol.* **2002,** *76* (9), 4666-4670.

63.   Franke, E. K.; Yuan, H. E. H.; Luban, J. Specific incorporation of cyclophilin A into HIV-1 virions. *Nature* **1994,** *372* (6504), 359-362.

64.   Bartz, S. R.; Pauza, C. D.; Ivanyi, J.; Jindal, S.; Welch, W. J.; Malkovsky, M., An Hsp60 related protein is associated with purified HIV and SIV. *J. Med. Primatol.* **1994,** *23* (2-3), 151-154.

65.   Matysiak, J.; Lesbats, P.; Mauro, E.; Lapaillerie, D.; Dupuy, J.-W.; Lopez, A. P.; Benleulmi, M. S.; Calmels, C.; Andreola, M.-L.; Ruff, M. et al. Modulation of chromatin structure by the FACT histone chaperone complex regulates HIV-1 integration. *Retrovirology* **2017,** *14* (1), 39.

66.   Quinlan, A. R.; Hall, I. M., BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010,** *26* (6), 841-842.

67.   Dobin, A.; Davis, C. A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T. R., STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **2013,** *29* (1), 15-21.

68.   Li, B.; Dewey, C. N., RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* **2011,** *12* (1), 323.

69.   Love, M. I.; Huber, W.; Anders, S., Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014,** *15* (12), 550.

70.   Huang, D. W.; Sherman, B. T.; Lempicki, R. A., Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **2008,** *37* (1), 1-13.

71.   Mootha, V. K.; Lindgren, C. M.; Eriksson, K.-F.; Subramanian, A.; Sihag, S.; Lehar, J.; Puigserver, P.; Carlsson, E.; Ridderstråle, M.; Laurila, E., PGC-1α-responsive genes involved in

oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* **2003,** *34* (3), 267.

Chapter 4: Conclusions and future directions

4.1. Conclusions

Recent pioneering work by our laboratory established the fundamental impact of host proteostasis on viral evolution, which defines the accessible evolutionary trajectories and mutational tolerance of viral proteins. My work targets the evolutionary implications of such impact. The work presented in this thesis demonstrates, for the first time, that interaction of viral proteins with host chaperones facilitates viral adaptation by assisting folding of destabilized viral mutants, which can possess important evolutionary advantages. Together with my colleague, Dr. Angela M. Phillips, we discovered the critical dependence of biophysically defective innate immune escape 1918 Spanish Flu pandemic variant on the Heat Shock Factor 1 (HSF1)-controlled proteostasis network assistance. Restricted access to such proteostasis assistance shifts the balance between phenotypical benefits and biophysical costs of viral protein mutants towards innate immunity-sensitive, but structurally intact variant. This discovery establishes the novel role of chaperones as potentiators of viral adaptation and illuminates the underlying molecular details, thus adding a new chapter to fundamental evolutionary biology and highlighting a potent therapeutic avenue, which may be less susceptible to resistance than conventional approaches.

My work in collaboration with Dr. Emmanuel E. Nekongo demonstrates the flip side of viral dependence on the host proteostasis assistance. We discovered the novel role of host HSF1 as an antiviral restriction factor. Stress-independent HSF1 activation inhibits human immunodeficiency virus 1 (HIV-1) replication. Not only is a smaller amount of viral particles produced, but also the infectivity of viral progeny is compromised. Although the exact molecular mechanism is yet to be established, it does not include global induction of antiviral response and likely targets the later stages of viral replication cycle, such as folding, assembly and egress of viral particles. Interestingly, we did not observe similar inhibition for influenza, suggesting that the effect is specific to HIV-1. This work demonstrates, for the first time, that HSF1 is capable of restricting replication of certain viruses.

Taken together, these discoveries highlight the complex consequences of viral interaction with the host proteostasis network. Viruses evolved to leverage the vital functions of the host chaperones for their biophysical and evolutionary benefit. However, any perturbation of the proteostasis network can have a drastic effect on viral proliferation. Reduction of chaperone activity and accessibility exposes the biophysical defects of mutating viral proteins. Excessive levels of cytosolic proteostasis network components, on the other hand, can selectively restrict viral replication. This dual role of proteostasis in the viral life cycle is similar to the "double-

edged sword" role of the immune system in the host's health: compromised immunity renders the host vulnerable to invading pathogens, but immune overreaction can directly harm the organism. Thus, just like the host immune system has to be tightly regulated to avoid harming the organism, interactions with the host proteostasis network have to be balanced for efficient viral replication. Any perturbation in this finely-tuned system can severely compromise viral replication.

4.2. Future directions

4.2.1 Molecular origins and mechanism of viral protein folding by the host chaperones

My discovery of the essential role of chaperones in addressing the biophysical defect associated with the destabilized adaptive Pro283 influenza nucleoprotein variant provokes numerous immediate follow-up questions. First, additional experimental evidence is needed to gain insight into the structural consequences of Pro283 acquisition. Our thermal stability data and molecular dynamics simulations, discussed in Chapter 2, provide initial insight into the molecular origins of selection against the wild-type residue in biophysically challenging host environments. However, the crystal structure of A/Aichi/2/1968 (H3N2) nucleoprotein, which has not yet been solved, could directly demonstrate the proposed disruption of the α-helix containing position 283 by the proline residue. A comparison of Pro283 and Ser283 nucleoprotein variants could also highlight any additional minor structural perturbations, which could not be detected or predicted using the preliminary data from Chapter 2.

In addition to obtaining a static structure by X-ray crystallography, the impact of structure perturbation on nucleoprotein protein dynamics and folding can be assessed using complementary biophysical and biochemical methods. For example, Nuclear Magnetic Resonance (NMR) spectroscopy can inform protein dynamics in solution, which is a critical piece of evidence for understanding the behavior of destabilized Pro283 NP in a biologically relevant state[1]. The success of protein folding and local structural changes introduced by the proline residue can be investigated using hydrogen/deuterium exchange mass spectrometry[2]. Additionally, protein folding kinetics can be assessed and compared between variants using protein refolding assays[3].

Next, my pilot proteomics interactome data discussed in Appendix can guide a systemic profiling of individual host factors involved into Pro283 variant folding. The relevant host factors can be either chaperones, providing direct folding assistance, or endogenous chaperone clients, the interaction of which with the viral proteins is perturbed by a remodeled proteostasis network. Investigation of the essential individual host components would require two sets of specific tools. First, the activity[4-7] and levels[8-10] of individual host factors have to be precisely regulated while minimizing the off-target perturbation of the system and avoiding induction of stress responses or compensatory mechanisms[9]. Second, the performance of viral protein variants has to be monitored in the perturbed host system, from transcript and protein levels[9] to assessing the fitness of complete viral particles. Elucidating the critical host factors and their specific roles in resolving the Pro283 biophysical defect will inform the possible molecular

mechanism of chaperone-mediated viral protein evolution. Such a mechanism can direct further studies of viral evolution, establishing the framework for predicting evolutionary trajectories and highlighting novel vulnerable steps in the viral lifecycle, which would be less prone to the rise of resistance than conventional therapeutic targets.

The future research directions outlined above are not limited exclusively to the Pro283 nucleoprotein variant. Deep mutational scanning, described in Chapter 2 identified other sites that experience strong selection in biophysically challenging environment. To note a few, sites 334, 353 and 377 experienced a strong selection against wild-type, while sites 50, 82 and 446 preferred wild-type over mutant residues. Our primary focus on position 283 is dictated by the largest observed effect and its historical role in the emergence of the 1918 pandemic strain. Pursuing other positions in influenza nucleoprotein, which are also sensitive to chaperone depletion at fever temperature would provide a more comprehensive view of the host chaperone assisted nucleoprotein folding mechanism.

Finally, the molecular mechanism of HIV-1 replication inhibition by the stress-independent HSF1 activation, discussed in Chapter 3 requires further investigation. First, the transcript and protein levels of viral proteins can be used to better inform the steps of a viral life cycle, which are affected. Next, the hypothesis of ZAP contributing to suppressing HIV-1 proliferation has to be tested by observing the extent of viral replication as a function of ZAP levels. Lastly, systematic investigation of other individual host components, guided by the provided transriptome profile can highlight additional players, which are involved in the observed replication inhibition.

4.2.2 The role of epistasis in accommodating destabilized adaptive mutations

Pursuing detailed structural studies on select protein variants cannot inform the possible importance of the genetic context of mutations of interest. Additional investigations, targeted on elucidating the role of epistasis in assisting fixation of destabilizing phenotypically beneficial mutations would provide an evolutionary context for the fundamental biophysical and biochemical studies proposed above.

The first prospective direction is to assess whether epistasis can stabilize the Pro283 nucleoprotein variant in the unfavorable environment created by reduced chaperone capacity and fever temperature. Approaching this question experimentally would involve a continuous propagation of Pro283 variant-carrying virus under biophysically challenging conditions and monitoring accumulation of mutations by long-read sequencing[11] of the entire nucleoprotein gene. One potential strategy would be to serially passage the wild-type virus for multiple passages[12]. An alternative approach would be to serially passage deep mutational scanning (DMS) viral libraries. The benefit of such an approach is in the high initial diversity of the nucleoprotein library upon which selection can act. The diverse starting population might already contain stabilizing mutations of interest, which would be enriched after a few serial passages, while numerous destabilizing and unfavorable variants would be rapidly purged from the population.

Analyzing long-read sequencing data from serial passaging experiments[13] may reveal a set of potentially stabilizing mutations. The follow-up biochemical and biophysical studies, including strategies proposed in the previous section, will test whether the identified mutations indeed contribute to accommodating Pro283 and partially or fully compensate the biophysical defect introduced by this adaptive mutation. Additional phylogenetic analyses could potentially inform and support experimental studies, however such an approach may be limited by the sparse available sequencing data of strains prior to A/Brevig Mission/1/18 (H1N1).

The experimental strategies, suggested above, are not limited to influenza nucleoprotein targeted by the human MxA. They can be expanded to other viral proteins experiencing strong selection. Of particular interest is influenza hemagglutinin (HA), which is the target of narrow and broad neutralizing antibodies[14]. Prior studies have mapped the antibody escape mutations of A/WSN/1933 (H1N1) influenza[14-15] and profiled the HA mutational tolerance as a function of the host proteostasis composition and activity[16]. A close comparative analysis of the two available datasets may identify positions and individual mutations, which, similar to Pro283 nucleoprotein variant, enable antibody escape but are biophysically compromised and critically

depend on the chaperone assistance. After verification of individual HA mutants, the same serial passaging approaches as ones proposed for Pro283 nucleoprotein variant can be applied to investigate the potential epistasis involved in stabilizing the antibody escape mutations. In addition, the extent of the available host chaperone assistance can justify and modulate the need for epistatic stabilization and should be addressed experimentally as well.

Deciphering the possibility and extent of epistasis in stabilizing the adaptive mutations and partially lifting the need for chaperone assistance would directly inform the evolutionary trajectories conferring influenza immune resistance and possible escape routes from therapeutic chaperone inhibition.

4.2.3   Addressing the biomedical importance of viral dependence on chaperone assistance

The discoveries reported in Chapter 2 and Appendix hold an important biomedical relevance. The first step towards translating those fundamental findings into therapeutically promising directions is to shift towards models that better reflect human physiology during viral infection. To begin with, the work from Chapter 2 and Appendix was performed in Madin Darby Canine Kidney (MDCK) cell lines, which are the typical workhorses of influenza research due to their ability to produce high yields of influenza[17]. Transitioning to more relevant human cell lines, such as immortalized human adenocarcinoma A549 and primary Normal Human Bronchial Epithelial (NHBE) cells, would test whether the effects and effect sizes discovered in MDCKs can also be observed in the human cells.

Next, adopting a more recent, clinically relevant influenza strain within our established experimental framework can inform the importance of chaperone assistance in evolution of emerging strains. While using laboratory adapted strains as models has its own advantages, rapid evolution of influenza keeps current circulating strains diverging from those models year after year. To address this issue, we have reached out to our collaborators and obtained the A/Perth/16/2009 (H3N2) influenza virus, as well as DMS libraries of this strain[18]. We have initiated the preparation for profiling the mutational tolerance of Perth/09 HA as a function of the host proteostasis. My graduate student colleagues Jimin Yoon and Jessica Patrick engineered MDCK cell lines that express proteostasis perturbation tools[16] and sustain efficient propagation of Perth/09 strain[18]. Performing DMS, following-up on individual variants, and exploring the underlying molecular mechanisms of the observed patterns would supply an invaluable and rich dataset. Such a dataset could be further placed in the context of previous antigenic profiling of this strain[19] and overlapped with the vastly sequenced preceding and following strains to inform the epidemiological importance of chaperone assistance in viral evolution.

Finally, we can ask interesting cell biology questions to better understand the effects of host proteostasis on viral infection progression. For example, influenza hemagglutinin remains the most heavily investigated viral target, possessing a complex, chaperone-dependent folding mechanism[20-21]. Our recent finding suggests that HA can accommodate mutations better in the host cell environment with enhanced proteostasis capacity due to increased HA mutational tolerance in such a setting[16]. This fact buries the important implication of generation and propagation of viruses containing destabilized HA mutant variants, which might be not viable in the host cells without additional folding assistance. Some of such variants can be potentially

126

conferring a beneficial phenotype, such as an improved immune escape ability. These destabilized adaptive HA variants can be identified by a functional screen of DMS libraries, which were generated in the host cells with increased chaperones levels. If the effect observed in a DMS batch experiment is validated in pairwise viral competitions, then the identified mutants should be closely investigated, including detailed biophysical and structural studies, as well as insights into potential epistasis.

In addition, glycans play an important role in HA folding process[22] and are directly involved in HA biological function[23]. A recent discovery by our group demonstrated that activation of the branch of the unfolded protein response (UPR) controlled by the spliced X-box binding protein 1 (XBP1s) globally remodels the cellular N-glycome signature[24]. This discovery suggests that HA, which is produced, folded and glycosylated by host cells, may be undergoing similar changes in its glycan shield. Also, given the role of the sialic acid on the host cells surface as HA receptor, host glycan remodeling can have a drastic impact on influenza infection dynamics and efficiency. Pursuing these directions using glycomics techniques could shed light on what may turn out to be a previously underestimated role of the UPR in defining host-virus interactions.

In conclusion, this thesis provides the starting point for a vast group of succeeding research projects including detailed biochemical and biophysical investigations, evolutionary and epidemiological analyses and biomedical research. Implementation of these projects and integration of the achieved results would create a complete picture of how host chaperones facilitate influenza infection and define its accessible evolutionary routes.

4.3 References

1. Markwick, P. R.; Malliavin, T.; Nilges, M. Structural biology by NMR: structure, dynamics, and interactions. *PLoS Comp. Biol.* **2008,** *4* (9).
2. Konermann, L.; Pan, J.; Liu, Y.-H. Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chem. Soc. Rev.* **2011,** *40* (3), 1224-1234.
3. Hingorani, K. S.; Gierasch, L. M. Comparing protein folding in vitro and in vivo: foldability meets the fitness challenge. *Curr. Opin. Struct. Biol.* **2014,** *24*, 81-90.
4. Ying, W.; Du, Z.; Sun, L.; Foley, K. P.; Proia, D. A.; Blackman, R. K.; Zhou, D.; Inoue, T.; Tatsuta, N.; Sang, J. Ganetespib, a unique triazolone-containing Hsp90 inhibitor, exhibits potent antitumor activity and a superior safety profile for cancer therapy. *Mol. Cancer Ther.* **2012,** *11* (2), 475-484.
5. Fewell, S. W.; Smith, C. M.; Lyon, M. A.; Dumitrescu, T. P.; Wipf, P.; Day, B. W.; Brodsky, J. L. Small molecule modulators of endogenous and co-chaperone-stimulated Hsp70 ATPase activity. *J. Biol. Chem.* **2004,** *279* (49), 51131-51140.
6. Papandreou, I.; Denko, N. C.; Olson, M.; Van Melckebeke, H.; Lust, S.; Tam, A.; Solow-Cordero, D. E.; Bouley, D. M.; Offner, F.; Niwa, M. Identification of an Ire1alpha endonuclease specific inhibitor with cytotoxic activity against human multiple myeloma. *Blood* **2011,** *117* (4), 1311-1314.
7. Stiegler, S. C.; Rübbelke, M.; Korotkov, V. S.; Weiwad, M.; John, C.; Fischer, G.; Sieber, S. A.; Sattler, M.; Buchner, J. A chemical compound inhibiting the Aha1–Hsp90 chaperone complex. *J. Biol. Chem.* **2017,** *292* (41), 17073-17083.
8. Qi, L. S.; Larson, M. H.; Gilbert, L. A.; Doudna, J. A.; Weissman, J. S.; Arkin, A. P.; Lim, W. A. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **2013,** *152* (5), 1173-1183.
9. Sebastian, R. M.; Shoulders, M. D. Chemical biology framework to illuminate proteostasis. *Annu. Rev. Biochem.* **2020,** *89.*
10. Alló, M.; Buggiano, V.; Fededa, J. P.; Petrillo, E.; Schor, I.; De La Mata, M.; Agirre, E.; Plass, M.; Eyras, E.; Elela, S. A. Control of alternative splicing through siRNA-mediated transcriptional gene silencing. *Nat. Struct. Mol. Biol.* **2009,** *16* (7), 717-724.
11. Mantere, T.; Kersten, S.; Hoischen, A. Long-read sequencing emerging in medical genetics. *Front. Genet.* **2019,** *10*, 426.
12. Phillips, A. M.; Gonzalez, L. O.; Nekongo, E. E.; Ponomarenko, A. I.; McHugh, S. M.; Butty, V. L.; Levine, S. S.; Lin, Y.-S.; Mirny, L. A.; Shoulders, M. D. Host proteostasis modulates influenza evolution. *eLife* **2017,** *6*, e28652.
13. Amarasinghe, S. L.; Su, S.; Dong, X.; Zappia, L.; Ritchie, M. E.; Gouil, Q. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.* **2020,** *21* (1), 1-16.
14. 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.* **2018,** *9* (1), 1-12.
15. Doud, M. B.; Hensley, S. E.; Bloom, J. D. Complete mapping of viral escape from neutralizing antibodies. *PLoS Pathog.* **2017,** *13* (3), e1006271.
16. Phillips, A. M.; Doud, M. B.; Gonzalez, L. O.; Butty, V. L.; Lin, Y.-S.; Bloom, J. D.; Shoulders, M. D. Enhanced ER proteostasis and temperature differentially impact the mutational tolerance of influenza hemagglutinin. *eLife* **2018,** *7*, e38795.
17. Seitz, C.; Frensing, T.; Höper, D.; Kochs, G.; Reichl, U. High yields of influenza A virus in Madin–Darby canine kidney cells are promoted by an insufficient interferon-induced antiviral state. *J. Gen. Virol.* **2010,** *91* (7), 1754-1763.
18. Lee, J. M.; Huddleston, J.; Doud, M. B.; Hooper, K. A.; Wu, N. C.; Bedford, T.; Bloom, J. D. Deep mutational scanning of hemagglutinin helps predict evolutionary fates of human H3N2 influenza variants. *Proc. Natl. Acad. Sci. U. S. A.* **2018,** *115* (35), E8276-E8285.

19.   Lee, J. M.; Eguia, R.; Zost, S. J.; Choudhary, S.; Wilson, P. C.; Bedford, T.; Stevens-Ayers, T.; Boeckh, M.; Hurt, A. C.; Lakdawala, S. S. Mapping person-to-person variation in viral mutations that escape polyclonal serum targeting influenza hemagglutinin. *eLife* **2019,** *8*, e49324.

20.   Braakman, I.; Hoover-Litty, H.; Wagner, K. R.; Helenius, A. Folding of influenza hemagglutinin in the endoplasmic reticulum. *J. Cell Biol.* **1991,** *114* (3), 401-411.

21.   Hebert, D. N.; Foellmer, B.; Helenius, A. Calnexin and calreticulin promote folding, delay oligomerization and suppress degradation of influenza hemagglutinin in microsomes. *EMBO J.* **1996,** *15* (12), 2961-2968.

22.   Daniels, R.; Kurowski, B.; Johnson, A. E.; Hebert, D. N. N-linked glycans direct the cotranslational folding pathway of influenza hemagglutinin. *Mol. Cell* **2003,** *11* (1), 79-90.

23.   Wilks, S.; de Graaf, M.; Smith, D. J.; Burke, D. F. A review of influenza haemagglutinin receptor binding as it relates to pandemic properties. *Vaccine* **2012,** *30* (29), 4369-4376.

24.   Wong, M. Y.; Chen, K.; Antonopoulos, A.; Kasper, B. T.; Dewal, M. B.; Taylor, R. J.; Whittaker, C. A.; Hein, P. P.; Dell, A.; Genereux, J. C. et al. XBP1s activation can globally remodel N-glycan structure distribution patterns. *Proc. Natl. Acad. Sci. U. S. A.* **2018,** *115* (43), E10089-E10098.

Appendix A: Elucidating the host factors essential for assisting the folding of the influenza nucleoprotein Pro283 variant

A.1   Author contributions

A.I.P., A.M.P. and M.D.S. conceived the project. A.I.P., L.J.P. and M.D.S. planned research, designed experiments, and analyzed data. A.I.P. performed experiments. L.J.P. and J.D.B. provided reagents. M.D.S. supervised research.

A.2   Introduction

As discussed in Chapter 2, we have demonstrated that acquisition of innate immune escape Pro283 mutation in influenza nucleoprotein (NP) during the emergence of the 1918 Spanish Flu pandemic strain destabilized the viral protein. These data indicate that some form of folding assistance from Heat Shock Factor 1 (HSF1)-controlled host factors is likely essential to enable effective propagation of this biophysically defective adaptive variant[1]. Inhibition of HSF1 at fever temperature renders Pro283 variant biophysically unfit and unable to efficiently propagate. Given the critical dependence of Pro283 fitness on the HSF1 proteostasis network, we set the goal to elucidate the individual host factors responsible for addressing the Pro283-induced biophysical defect and enabling proper NP folding.

Introduction of proline into the polypeptide chain has certain implications for the protein structure. The absence of hydrogen on the α amino group prevents proline from forming a hydrogen bond with the adjacent amino acid residues[2]. This property renders proline highly unfavorable in the middle of α helices and β sheets because it disrupts both of these regular secondary structures[2]. Incorporation of proline in the middle of an α helix can create up to a 30° kink in the helix[3-4], often exposing the proline side chain on the protein surface and positioning the proline ring away from the neighboring amino acid residues. In a more extreme scenario, proline can completely break the secondary structure element by introducing a 180° turn[5]. The turn formation is more favorable if residues next to proline include glycine, which contributes to the flexibility of the peptide chain, and amino acids with aromatic side chains, which can stabilize the turn by aromatic stacking[2, 5]. This property makes proline residues naturally prevalent in turns, ends of strands and helices.

The unique cyclic structure of proline side chain imposes constraints on rates of protein synthesis and folding. Incorporation of proline into the growing peptide chain slows down translation[6], and the presence of multiple proline codons in a row can stall the ribosome and require assistance of translation elongation factor eIF5A to resume and facilitate translation[7]. Folding of the newly synthesized polypeptide is also complicated due to proline *cis*/*trans* isomerization. The partial double bond character of proline peptide bond creates a high activation energy barrier for *cis*/*trans* isomerization, which is usually a rate limiting step in protein folding[8].

Multiple chaperones from the peptidylprolyl isomerase (PPIase) family facilitate *cis*/*trans* isomerization and play important roles in replication of viruses[9], including influenza[10]. For example, the isomerase activity of cyclophilin A (CypA), encoded by the PPIA gene, is vital for

assisting folding of hepatitis C virus (HCV)[11] and human immunodeficiency virus (HIV)[12] proteins, but is not essential for influenza replication[13]. Moreover, CypA also restricts influenza replication by targeting the viral protein M1 to degradation[14]. Another PPIase, Pin1, which specifically recognizes phospho-Ser/Thr-Pro motif, stabilizes the human T-cell leukemia virus type 1 (HTLV-1) Tax protein by catalyzing the phosphorylation-dependent prolyl isomerization[15]. In addition to post-translational folding assistance, host factors enabling co-translational folding might also be important for assisting Pro283 variant. For example, ribosome-associated molecular chaperone trigger factor (TF) has a PPIase domain, which recognizes the proline residue in the unfolded client protein and accelerates peptidyl-prolyl bond isomerization[16].

Elucidating the exact host factors that assist NP in dealing with the structural consequences of incorporating the Pro283 residue requires a systematic approach. We have chosen mass-spectrometry based proteomics for identification of these host proteins that NP actively interacts with in different proteostasis environments. We have developed and optimized an NP co-immunoprecipitation protocol, successfully evaluated our experimental system in a pilot proteomics screen and are now well-positioned for future detailed interactome investigation.

A.3   Results and Discussion

We set the goal to examine NP on the protein level and investigate the details of its interaction with host factors. We began by developing an NP expression system in mammalian cells to assess whether the depletion of HSF1-controlled host factors affects NP protein expression. Our initial approach was to use transient transfection to deliver the nucleoprotein gene into MDCK$^{dn\text{-}cHSF1}$ cells with perturbed proteostasis environments. The pilot transfection experiments using FLAG-tagged NP constructs demonstrated a noticeable decrease in NP protein expression upon raising temperature, especially when coupled with cytosolic chaperone depletion (**Figure A.1A**). These preliminary results correlated with our expectations based on the previously observed decrease in viral fitness at 39 °C, discussed in Chapter 2.
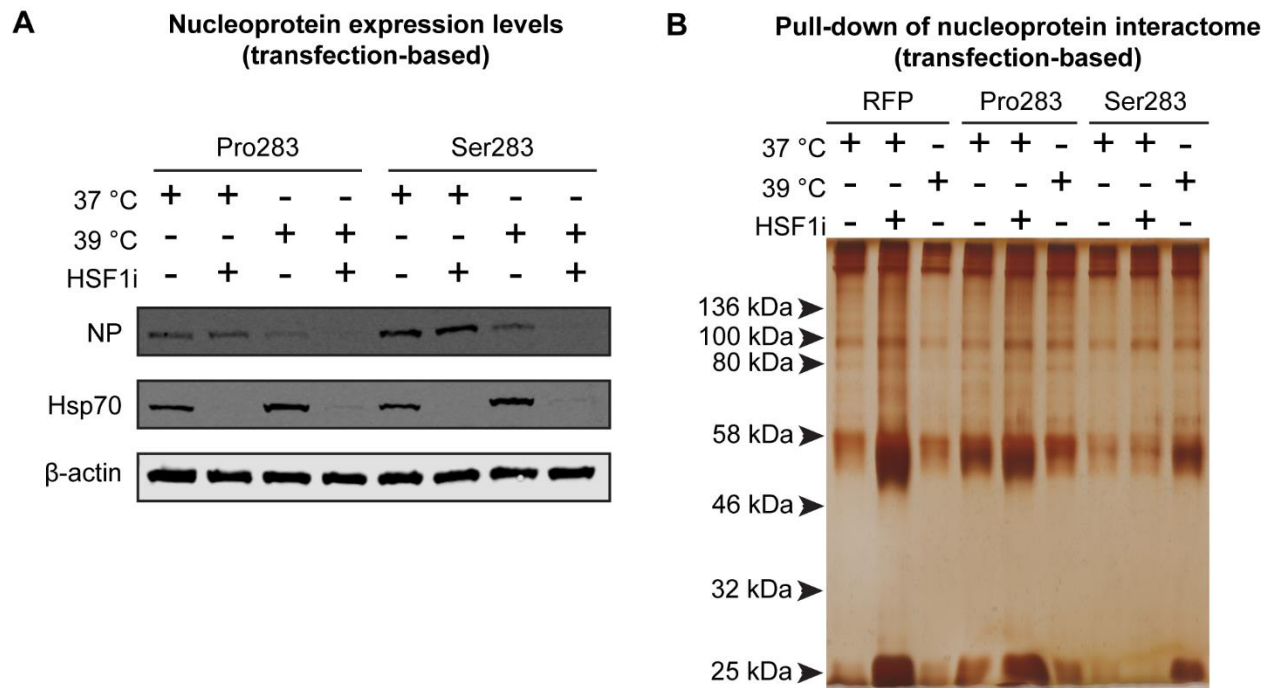

Figure A.1. Initial transfection-based attempts to assess influenza nucleoprotein expression levels in the host cells with perturbed proteostasis.
(A) Representative immunoblot demonstrating differential nucleoprotein expression in perturbed host proteostasis environments, achieved using transient transfection. Influenza nucleoprotein was probed using the N-terminal FLAG tag. The levels of a classical heat shock marker Hsp70 were probed to assess the heat shock induction at 39 C and decrease in expression of cytosolic chaperones upon activation of HSF1i construct upon treatment with 1 µg/mL dox for 48 h. β-actin was probed as a loading control. **(B)** The silver stained SDS-PAGE gel of nucleoprotein co-immunoprecipitation from the perturbed host proteostasis environments. Transfection with RFP was used as a negative control.

However, NP expression levels, detected by immunoblotting, remained low even in the absence of proteostasis perturbation. We attempted to increase the NP signal by performing pull-down experiments. In order to simultaneously get an insight into the extent of NP interaction with the host factors, we performed co-immunoprecipitation and used silver staining of the SDS-PAGE gel to detect the total protein with high sensitivity (**Figure A.1B**). While we did observe the expected enrichment of NP, the difference between NP pull-down samples and negative controls remained very subtle even despite incorporating additional experimental steps for minimizing the non-specific binding of the host proteins. We attribute the cause of such low target signal, which is almost indistinguishable from the negative control, to poor transfection efficiency. Using the red fluorescent protein (RFP) transfection as a negative control allowed us to estimate the transfection efficiency upon visual examination of the fluorescence of transfected cells. The maximum efficiency of MDCK cells transfection did not exceed ~30%. After numerous efforts to improve the quality of our transfections, the persistent low efficiency and poor reproducibility rendered our transfection approach not useful for future investigations.

In an effort to overcome the limitations of transfection-based approach, we switched towards using virus-mediated gene delivery. We successfully engineered the human adenovirus type 5 to encode the Pro283 and Ser283 NP variants and adapted our experimental protocols accordingly. We incorporated a hemagglutinin (HA) epitope tag at the C terminus of NP gene for detection by immunoblotting and subsequent immunoprecipitation for proteomics experiments. To test the performance of the engineered nucleoprotein-encoding adenovirus, we used it to infect MDCK$^{dn-cHSF1}$ cells and assessed the NP expression levels by immunoblotting (**Figure A.2**). We successfully achieved a reproducibly high NP expression, thus overcoming the limitation of the transient transfection approach. Next, we sought to validate whether the dependence of NP expression levels on the host proteostasis environment, observed using transient transfections (**Figure A.1A**), would replicate using adenovirus as a gene delivery tool. In contrast to the results obtained by transfection, we did not observe significant changes in protein levels of both Pro283 and Ser283 NP variants upon temperature increase and depletion of HSF1-controlled proteostasis network (**Figure A.2B)**. Therefore, perturbation of the host proteostasis does not affect steady state NP protein expression levels under these experimental conditions.
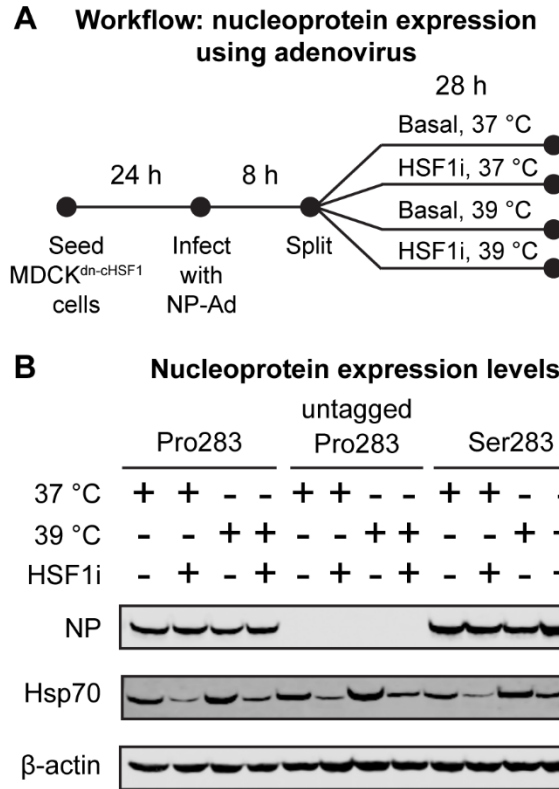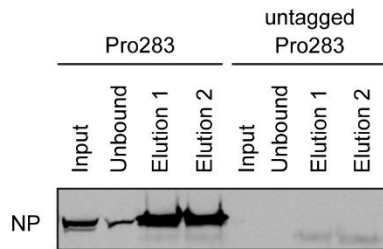
135

A **Workflow: nucleoprotein expression**
using adenovirus

B **Nucleoprotein expression levels**

Figure A.2. Influenza nucleoprotein expression in perturbed proteostasis environments using adenoviral gene delivery.
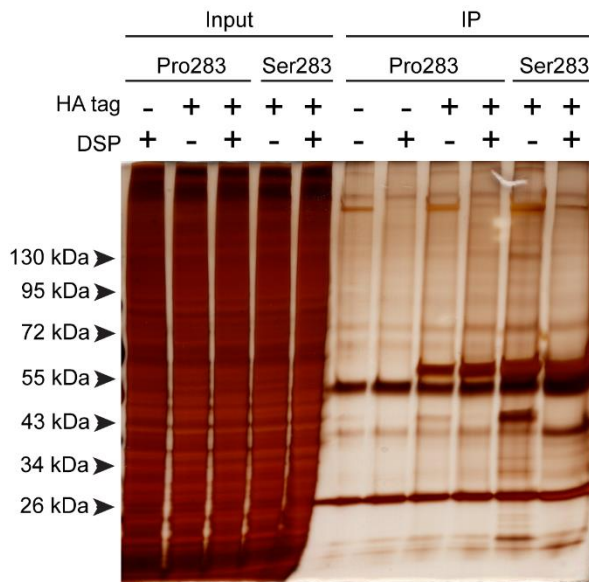**(A)** Experimental workflow and **(B)** Immunoblot demonstrating nucleoprotein expression in perturbed host proteostasis environments, achieved using adenoviral NP gene delivery. Influenza nucleoprotein was probed using the C-terminal HA tag. The levels of a classical heat shock marker Hsp70 were probed to assess the heat shock induction at 39 C and decrease in expression of cytosolic chaperones upon activation of HSF1i construct upon treatment with 1 µg/mL dox for 48 h. β-actin was probed as a loading control.

Next, we revisited our pull-down protocol and optimized it for subsequent interactome investigation. Our successful pilot immunoprecipitation experiment achieved a significant enrichment of NP bait (**Figure A.3A**). In fact, the much higher expression level of NP, achieved by using adenoviral gene delivery, prompted us to increase the amount of beads used for pull-down after detecting residual NP in lysate after overnight immunoprecipitation (the unbound fraction) for a more efficient bait enrichment. In addition, we performed the elution procedure twice to ensure the complete removal of NP from the agarose beads. After implementing the above mentioned modifications to our pull-down protocol, we proceeded with optimizing the cross-linking conditions for co-immunoprecipitation. We tested a range of the cross-linking reagent amounts in search of the optimal concentration that would allow us to efficiently immortalize interactions of NP with the host factors without a significant loss of NP signal

A Optimization of NP immunoprecipitation (adenovirus-based)

B Cross-likning optimization for NP immunoprecipitation

C Nucleoprotein co-immunoprecipitation for a pilot mass-spectometry based proteomics
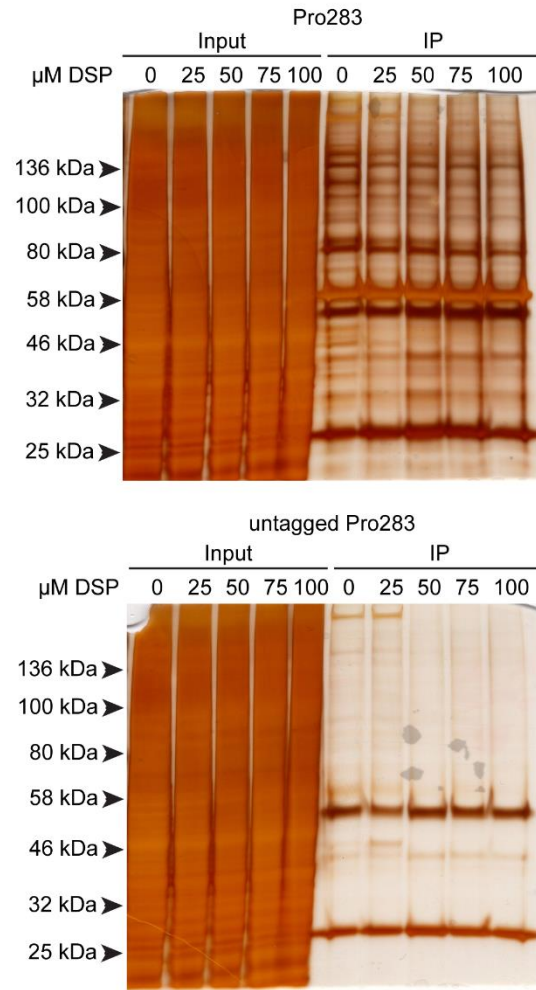
Figure A.3. Nucleoprotein pull-down for subsequent interactome investigation.
(A) Optimization of nucleoprotein (NP) immunoprecipitation conditions. To assess the efficiency of pull-down, NP levels were compared in pre-immunoprecipitation cell lysate (input lane) and supernatants from the two consecutive elutions (elution 1 and elution 2 lanes). The amount of residual NP in post- immunoprecipitation lysate indicates that pull-down enrichment was incomplete. **(B)** Optimization of dithiobis-(succinimidyl propionate) (DSP) cross-linker concentration for co-immunoprecipitation of NP with its host interactome. The silver stain analysis of the SDS-PAGE gel was used to visualize total protein. The untagged Pro283 NP was used as a negative control to assess the extent of non-specific binding. **(C)** Silver stained SDS-PAGE gel of the optimized NP pull-down with or without 50 μM DSP cross-linking.

(**Figure A.3B**). We also expected that performing cross-linking prior to pull-down would not only enrich the true NP interactors, but would also remove non-specific binding of other cellular proteins. However, we detected only a few minor changes between samples with the chosen concentration of the cross-linking reagent (50 μM) and no cross-linker (**Figure A.3C**), thus questioning the need for cross-linking.

The promising results achieved by applying viral delivery for NP expression and immunoprecipitation encouraged us to proceed with investigating the cellular components required for resolving the deleterious effect of incorporating Pro283 into NP structure. Our goal was to pursue mass-spectrometry based proteomics to identify the individual chaperones that are preferentially recruited by Pro283 NP variant. We performed a pilot mass-spectrometry (MS) based proteomics screen to assess the efficiency of our immunoprecipitation protocol and gain a first insight into NP host interactome. Encouraged by detecting sufficient levels of NP bait in our target samples (**Table A.1**), we proceeded with analyzing the levels of relevant chaperones. To account for different levels of NP bait in Pro283 and Ser283 samples, we normalized the raw spectral counts of each interactor from Ser283 sample by multiplying them by the ratio of NP bait counts in Pro283 over Ser283 sample (equal to 345/689 = 0.5). After normalization, we calculated the fold change in spectral counts of Pro283 interactors relative to counts of Ser283 interactors. We especially focused on heat shock proteins and PPIases, which might be involved in stabilizing the defective Pro283 variant and thus are expected to be enriched in Pro283 relative to Ser283 interactome. Especially intriguing, we have noticed a ~2-fold increase in Hsc70 (HSPA8 gene) and Hsp90β (HSP90AB1 gene) chaperone recruitment to Pro283 variant over Ser283, as well as exclusive binding of PPIase G (PPIG gene) to Pro283, but not the Ser283 variant. While these host factors definitely have the potential to assist folding of defective Pro283 variants, these results were obtained from a pilot qualitative screen of a single replicate samples. We will use these initial insights to guide our analysis of the future systematic characterization of NP interactome in perturbed proteostasis environments.

Table A.1. Raw spectral counts and interactors enrichment fold change detected in the pilot NP interactome screen.

Gene	Raw spectral counts			Enrichment fold change: Pro283 over Ser283 (normalized to Pro283 counts)
	Negative control	Pro283	Ser283	
NP (bait protein)	0	345	689	1.0
Heat shock proteins				
HSPA8	28	73	81	1.8
HSP90AB1	6	23	20	2.3
HSPD1	0	29	46	1.3
HSP90AA1	5	25	32	1.6
Peptidyl-prolyl isomerases				
PPIA	8	9	13	1.4
PPIB	0	1	1	2.0
PPIG	1	4	0	--

In preparation for our upcoming proteomics screen we have examined the transcript levels of PPIAses, using transcriptome profiling of MDCK$^{dn\text{-}cHSF1}$ cells, discussed in Chapter 2. Interestingly, the use of our chemical genetic construct for decreasing the levels of HSF1-controlled chaperones (termed HSF1i environment) did not drastically perturb expression of select PPIases (**Table A.2**). In contrast, raising temperature to 39 °C increases expression of PPIase G by 1.59-fold, which is comparable to changes in expression levels of such classic HSF1 chaperone targets as HSP70 and HSPB1, which are upregulated 2.04 and 3.16-fold respectively (**Table A.2**). As noted above, we have detected a more active recruitment of this isomerase by Pro283 NP variant compared to Ser283 already at 37 °C in our pilot proteomics screen (**Table A.1**). It is possible that the difference in the extent of interaction between NP variants and PPIase G is even more drastic at 39 °C. If NP depends on the increased levels of PPIase G to address the Pro283-induced biophysical defect at 39 °C, then decreased expression of this isomerase in HSF1i 39 °C environment might be responsible for the fitness loss of wild-type variant, discussed in details in Chapter 2.

The pilot proteomics experiment also allowed us to evaluate the need for cross-linking in our interactome screen. Previously we did not observe a drastic decrease in the levels of non-specific interactors using the silver staining of corresponding SDS-PAGE gels (**Figure A.3**). However, inspecting the raw spectral counts of common non-specific binders, such as ribosomal and cytoskeleton proteins (**Table A.3**), clearly demonstrates at least partial reduction of their levels with the use of cross-linking agent. This reduction can be attributed to a more rigorous wash of the beads after immunoprecipitation. Cross-linking NP with its interactors allows the use of a wash buffer with denaturing detergents for more efficient removal of non-specific binding proteins without losing true interactors.

Table A.2. Differential expression of PPIases and heat shock protein genes in perturbed proteostasis environments.

Gene	HSF1i 37 °C vs. Basal 37 °C		HSF1i 39 °C vs. Basal 39 °C		Basal 39 °C vs. Basal 37 °C		HSF1i 39 °C vs. Basal 37 °C	
	Fold change	p-value	Fold change	p-value	Fold change	p-value	Fold change	p-value
PPIB	1.08	0.5188	0.97	0.9609	0.92	0.0399	0.90	0.0046
PPIC	0.93	0.8983	0.98	0.9991	1.22	0.0307	1.20	0.0593
PPIE	1.16	0.5603	0.97	0.9958	0.81	0.0075	0.79	0.0023
PPIG	1.03	0.9598	1.01	1.0000	1.58	2.58E-09	1.59	1.52E-09
PPIH	1.02	0.9667	1.07	0.8943	0.82	0.0025	0.88	0.0650
PPIL1	0.94	0.8638	1.01	1.0000	0.85	0.0086	0.86	0.0171
HSP70	0.48	7.58E-11	0.17	1.72E-126	2.04	1.46E-19	0.34	5.36E-36
HSP90AA1	0.70	1.71E-18	0.63	1.47E-34	1.15	0.0007	0.72	1.11E-17
HSP90AB1	1.04	0.7489	1.00	1.00	0.83	5.73E-13	0.83	9.88E-13
HSPB1	0.62	0.0076	0.44	7.33E-23	3.16	4.96E-32	1.38	0.0036

Table A.3. Raw spectral counts of representative non-specific interactors in presence or absence of cross-linker.

Protein	Negative control		Pro283		Ser283	
	No cross-linker	With cross-linker	No cross-linker	With cross-linker	No cross-linker	With cross-linker
NP (bait)	0	0	392	345	494	689
Myosin-9	855	7	804	114	736	158
Plectin	259	0	153	5	180	0
Actin	118	28	117	57	99	56
Myosin-14	142	0	148	8	130	10
Myosin-10	128	2	119	20	109	27
40S ribosomal	63	72	65	61	62	66
60S ribosomal	94	64	83	60	86	65

With the first successful proteomics experiment we are now prepared for a future systematic analysis of individual host factors required for stabilization of the biophysically defective adaptive Pro283 NP variant. The ongoing efforts include analyzing NP interactome in perturbed host proteostasis environments and evaluating the differences observed for adaptive Pro283 and more biophysically stable Ser283 variants.

A.4   Conclusion

We have successfully developed a protocol for expressing NP in the host cells and investigating its interactome using mass-spectrometry based proteomics. Using adenovirus for NP gene delivery enables a reproducibly sufficient expression, thus overcoming the main challenges of a transfection-based approach. Our successful pilot mass-spectrometry based proteomics screen validated the co-immunoprecipitation procedure that we developed and provided a first insight into the recruitment of the host chaperones for nucleoprotein folding. We are now well-prepared for a systematic analysis of NP interactome in perturbed host proteostasis environments. We anticipate that applying the adenovirus-based gene delivery system coupled with quantitative mass-spectrometry based proteomics would reveal the individual HSF1-controlled host factors, which are essential for enabling proper folding of destabilized adaptive Pro283 NP variant.

A.5   Materials and Methods

Plasmids. HDM_FLAG_Aichi68_NP_IRES_mCherry nucleoprotein expression plasmid used for transient transfection was a kind gift from Professor Jesse D. Bloom (Fred Hutchinson Cancer Research Center, Seattle WA U.S.A.). Plasmids pAd-CMV/V5-DEST and pENTR1A, used for construction of NP-encoding adenovirus were purchased from Thermo Fisher.

Antibodies. Antibodies used were as follows: mouse monoclonal anti-HA (Thermo Fisher; 26183), mouse monoclonal anti-β-actin (Sigma; A1978), rat monoclonal anti-Hsp70 (Cell Signaling; 4873). IRDyes 800CW goat anti-rat and 680LT goat anti-mouse secondary antibodies were obtained from LI-COR (926-32219 and 926-68020, respectively).

Cell culture. HEK 293A and MDCK$^{dn\text{-}cHSF1}$ cells were cultured at 37 °C in a 5% CO_2 atmosphere in DMEM (CellGro) supplemented with 10% fetal bovine serum (CellGro) and 1% penicillin/streptomycin/ glutamine (CellGro). Construction of MDCK$^{dn\text{-}cHSF1}$ cells was described previously in Chapter 2.

Engineering NP-encoding adenovirus. Influenza nucleoprotein coding sequence was cloned from plasmid HDM_FLAG_Aichi68_NP_IRES_mCherry using Q5 polymerase (NEB) and primer pairs KpnI_NP_N_term and XhoI_HA_C_term_NP to amplify NP with C-terminal HA tag and KpnI_NP_N_term and XhoI_NP_C_term_no_tag to amplify the untagged NP (**Table 4.4**). After PCR NP amplicons were purified using the E.Z.N.A. Cycle Pure kit (Omega) and inserted into pENTR1A vector (Thermo Fisher) using KpnI and XhoI (NEB). The resulting pENTR1A-NP constructs were verified by Sanger sequencing using primers pENTR1A_Sanger_2, NP_Seq_primer_7 and NP_Seq_primer_75. Next, 100 µg of pENTR1A-NP constructs were recombined with the pAd-CMV/V5-DEST vector (Thermo Fisher) using the Gateway LR Clonase Enzyme Mix (Thermo Fisher), following manufacturer's instructions to form pAdDEST-NP plasmid. The resulting pAdDEST-NP constructs were Sanger sequenced using primers CMV_0106, NP_Seq_primer_6, NP_Seq_primer_75, NP_Seq_primer_7.

To generate adenovirus, 1 µg of the resulting AdDEST-NP construct was mixed with 6 µL of polyethylenimine (PEI) (Sigma) and 100 µL of Opti-Mem I (Invitrogen), incubated at room temperature for 20 min and used to transfect a confluent 15 cm plate of HEK 293A cells. The media was replaced 6 h post-transfection. The viral media and cells were collected upon observing a complete cytopathic effect and subjected to 3 × freeze/thaw cycles to extract the

remaining virus from cells. Next the cell debris was removed by centrifugation at 2,000 × g for 15 min, the supernatant was aliquoted and stored at – 80 °C prior to use. 150 µL of the viral transfection supernatant (passage 0, or P0) was used to amplify the generated adenovirus by passaging it on a confluent 15 cm plate of HEK 293A cells. The viral supernatant was harvested, freeze-thawed and clarified as described above. Amplification of the resulting passage 1 (P1) adenovirus was repeated in the same way to yield passage 2 (P2) adenovirus, which was tittered using a plaque assay (described below) and used for subsequent experiments.

Table A.4. Primers used for cloning NP into adenoviral genome and Sanger sequencing.

Primer name	Sequence
KpnI_NP_N_term	AAAAAAGGTACCGCCACCATGGCGTCCCAAGGCAC
XhoI_HA_C_term_NP	AAAAAACTCGAGTTAAGCGTAATCTGGAACATCGTATGGGTAATTGTCGTACTCCTCTGCATTG
XhoI_NP_C_term_no_tag	AAAAAACTCGAGTTAATTGTCGTACTCCTCTG
pENTR1A_Sanger_2	CCACTGCTTACTGGCTTATCG
NP_Seq_primer_7	GGAATCAAGTACTCTTGAACTG
NP_Seq_primer_75	GACGAAAGAAGGAATAAATATC
NP_Seq_primer_6	CAAAGGAGTTGGGACAATG
CMV_0106	CGCAAATGGGCGGTAGGCGTG

Adenovirus plaque assay. HEK 293A cells were seeded at a density 300,000 cells/well in a poly-D-lysine (PDL) coated 12-well plate at 24 h prior to infection with a serial 10-fold dilution of P2 adenoviruses. At 18 h post-infection the media was changed on 0.4% agarose in DMEM. Formation of plaques was observed daily. Once the plaques were observed, typically at 7 – 9 days post infection, 100 µL of 5 mg/mL solution of 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) stain (Research Products International) in PBS was added to wells. After 16 h the plaques were counted manually and viral titer was calculated, accounting for the viral dilution.

Immunoblotting. For transient transfection experiments MDCK$^{dn-cHSF1}$ cells in confluent 10 cm plates were pre-treated with 1 µg/mL aqueous solution of doxycycline (dox) or water (vehicle) at either 37 °C or 39 °C for 18 h. Then the cells were split into 3 wells of a 6-well plate, plating at a density 350,000 cells/well. The cells were transfected at the time of seeding using the 2.2 µg of HDM_FLAG_Aichi68_NP_IRES_mCherry plasmid, pre-incubated with 8.8 µL of TransIt X2 transfection reagent (Mirus Bio) and 215 µL of OptiMem I (Invitrogen). The media was replaced

with fresh DMEM with or without 1 µg/mL dox at 24 h post-transfection. The cells were harvested at 48 h post-transfection.

For adenoviral gene delivery, 1,000,000 MDCK$^{dn-cHSF1}$ cells were seeded in a 10 cm plate and infected with P2 NP-encoding adenovirus at a multiplicity of infection (MOI) of 10 at 24 h after seeding. At 8 h post-infection the cells were split into a 6-well plate at a density 350,000 cells/well and treated with either 1 µg/mL dox or water at either 37 °C or 39 °C for 28 h prior to harvesting.

For immunoblotting analysis, the cells were harvested and lysed using Triton lysis buffer (1% Triton, 200 mM NaCl, 50 mM Tris at pH 7.5, 1 mM EDTA, 1.5 mM MgCl$_2$, and protease inhibitor tablet from Thermo Fisher Scientific). The lysate was cleared by centrifugation at 21,100 × g for 15 min at 4 °C and quantified using the Protein Assay Dye reagent (Bio-Rad). For transfected samples, 120 µg of cell lysates were separated on a homemade 12 % SDS PAGE gel for 2 h at 140 V. For adenovirus-infected samples, 150 µg of cell lysates were separated on a 4 – 12 % NuPAGE Bis-Tris gel (Invitrogen) in a NuPAGE MOPS buffer (Invitrogen) for 2 h at 120 V. The gels were transferred on a nitrocellulose membrane using TransBlot Turbo (Bio-Rad) high MW setting for 8 min. The membranes were then probed for HA-tagged nucleoprotein, Hsp70 and β-actin, and imaged using an Odyssey Infrared Imaging System (Li-Cor).

Co-immunoprecipitation. MDCK$^{dn-cHSF1}$ cells were seeded in 10 cm plates at 1,000,000 cells/plate. After 8 h the cells were infected with NP-encoding adenovirus at a multiplicity of infection (MOI) of 10, using the titers determined by the plaque assay. After 44 h the cells were harvested and washed 3× with PBS. The cells from each of 10 cm plates were resuspended in 10 mL PBS in a 15 mL Falcon tube and cross-linked using dithiobis(succinimidyl propionate) (DSP) (Thermo Fisher Scientific). To cross-link, 5 µL of a freshly made 100 mM DSP solution in DMSO was added to a suspension of cells in 10 mL PBS. Alternatively, 5 µL of DMSO was added to no cross-linker samples. The cells were cross-linked by rotating at room temperature for 30 min, then the cross-linker was neutralized using 200 µL of 1M Tris, pH 8. The cells were rotated at room temperature for 15 min and pelleted by centrifugation at 1,000 × g for 5 min. Next, the cells were lysed using Triton lysis buffer (1% Triton, 200 mM NaCl, 50 mM Tris at pH 7.5, 1 mM EDTA, 1.5 mM MgCl$_2$, and protease inhibitor tablet from Thermo Fisher Scientific), the lysate was cleared by centrifugation at 21,100 × g for 15 min at 4 °C and quantified using the Protein Assay Dye reagent (Bio-Rad). 50 µg of lysate was saved as input samples and 3 µg

lysate was used for immunoprecipitation. First, to eliminate the non-specific binding proteins, the lysate was pre-cleared by incubating with 60 µL of protein A/G agarose beads slurry (Santa Cruz Biotechnology) while rotating at room temperature for 45 min. Then the lysate was incubated with 60 µL of monoclonal anti-HA agarose beads slurry (Sigma) while rotating at 4 °C overnight. The beads were washed 3× with RIPA buffer (10 mM Tris, 1 mM EDTA, 140 mM NaCl, 0.2 % SDS, 0.1 % sodium deoxycholate, 1 % Triton X-100, pH 8) and the residual buffer was removed using a 30 G needle. Nucleoprotein and its interactors were eluted by boiling the beads with 70 µL of denaturing buffer (300 mM Tris, 6% SDS, pH 7.5) for 10 min. The eluent was collected using a 30 G needle and elution was repeated using 70 µL of denaturing buffer. The two elution supernatants were combined and analyzed on a BOLT 4-12% Bis-Tris SDS-PAGE (Invitrogen) in BOLT MOPS SDS buffer (Invitrogen). Silver staining of the gel was performed using a Silver Stain Kit for Mass Spectrometry (Pierce). For Western blotting the gels were transferred to a nitrocellulose membrane using the Trans-Blot Turbo system (Bio-Rad). Membranes were then probed for HA-tagged nucleoprotein and imaged using an Odyssey Infrared Imaging System (Li-Cor).

Mass-spectrometry based interactome investigation. The co-immunoprecipitation was performed as described above. After pull-down, the beads were washed 3× with either RIPA buffer for cross-linked samples, or with Triton lysis buffer for samples without cross-linker. Next the beads were washed 3× with buffer containing 200 mM NaCl, 50 mM Tris at pH 7.5, 1 mM EDTA, 1.5 mM $MgCl_2$ to remove the residual detergent. The beads were dried using a 30 G needle and stored at -20 °C prior to submission to MIT Biopolymers & Proteomics Core for mass-spectrometry analysis.

A.6　Acknowledgements

A.7 References

1.   Phillips, A. M.; Ponomarenko, A. I.; Chen, K.; Ashenberg, O.; Miao, J.; McHugh, S. M.; Butty, V. L.; Whittaker, C. A.; Moore, C. L.; Bloom, J. D., Lin, y.-S., Shoulders, M.D. Destabilized adaptive influenza variants critical for innate immune system escape are potentiated by host chaperones. *PLoS Biol.* **2018,** *16* (9), e3000008.

2.   Richardson, J. S.; Richardson, D. C. Principles and patterns of protein conformation. In *Prediction of protein structure and the principles of protein conformation*, Springer: **1989**; pp 1-98.

3.   Wilman, H. R.; Shi, J.; Deane, C. M. Helix kinks are equally prevalent in soluble and membrane proteins. *Proteins* **2014,** *82* (9), 1960-1970.

4.   Kumar, S.; Bansal, M. Geometrical and sequence characteristics of α-helices in globular proteins. *Biophys. J.* **1998,** *75* (4), 1935-1944.

5.   Richardson, J. S. The anatomy and taxonomy of protein structure. In *Advances in protein chemistry*, Elsevier: 1981; Vol. 34, pp 167-339.

6.   Pavlov, M. Y.; Watts, R. E.; Tan, Z.; Cornish, V. W.; Ehrenberg, M.; Forster, A. C. Slow peptide bond formation by proline and other N-alkylamino acids in translation. *Proc. Natl. Acad. Sci. U.S.A.* **2009,** *106* (1), 50-54.

7.   Pelechano, V.; Alepuz, P. eIF5A facilitates translation termination globally and promotes the elongation of many non polyproline-specific tripeptide sequences. *Nucleic Acids Res.* **2017,** *45* (12), 7326-7338.

8.   Wedemeyer, W. J.; Welker, E.; Scheraga, H. A. Proline cis−trans isomerization and protein folding. *Biochemistry* **2002,** *41* (50), 14637-14644.

9.   Frausto, S. D.; Lee, E.; Tang, H. Cyclophilins as modulators of viral replication. *Viruses* **2013,** *5* (7), 1684-1701.

10.   Liu, X.; Zhao, Z.; Liu, W. Insights into the roles of cyclophilin A during influenza virus infection. *Viruses* **2013,** *5* (1), 182-191.

11.   Chatterji, U.; Bobardt, M.; Selvarajah, S.; Yang, F.; Tang, H.; Sakamoto, N.; Vuagniaux, G.; Parkinson, T.; Gallay, P. The isomerase active site of cyclophilin A is critical for hepatitis C virus replication. *J. Biol. Chem.* **2009,** *284* (25), 16998-17005.

12.   Bosco, D. A.; Eisenmesser, E. Z.; Pochapsky, S.; Sundquist, W. I.; Kern, D. Catalysis of cis/trans isomerization in native HIV-1 capsid by human cyclophilin A. *Proc. Natl. Acad. Sci. U.S.A.* **2002,** *99* (8), 5247-5252.

13.   Liu, X.; Sun, L.; Yu, M.; Wang, Z.; Xu, C.; Xue, Q.; Zhang, K.; Ye, X.; Kitamura, Y.; Liu, W. Cyclophilin A interacts with influenza A virus M1 protein and impairs the early stage of the viral replication. *Cell. Microbiol.* **2009,** *11* (5), 730-741.

14.   Liu, X.; Zhao, Z.; Xu, C.; Sun, L.; Chen, J.; Zhang, L.; Liu, W. Cyclophilin A restricts influenza A virus replication through degradation of the M1 protein. *PLoS One* **2012,** *7* (2), e31063.

15.   Jeong, S.-J.; Ryo, A.; Yamamoto, N. The prolyl isomerase Pin1 stabilizes the human T-cell leukemia virus type 1 (HTLV-1) Tax oncoprotein and promotes malignant transformation. *Biochem. Biophys. Res. Commun.* **2009,** *381* (2), 294-299.

16.   Kawagoe, S.; Nakagawa, H.; Kumeta, H.; Ishimori, K.; Saio, T. Structural insight into proline cis/trans isomerization of unfolded proteins catalyzed by the trigger factor chaperone. *J. Biol. Chem.* **2018,** *293* (39), 15095-15106.