



HAL
open science

Quantitative Theory of Viral-Immune Coevolution May Be within Reach

Thierry Mora, Aleksandra Walczak

► **To cite this version:**

Thierry Mora, Aleksandra Walczak. Quantitative Theory of Viral-Immune Coevolution May Be within Reach. PRX Life, 2023, 1 (1), pp.011001. 10.1103/PRXLife.1.011001 . hal-04252244

HAL Id: hal-04252244

<https://hal.science/hal-04252244v1>

Submitted on 11 Sep 2024


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

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



Distributed under a Creative Commons Attribution 4.0 International License

Quantitative Theory of Viral-Immune Coevolution May Be within Reach

Thierry Mora and Aleksandra M. Walczak 

*Laboratoire de Physique de l'École Normale Supérieure, CNRS, PSL University,
Sorbonne Université, and Université Paris-Cité, 75005 Paris, France*



(Received 2 June 2023; published 22 August 2023)

Pathogens drive changes in host immune systems that in turn exert pressure for pathogens to evolve. Quantifying and understanding this constant coevolutionary process has clear practical global health implications. Yet its relatively easier accessibility compared to macroevolution makes it a fascinating system to learn about the basic laws of evolution. Focusing on immune-viral evolution, we present an overview of theoretical and experimental approaches that have recently started coming together to build the foundations for a quantitative and predictive coevolutionary theory.

DOI: [10.1103/PRXLife.1.011001](https://doi.org/10.1103/PRXLife.1.011001)

I. INTRODUCTION

Vertebrates, including us, use their adaptive immune system to protect themselves against recurring or endemic pathogens such as influenza and SARS-CoV-2, or to battle chronic infections such as HIV. Understanding and quantitatively predicting when and how pathogens evolve and escape immunity through escape mutations is of great importance for the design of vaccine strategies, diagnostic tools, and treatment. The evolution of pathogens is partly driven by the immune system of the hosts. This happens on several time and spatial scales, from within-host evolution in chronic diseases, to continuous interaction with many hosts at the population level in acute infections. The immune systems, in turn, update their memory repertoires to protect hosts against future infections by the pathogen or its close variants. This leads to a process of coevolution between immune systems and pathogens that spans many scales, from the molecular and cellular interactions between immune receptors and pathogenic epitopes, to the organismal and epidemiological levels (Fig. 1).

Protection against pathogens involves all branches of the immune system working together. The ensemble of antibodies is one branch that has the ability to somatically mutate and adapt rapidly, making it an important element of coevolution occurring on fast timescales. Antibodies, or more generally immunoglobulins, specifically bind to bits of molecules called antigens, and recognize them as either natural to the organism (self-proteins) or foreign. They can either be presented on the surface of B cells (where they are part of the B-cell receptor) or found in a free soluble form (antibodies). This process and the ones that follow can involve interactions with other arms of the immune system (e.g., T cells, cells of the innate immune system). Recognition of foreign antigens also induces migration of these B cells to germinal centers, where they

somatically hypermutate and are selected for binding to this specific antigen. Along this process, hypermutated B cells are released to the blood stream and secrete soluble versions of their antibody that bind to pathogens and help combat the infection more rapidly. This results in a constantly changing, diverse set of expressed antibodies, termed the repertoire.

Viral-antibody coevolution, in addition to its practical importance for health, offers a test bed for basic questions of evolution and coevolution. As is becoming clear in many fields, organisms do not evolve in steady conditions, competing only with members of their own species. The outcome of evolution is shaped by environmental conditions, resource availability, and surrounding species. Viral-immune evolution occurs on fast timescales, with both the viral population and the B-cell repertoires updating their diversity within many weeks. The rapidity of the process, combined with relative ease of sampling compared to other systems, and in some cases the well-defined perimeter where the interactions take place, allows us to test and develop quantitative evolutionary theories.

Quantitative evolutionary theories have been around for over 100 years, epidemiological models of susceptible-infected and recovered (SIR) populations close to 100 years [2], and theories of host-pathogen coevolution date back at least 30 years [3]. In recent years, in particular in the wake of the COVID-19 pandemic, troves of genomic data and ever more powerful parallelized biophysical assays have offered new opportunities to test these theories experimentally, and to refine quantitative models. We describe how these advances in theoretical models and in experimental and population data have all contributed to our quantitative understanding of coevolutionary processes. Yet, taken separately, these amazing feats of technology and imagination often fall short of providing a predictive theory of evolution. While evolution is clearly driven by rare events, coevolution does not allow for all possible subsequent events, and it may significantly limit actual evolutionary trajectories. Making concrete statistical predictions about future dominant strains and estimating the errors and timescales for the reliability of these predictions is now within reach. Bringing together data, quantitative controlled experiments, and theory, we can change our understanding of

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

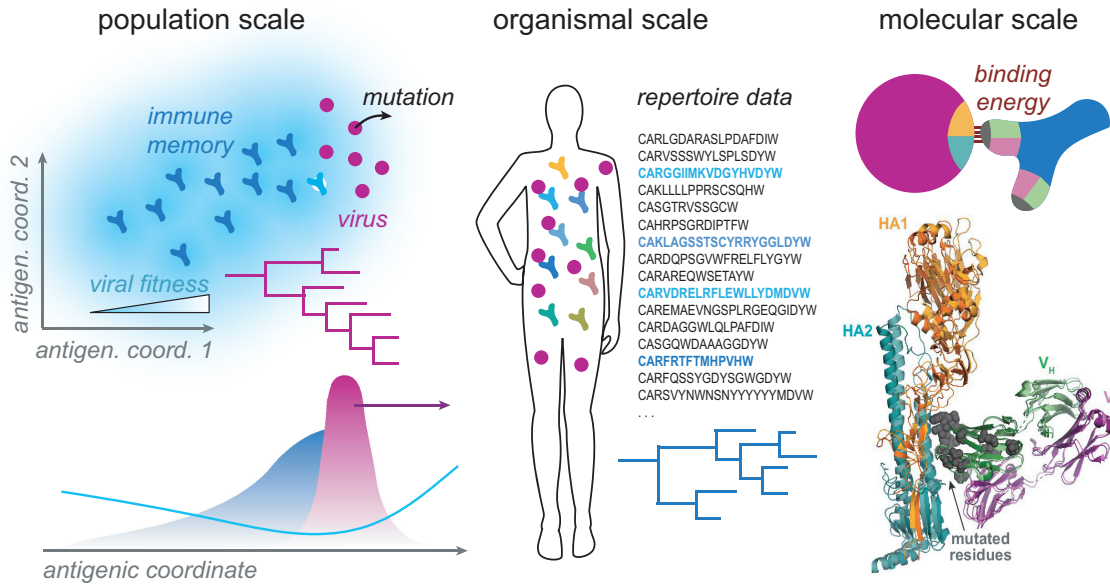


FIG. 1. Coevolution from the population to the molecular scale. Recognition starts at the molecular level through the specific binding of antibodies to bits of the viral proteins called epitopes, illustrated on the right by the co-crystal structure of an antibody fragment and the hemagglutinin protein of influenza (from [1]). At the other end of the population scale (on the left), effective descriptions of antibodies and viruses in a common antigenic space explain how viruses (in purple) are pushed to mutate continually under the pressure of the immune system of its target hosts (in blue), growing faster in regions of the antigenic space where the population is susceptible. At the intermediate scale of the organism (middle), the repertoire of antibodies, represented here by a list of sequences of their third heavy-chain Complementarity Determining Loop, evolves under the selective pressure of successive acute infections (or of a chronic infection) to optimize protection and to prepare the organism for future challenges. Only a small fraction of the whole repertoire (shown in blue) is specific to the viral epitope, and is structured into distinct lineages.

antibody-viral coevolution and propose a predictive quantitative theory of coevolution.

II. VIRAL-ANTIBODY COEVOLUTIONARY PROCESS

Viruses follow basic Darwinian evolutionary rules. They mutate randomly, and their reproductive success depends on their ability to survive in hosts or to infect new ones. The viral capsid envelope and other externally displayed proteins (such as the spike protein in SARS-CoV-2) breach the host’s protective layers, allowing the virus to hijack host cells to reproduce. For this reason, host immune systems target these proteins. Upon successful recognition of a viral protein as a foreign protein, B cells are recruited to germinal centers where the receptor version of their antibody gets hypermutated and selected for better binding to this specific antigen [4,5]. The affinity maturation process goes through a few rounds of mutation and selection [4,6,7], leading to long-lived plasma cells producing highly efficient neutralizing antibodies. At the same time, B cells are also released into the periphery at intermediate stages as plasma or memory B cells [8]. As a result, the outcome of affinity maturation is a set of B cells with a range of binding affinities [9], and not just the best binders for this specific antigen.

Some antibodies are also cross-reactive, meaning that they can recognize many similar antigenic strains of the virus. Both cross-reactivity and the diversity of antibodies encoded by memory B cells guarantees broad protection against the viral strain as well as many of its antigenic variants. For this reason, the virus needs to accumulate enough mutations on its

antigenic sites to avoid recognition by the immune systems of previously infected hosts. In the case of acute infections, it needs these escape mutations to be able to reinfect individuals. Without these escape mutations, it can and does infect susceptible individuals, but after some time a sufficient fraction of the population develops protecting antibodies (herd immunity), making escape mutations necessary for the virus to continue spreading. A similar process occurs for chronic infections such as HIV, but within a single host, and on a faster timescale. The speed of these antigenic substitutions depends on the mutation rate of the virus, as well as on the strength of selection exerted by the hosts’ immune systems.

III. PHENOMENOLOGICAL THEORIES

Most theories of immune-viral coevolution start with dynamical models of susceptible, infected, and recovered (SIR) compartments of hosts [2], repeated for each antigenic strain. The most basic version of this model considers strains in a one-dimensional discrete arrangement called the stepping-stone model, where viral strains can mutate into their neighbors along that single antigenic dimension [3], with possible cross-reactivity (also called cross-immunity) protecting recovered individuals against neighboring strains [10].

In their linearized, deterministic version, these models admit traveling-wave solutions along the antigenic axis that fall into the same class as the Fisher-Kolmogorov-Petrovsky-Piskunov (Fisher-KPP) equation of reaction diffusion [11]. Viruses at the front of the wave enjoy a susceptible host population, giving them a growth advantage over strains in the

bulk of the wave, which suffer from herd immunity. As they infect more individuals, these strains become less and less fit, and are in turn outcompeted by more advanced variants along the antigenic axis. This leads to a nonequilibrium steady state where the average fitness of the viral population remains constant, in a striking realization of the “Red Queen” hypothesis, where the fitness benefits obtained by escape mutations are continually compensated by building herd immunity. Adding cross-immunity may yield nonlinear solutions of oscillating waves, leading to clustered viral populations and periods of quiescence [10,12].

Since these early works, other theories of immune-viral coevolution [13–15] have assumed that viral evolution operates in a different regime, where small-number (demographic) noise at the most advanced tip of the wave drives its motion, giving a striking example of how macroscopic behavior can be determined by the statistics of rare events [16–20]. The crossover between this stochastic regime and the Fisher-KPP regime is governed by the range of cross-reactivity [21].

The two regimes lead to very different predictions for the rate of adaptation and its dependence upon the antigenic mutation rate and host population size [21]. In the regime of small cross-reactivity, the virus spreads in an essentially susceptible population, and the wave falls in the Fisher-KPP regime. The stochastic regime is achieved at large cross-reactivities, and very small fitness differences lead to a large advantage to “jumping before the crowd” and infecting new hosts. The two regimes correspond to distinct predictions for the evolution of non-antigenic traits such as the mutation rate itself, or the virulence of the pathogen. In the stochastic regime, viruses strive to maximize their basic reproductive number, while in the Fisher-KPP regime they tend to maximize the speed of the wave. Rough estimates seem to place the evolution of influenza and possibly SARS-CoV-2 in the stochastic regime [21].

A limit of these early models is the assumption that the antigenic space is one-dimensional. Dating back to the work of Perelson and collaborators [22], physicists have proposed models of coupled nonlinear equations in high-dimensional space to describe the effective ecology of repertoires, and to predict their response to immune challenges. Within this description, antigens and antibodies coexist in the same space, and their ability to interact depends on their mutual distance. The cross-reactivity range defines the distance within which an antibody can recognize an antigen. In this view, antibodies “cover” the space of antigens with cross-reactivity hyperspheres. Hemagglutination inhibition (HI) assays performed in the mid-2000s [23] measured the ability of different animal sera to neutralize influenza viruses found in human populations across the second part of the 20th century. The neutralization trajectory could be computationally mapped onto a low-dimensional trajectory, justifying the approximation made by many phenotypic models.

With these observations in mind, models of antigenic waves were extended to multiple or even infinite dimensions (infinite allele model), leading to mathematical developments with direct biological relevance, in particular about the predictability of viral evolution. The theory predicts that the immune pressure pushes the viral wave forward, channeling its inertial course into an effectively one-dimensional track

[24], but the direction of this evolutionary trace diffuses rotationally in antigenic space [15,25]. This observation may be used to retrospectively explain and justify the assumption that antigenic evolution is low-dimensional. However, the rotational diffusion limits the ability to predict future strains, especially in high dimensions.

Accounting for the finite size of the population of infected hosts leads to the possibility that the viral population goes to extinction, or splits into two antigenically independent lineages [14]. The latter happens when two viral strains at the tip of the wave jump away from the immune systems, but in different directions, and far enough to be outside the cross-reactivity range of each other. This effect, which is reminiscent of the split between the Yamagata and Victoria strain of influenza B in 1983 [26], can only occur in multiple dimensions, but is predicted to be a very rare and essentially unpredictable event, with exponential sensitivity to the model parameters [14,15]. The Red Queen state of constant escape is not asymptotically stable, but it is an incredibly long-lived transient [14]. This result proposes a more subtle interpretation of the neutralization data, where the emergence of an effective low-dimensional coordinate for viral escape is not guaranteed but results from the parameters of the coevolutionary process.

In summary, existing models show the importance of dimensionality for coevolutionary trajectories and their stability. For this reason, estimating the dimensions of this effective phenotypic space is an important challenge for understanding the space and span of possible coevolutionary trajectories and will help connect the evolution of real viruses to these models.

IV. POPULATION LEVEL VIRAL DATA

What does the evolution of real viruses look like? Clearly the impact of viral evolution on our lives is more complex than predicted by theoretical models described above, due to geographic effects [27–30], seasons [31], immune history [32–34], and social behavior [35,36]. Tracking viral evolution based on sampled influenza, HIV, and SARS-CoV-2 data has become an important endeavor for both global health and policy making [37–41], as well as understanding coevolution [42,43].

Since viruses evolve continuously by acquiring unique sets of mutations, these mutations can be used as markers to track evolution and disease spreading. Tracking viral genomes and reconstructing their phylogenies allows us to identify how viral populations spread in host populations, and how and where epidemics arise. Bedford and Neher developed the comprehensive NextStrain open source project [44], which continually updates publicly available pathogen genome data from many species, and it offers analysis and interpretation tools [45]. Collecting data and making them easily accessible is a necessary step for understanding evolutionary processes.

Previously, Neher *et al.* used tree branching statistics to successfully predict the future success of influenza strains [46]. In parallel efforts, Łuksza and Lässig used fitness models based on physical quantities such as protein stability and antigenic binding to predict the evolution of influenza [47]. Such approaches, although currently limited to a one-year prediction window, are now used by global health agencies to

inform influenza vaccine design. These two models focus on different effective signatures of viral evolution: Neher *et al.* [46] exploit signatures of natural selection encoded in the branching structure of viral lineages; Łuksza and Lässig [47] integrate information about viral stability and antigenicity to build a predictive fitness model. Incidentally, similar fitness models generalize well to a different immune context where coevolution plays a similar conceptual role: the arms race between tumor cells and the T cells that eliminate them. These models are able to predict the immunogenicity of neoantigens [48], and to detect the selection or elimination of immunogenic cancer cells by the immune system (immuno-editing) [49]. They are now being used in the design of patient-specific vaccines for pancreatic cancer [50]. Similarly, fitness models have been used to show that selection induced by vaccination has driven the antigenic evolution of SARS-CoV-2 [43], or to predict the vulnerability of HIV variants with implications for vaccine design [51–53], even leading to a preclinical trial [54]. Despite these important applications, much work needs to be done to relate these models to more precise descriptions of pathogen-immune interactions, include the feedback of the hosts' immunity on viral evolution, and integrate the events at the molecular and population level into a general theory of pathogen-immune coevolution.

V. IMMUNE REPERTOIRE ANALYSIS

On the immune side, recent advances in the profiling of antibody repertoires by high-throughput sequencing [55] have opened opportunities for quantitative understanding of pathogen-immune interactions and coevolution. In particular, how the immune repertoire is reshaped following infections or reinfections by the same or mutated pathogens is a crucial building block towards building multiscale models of the coevolutionary arms race between an ensemble of immune systems (from the host population) and a population of evolving pathogens.

The initial diversity of the repertoire is generated by a random process of recombination of the antibody genes, which has been extensively characterized using data-driven biophysical models of recombination and function selection [56,57]. The second process of antibody diversification is through biased hypermutations [58,59], which are generated and selected for better binding during the affinity maturation process. Hypermutations create a large number of lineages with distinct ancestors, and each lineage is shaped by selection through their ability to recognize its cognate antigen [4]. Understanding and quantifying the underlying evolutionary process within antibody repertoires is still a major challenge. Normative or descriptive theories have been proposed to understand how antibody repertoires self-organize to improve their specificity to the target through affinity maturation [60–62], and efficiently encode the memory of past infections to prepare for the next one [63,64]. However, it has been hard to directly relate these predictions to actual repertoire data.

Disease-specific patient cohorts are now routinely subjected to antibody repertoire sequencing [65–70], with the goal of providing insight into how immunity is shaped and acquired. However, most of these experiments sample whole antibody repertoires from patient blood, which is a mixture

of many lineages stemming from multiple past infections. To correctly describe the evolutionary process and gain interpretability, the bulk antibody repertoire data must first be partitioned into clonal families that share common ancestors, and organized into phylogenetic trees retracing the ordering of hypermutations. Because of the large initial diversity of possible ancestors, lineage clustering is a difficult problem for which recent progress has been made [71–73]. In addition, the specific structure of antibody receptor sequences, with varying lengths, small lineage sample sizes due to limited sampling, and heterogeneous mutation rates, makes the problem of phylogenetic reconstruction different and more difficult than in traditional genetics.

Attempts to characterize and quantify the selection on antibody lineages have been inspired by traditional methods from molecular evolution, either by inspecting the spectrum of observed substitutions [74,75], by analyzing the statistics of the lineage structure and phylogenies [66,76], or by exploiting longitudinal data to quantify evolution dynamically [65–67]. These analyses have returned global signatures of selection that are statistically significant but hard to interpret, calling for new approaches.

For example, typical measures of selection, such as skewed site frequency spectra, abnormal ratios of synonymous to non-synonymous mutations, or tree imbalance, are detected even in the repertoire of healthy individuals [66,77], since these repertoires also contain lineages that have been expanded during past infections. This suggests that we should refine our null hypotheses of what makes a normal healthy repertoire, beyond the standard neutral model of evolution. A detailed analysis [73] of B-cell repertoire data sets from healthy people [78] shows that evolutionary properties of the lineages do not depend on the global properties of the ancestral sequence, such as its junctional length or generation probability, suggesting that the selection process in germinal centers treats every sequence in the same way. Estimated selection pressures (through ratios of synonymous to nonsynonymous mutations along the trees) span two orders of magnitude, suggesting a broad range of selection forces that go beyond simply purging weak binders (purifying selection). These results based on human data are in agreement with the direct observation of diversity in mouse lineages [6], where the binding properties of cells exiting germinal centers is large.

Coming back to the analysis of repertoires in infected individuals, progress has been made possible through the longitudinal analysis of lineages across multiple time points. This has allowed us to detect [66] and subsequently validate [79] antigen-specific lineages, or to demonstrate the accumulation of hypermutations with time in acute and chronic infections [80]. Combining longitudinal data with phylogenetic analyses has also helped us to go beyond traditional genetics approaches by using a propagator-based approach originally designed to study viral evolution [81]. Applied to repertoire data from HIV patients, this method identified which part of the antibody sequence was selected [77], and it demonstrated the existence of clonal interference within antibody lineages. In this regime of clonal interference, the best antibodies have a selective advantage driving the growth of their relative frequencies, but they are constantly challenged by new variants before they have the time to invade the population. These

results reveal a complex dynamics with a continual production of ever better antibodies competing with each other, maintaining a large clonal diversity and potentially slowing down adaptation.

Despite these advances, repertoire analyses are generally hampered by the fact that the antigenic targets of antibodies are mostly unknown, making their practical use limited. New experimental techniques and computational tools are needed to identify responding or antigen-specific lineages directly from repertoire data, using either longitudinal data [79] or large cohorts of donors infected with a common disease [82].

VI. COEVOLUTION

A recently published HIV data set was the first to track in depth both antibody repertoires and viral strains in the same patients [83]. While initial analysis does not seem to reveal any signatures of coevolution, pulling the data from different patients and statistically correcting for different sampling depths shows significant anticorrelation between viral and antibody turnover in all patients [84]. This suggests that fast viral evolution is accompanied by a slowdown in the change of the composition of the immune repertoire. This is then followed by a fast immune change when the viral population is stable. The results may appear surprising, since we expect the immune system to track the virus, so that their evolutions should be correlated. However, simple string models such as those proposed earlier for HIV evolution [85] (see below) show that both this initial expectation and the observed anticorrelations are correct, depending on the sampling timescale: if it is of the order of the switching time, as it is in the experiments, we observe negative correlations, since both the viral and immune populations experience the largest selection pressure, hence the largest turnover, when the other population is stably adapted, whereas for longer sampling timescales we would observe positive correlations.

Despite this exception and the weak coevolutionary signal that could be extracted from it, we lack more longitudinal data sets of repertoires and infecting viruses over long timescales. Such experiments would allow us to test coevolution theories quantitatively at the genetic level, with the hope of relating specific viral mutations to specific changes of the repertoire.

Coevolution involves the molecular interaction of viral proteins with immune receptors. While for many purposes a phenomenological description suffices, the molecular nature of these interactions matters, since certain regions of viral proteins are under stronger stability constraints, making them more conserved and more stable targets for the immune system. Conserved regions are usually harder to access from a conformational standpoint, and antibodies primarily target strain-specific and variable regions of the virus [86]. However, antibodies that do neutralize conserved regions are likely to neutralize more than one viral strain. They are known as broadly neutralizing antibodies (BnAbs), and they emerge as a result of natural affinity maturation [87]. Yet this happens rarely and usually after long periods of coevolution, such as in the case in individuals infected with HIV.

String models have been introduced to describe the viral protein-antibody interface as strings of interacting amino

acids. Despite their simplicity, these models have been successfully used to understand viral-antibody coevolution [85], and to explain how broadly neutralizing antibodies emerge during affinity selection [88,89]. Using time-shifted neutralization assays in HIV-infected patients [90], fixation of a lineage was shown to be determined by competition between circulating antibody lineages, and BnAbs were found to be more likely to emerge when confronted with diverse viral populations. String models have also helped design vaccination schedules that force affinity maturation into a regime where antibody evolution is focused on the conserved region of the virus (rather than its antigenic sites, which the virus can mutate without much harm), leading to the generation of antibodies with broader affinity [62,91–93]. Using string models combined with data from humans who received COVID vaccines, Yang *et al.* [94] have shown that feedback from newly produced antibodies that transport antigens more efficiently to germinal centers and mask existing immunodominant epitopes alter the recall response and can allow for the generation of cross-reactive antibodies, altering the path of coevolution.

VII. IMMUNE-ANTIGEN INTERACTION AT THE MOLECULAR LEVEL

Quantifying immune-pathogen interaction at the molecular level remains an open challenge in understanding their coevolution. The general problem is to predict binding affinity, or possibly a more functional readout such as antibody neutralization, from the pair of sequences of the antibody on the one hand, and of the pathogenic protein on the other hand.

Many experiments exist to get insight into the molecular aspects of antibody-antigen interactions in a high-throughput manner, as recently reviewed in [95]: titration-based binding assays based on yeast display (Tite-Seq), which give the binding affinity of antibody-antigen pairs in the physical units of a dissociation constant (K_d) [96], neutralization assays [97], epitope-specific immunoprecipitation assays [98], hemagglutination assays [23], and directed evolution experiments [99]. Currently these methods allow for testing either one antibody against many antigens, or one antigen against many antibodies. A lot of effort is currently being put into large high-throughput neutralization and binding affinity assays that would allow us to scan many antibody-antigen pairs at a time.

While these methods may seem similar, they probe different phenotypic properties, such as neutralization potency or binding affinity. They also have different scopes: directed evolution experiments aim to find the best binding antibodies [99], or to predict the future course of viral evolution [100]. Yet affinity maturation is known to keep more than just the best binders [6], and tracing evolutionary trajectories while maintaining antibody diversity is essential for describing how coevolution works. High-throughput assays such as Tite-Seq measure the phenotypic properties of a large library of variants. The choice of the library itself may be different across methods: one may want to assay all possible single mutation variants of the virus (deep mutational scan) to get a complete but local map of mutational effects [101],

or explore all possible intermediates between an initial and an evolved variant, either of the antibody [1] or the viral [102] sequence. Combining global and local exploration is essential to map out the constantly changing coevolutionary landscape.

The application of Tite-Seq to libraries of anti-influenza antibody variants has highlighted the effects of epistatic (non-additive) interactions between mutations at many orders, and quantified the constraints that epistasis puts on the evolution of broadly neutralizing antibodies [1]. Applied to influenza and SARS-CoV-2 evolution, deep mutational scans based on Tite-Seq have shown that the receptor-binding domain (RBD) of the spike protein can support many mutations while maintaining its affinity for ACE2 [101], and they have been used to assess monoclonal antibodies for their robustness against antigenic escape, as measured by their affinity to all single mutants of RBD [103]. Antibody binding measurements of all possible intermediates between the Wuhan strain and Omicron BA.1 have further revealed the diversity of escape strategies available to the virus, as well as the importance of compensatory mutations [102].

While the limiting factor for quantifying antibody-antigen evolution has been mostly experimental, computational scientists have been developing approaches to generalize existing measurements to predict new interacting pairs. The traditional approach based on protein costructures [104] has proven difficult to generalize into a high-throughput assay, although as more data sets become available [105] machine learning approaches are showing potential for generalization. Obtaining a structure does not directly give information about binding affinity, let alone neutralization, and many models aim to bypass the structural step. Some of these efforts have been powered by machine learning techniques [106], but simpler sequence-based biophysical models of residue preferences and pairwise interactions, such as the string models described above, offer ways of integrating sequencing data sets that are easier to obtain. These models are often trained on deep mutational scan data [107]. Since deep mutational scans explore only single single mutation variants, they offer only local exploration, making it hard to learn models that generalize to completely unseen antibodies or antigens. Combining different experimental techniques with collected data samples that explore local and global mutations is needed to allow for generalization [108].

Mutations first influence binding, however protection is better related to neutralization [97]. Understanding the coevolutionary landscape beyond simple directed evolution experiments and linking binding and neutralization assays is needed to predictively describe antigen-antibody coevolution. The results of these assays are huge data sets, and there is a strong need for theory to make sense of them and to integrate them into predictive models of immune-pathogen coevolution, grounded in their biophysical interactions.

VIII. CONCLUSIONS

The coevolution of immune repertoires and viruses proceeds on a complex, rugged binding-affinity landscape, with potential barriers limiting the availability of evolutionary paths. From the point of view of either a single viral strain or a single antibody, this landscape is constantly changing as the other component is evolving, creating a dynamic “seascape” [109]. Both the viral and antibody binding landscapes can be mapped out, at least locally, thanks to a combination of molecular evolution binding and neutralization assays and by sampling coevolving immune repertoires and viruses in individuals. Quantitatively predicting and understanding the evolution of both the immune and viral components, and learning how to manipulate this coevolution, could help to design vaccines with a broad cross-reactivity (e.g., universal influenza or corona virus vaccines, or HIV vaccines).

Evolution happens by small molecular changes over long times, leading to new functional solutions. These steps are mutations, insertions, or deletions of nucleotides or segments of nucleotides. Most modifications are deleterious to function, and most of these are selected against. As a result, not all paths allowed by local evolutionary moves can be implemented in an evolving organism. For example, for a binding site with L amino acids, not all $19 \times L$ amino acid changes will be viable for the next mutational step. The fact that evolution is really a coevolution between molecules, viral proteins, and immune receptors further restricts the space of realized moves.

We are at a point where, by combining data, laboratory experiments, and theory, we can try to estimate the space of possible evolutionary trajectories and figure out what are the right measures in which to quantify this space. This does not mean we will be able to deterministically foresee the future, but we can put bounds on the scale and timescale of our uncertainty. This theory should give estimates about the diversity and overlap of circulating strains as a function of time (with confidence intervals), predict the probability of specific immune repertoires recognizing a given strain, and put a likelihood on how surprised we should be to see a very diverse strain appear supposedly “out of nowhere.”

From a practical public health perspective, a transient state that persists for millions of generations can be viewed as stable. Current phenomenological models of coevolution give us predictions for how stable the coevolutionary coexistence of the viral and host populations is, but they usually ignore the heterogeneity of the recognition landscape. One important remaining challenge is to figure out in what regime particular viruses are in order to be able to assess the likelihood that they split into distinct substrains, or eventually go extinct, as a function of their sequence. To do that, we need to better calibrate these phenomenological models using data—both population data sampled from individuals, and molecular data from evolutionary experiments. The goal of having a quantitative falsifiable theory of coevolution is within our reach.

[1] A. M. Phillips *et al.*, Binding affinity landscapes constrain the evolution of broadly neutralizing anti-influenza antibodies, *eLife* **10**, e71393 (2021).

[2] W. Kermack and A. McKendrick, Contributions to the mathematical theory of epidemics, *Bltm. Mathcal. Biol.* **53**, 33 (1927).

- [3] A. Sasaki, Evolution of antigen drift/switching: Continuously evading pathogens, *J. Theor. Biol.* **168**, 291 (1994).
- [4] G. D. Victora and M. C. Nussenzweig, Germinal centers, *Annu. Rev. Immunol.* **30**, 429 (2012).
- [5] A. K. Chakraborty, A Perspective on the role of computational models in immunology, *Annu. Rev. Immunol.* **35**, 403 (2017).
- [6] J. M. J. Tas *et al.*, Visualizing antibody affinity maturation in germinal centers, *Science* **351**, 1048 (2016).
- [7] F. Weisel and M. Shlomchik, Memory B cells of mice and humans, *Annu. Rev. Immunol.* **35**, 255 (2017).
- [8] M. Akkaya, K. Kwak, and S. K. Pierce, B cell memory: Building two walls of protection against pathogens, *Nat. Rev. Immunol.* **20**, 229 (2020).
- [9] K. G. C. Smith, A. Light, G. J. V. Nossal, and D. M. Tarlinton, The extent of affinity maturation differs between the memory and antibody-forming cell compartments in the primary immune response, *EMBO J.* **16**, 2996 (1997).
- [10] J. R. Gog and B. T. Grenfell, Dynamics and selection of many-strain pathogens, *Proc. Natl. Acad. Sci. USA* **99**, 17209 (2002).
- [11] R. A. Fisher, The wave of advance of advantageous genes, *Ann. Eugen.* **7**, 355 (1937).
- [12] K. Koelle, S. Cobey, B. Grenfell, and M. Pascual, Epochal evolution shapes the phylodynamics of interpanemic influenza A (H3N2) in humans, *Science* **314**, 1898 (2006).
- [13] I. M. Rouzine and G. Rozhnova, Antigenic evolution of viruses in host populations, *PLOS Pathogens* **14**, e1007291 (2018).
- [14] L. Yan, R. A. Neher, and B. I. Shraiman, Phylodynamic theory of persistence, extinction and speciation of rapidly adapting pathogens, *eLife* **8**, e44205 (2019).
- [15] J. Marchi, M. Lässig, A. M. Walczak, and T. Mora, Antigenic waves of virus—immune coevolution, *Proc. Natl. Acad. Sci. USA* **118**, e2103398118 (2021).
- [16] L. S. Tsimring, H. Levine, and D. A. Kessler, RNA Virus Evolution via a Fitness-Space Model, *Phys. Rev. Lett.* **76**, 4440 (1996).
- [17] M. M. Desai, D. S. Fisher, and A. W. Murray, The speed of evolution and maintenance of variation in asexual populations, *Curr. Biol.* **17**, 385 (2007).
- [18] I. M. Rouzine, É. Brunet, and C. O. Wilke, The traveling-wave approach to asexual evolution: Muller's ratchet and speed of adaptation, *Theor. Pop. Biol.* **73**, 24 (2008).
- [19] É. Brunet, I. M. Rouzine, and C. O. Wilke, The stochastic edge in adaptive evolution, *Genetics* **179**, 603 (2008).
- [20] B. H. Good, I. M. Rouzine, D. J. Balick, O. Hallatschek, and M. M. Desai, Distribution of fixed beneficial mutations and the rate of adaptation in asexual populations, *Proc. Natl. Acad. Sci. USA* **109**, 4950 (2012).
- [21] V. Chardès, A. Mazzolini, T. Mora, and A. M. Walczak, Evolutionary stability of antigenically escaping viruses, [arXiv:2304.11041](https://arxiv.org/abs/2304.11041).
- [22] L. A. Segel and A. S. Perelson, Shape space: An approach to the evaluation of cross-reactivity effects, stability and controllability in the immune system, *Immunol. Lett.* **22**, 91 (1989).
- [23] D. J. Smith, Mapping the antigenic and genetic evolution of influenza virus, *Science* **305**, 371 (2004).
- [24] T. Bedford, A. Rambaut, and M. Pascual, Canalization of the evolutionary trajectory of the human influenza virus, *BMC Biol.* **10**, 38 (2012).
- [25] J. Marchi *et al.*, Size and structure of the sequence space of repeat proteins, *PLoS Comput. Biol.* **15**, e1007282 (2019).
- [26] P. A. Rota *et al.*, Cocirculation of two distinct evolutionary lineages of influenza type B virus since 1983, *Virology* **175**, 59 (1990).
- [27] B. D. Dalziel *et al.*, Urbanization and humidity shape the intensity of influenza epidemics in U.S. cities, *Science* **362**, 75 (2018).
- [28] T. Bedford, S. Cobey, P. Beerli, and M. Pascual, Global migration dynamics underlie evolution and persistence of human influenza A (H3N2), *PLOS Pathogens* **6**, e1000918 (2010).
- [29] T. Bedford *et al.*, Global circulation patterns of seasonal influenza viruses vary with antigenic drift, *Nature (London)* **523**, 217 (2015).
- [30] F. Wen, T. Bedford, and S. Cobey, Explaining the geographical origins of seasonal influenza A (H3N2), *Proc. R. Soc. B* **283**, 20161312 (2016).
- [31] V. N. Petrova and C. A. Russell, The evolution of seasonal influenza viruses, *Nat. Rev. Microbiol.* **16**, 47 (2018).
- [32] J. A. Lewnard and S. Cobey, Immune history and influenza vaccine effectiveness, *Vaccines* **6**, 28 (2018).
- [33] H. L. Dugan *et al.*, Preexisting immunity shapes distinct antibody landscapes after influenza virus infection and vaccination in humans, *Sci. Translation. Med.* **12**, eabd3601 (2020).
- [34] T. K. Tsang *et al.*, Investigation of the impact of childhood immune imprinting on birth year-specific risk of clinical infection during influenza A virus epidemics in Hong Kong, *J. Infect. Dis.* **228**, 169 (2023).
- [35] G. D. Luca *et al.*, The impact of regular school closure on seasonal influenza epidemics: A data-driven spatial transmission model for Belgium, *BMC Infect. Dis.* **18**, 29 (2018).
- [36] D. Balcan *et al.*, Multiscale mobility networks and the spatial spreading of infectious diseases, *Proc. Natl. Acad. Sci. USA* **106**, 21484 (2009).
- [37] A. S. Perelson, A. U. Neumann, M. Markowitz, J. M. Leonard, and D. D. Ho, HIV-1 dynamics in vivo: Virion clearance rate, infected cell life-span, and viral generation time, *Science* **271**, 1582 (1996).
- [38] B. Korber *et al.*, Evolutionary and immunological implications of contemporary HIV-1 variation, *Brit. Med. Bull.* **58**, 19 (2001).
- [39] G. Pullano *et al.*, Underdetection of cases of COVID-19 in France threatens epidemic control, *Nature (London)* **590**, 134 (2021).
- [40] K. M. Bubar *et al.*, Model-informed COVID-19 vaccine prioritization strategies by age and serostatus, *Science* **371**, 916 (2021).
- [41] L. Ferretti *et al.*, Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing, *Science* **368**, eabb6936 (2020).
- [42] C. J. R. Illingworth, A. Fischer, and V. Mustonen, Identifying Selection in the within-host evolution of influenza using viral sequence data, *PLoS Comput. Biol.* **10**, e1003755 (2014).

- [43] M. Meijers, D. Ruchnewitz, M. Łuksza, and M. Lässig, Vaccination shapes evolutionary trajectories of SARS-CoV-2, *Biophys.* (2022).
- [44] <https://nextstrain.org/>.
- [45] J. Hadfield *et al.*, Nextstrain: Real-time tracking of pathogen evolution, *Bioinformatics* **34**, 4121 (2018).
- [46] R. A. Neher, C. A. Russell, and B. I. Shraiman, Predicting evolution from the shape of genealogical trees, *eLife* **3**, e03568 (2014).
- [47] M. Łuksza and M. Lässig, A predictive fitness model for influenza, *Nature (London)* **507**, 57 (2014).
- [48] M. Łuksza *et al.*, A neoantigen fitness model predicts tumour response to checkpoint blockade immunotherapy, *Nature (London)* **551**, 517 (2017).
- [49] M. Łuksza *et al.*, Neoantigen quality predicts immunoeediting in survivors of pancreatic cancer, *Nature (London)* **606**, 389 (2022).
- [50] L. A. Rojas *et al.*, Personalized RNA neoantigen vaccines stimulate T cells in pancreatic cancer, *Nature (London)* **618**, 144 (2023).
- [51] R. H. Y. Louie, K. J. Kaczorowski, J. P. Barton, A. K. Chakraborty, and M. R. McKay, Fitness landscape of the human immunodeficiency virus envelope protein that is targeted by antibodies, *Proc. Natl. Acad. Sci. USA* **115**, E564 (2018).
- [52] J. P. Barton *et al.*, Relative rate and location of intra-host HIV evolution to evade cellular immunity are predictable, *Nat. Commun.* **7**, 11660 (2016).
- [53] A. L. Ferguson *et al.*, Translating HIV sequences into quantitative fitness landscapes predicts viral vulnerabilities for rational immunogen design, *Immunity* **38**, 606 (2013).
- [54] D. K. Murakowski *et al.*, Adenovirus-vectored vaccine containing multidimensionally conserved parts of the HIV proteome is immunogenic in rhesus macaques, *Proc. Natl. Acad. Sci. USA* **118**, e2022496118 (2021).
- [55] G. Georgiou *et al.*, The promise and challenge of high-throughput sequencing of the antibody repertoire, *Nat. Biotechnol.* **32**, 158 (2014).
- [56] Q. Marcou, T. Mora, and A. M. Walczak, IGoR: A tool for high-throughput immune repertoire analysis, *Nat. Commun.* **9**, 561 (2018).
- [57] G. Isacchini, A. M. Walczak, T. Mora, and A. Nourmohammad, Deep generative selection models of T and B cell receptor repertoires with soNNia, *Proc. Natl. Acad. Sci. USA* **118**, e2023141118 (2021).
- [58] G. Yaari *et al.*, Models of somatic hypermutation targeting and substitution based on synonymous mutations from high-throughput immunoglobulin sequencing data, *Front. Immunol.* **4** (2013).
- [59] N. Spisak, A. M. Walczak, and T. Mora, Learning the heterogeneous hypermutation landscape of immunoglobulins from high-throughput repertoire data, *Nucl. Acids Res.* **48**, 10702 (2020).
- [60] T. B. Kepler and A. S. Perelson, Cyclic re-entry of germinal center B cells and the efficiency of affinity maturation, *Immunol. Today* **14**, 412 (1993).
- [61] M. Meyer-Hermann *et al.*, A theory of germinal center B cell selection, division, and exit, *Cell Rep.* **2**, 162 (2012).
- [62] S. Wang *et al.*, Manipulating the selection forces during affinity maturation to generate cross-reactive HIV antibodies, *Cell* **160**, 785 (2015).
- [63] O. H. Schnaack and A. Nourmohammad, Optimal evolutionary decision-making to store immune memory, *eLife* **10**, e61346 (2021).
- [64] V. Chardès, M. Vergassola, A. M. Walczak, and T. Mora, Affinity maturation for an optimal balance between long-term immune coverage and short-term resource constraints, *Proc. Natl. Acad. Sci. USA* **119**, e2113512119 (2022).
- [65] K. B. Hoehn *et al.*, Dynamics of immunoglobulin sequence diversity in HIV-1 infected individuals, *Philos. Trans. R. Soc. B* **370**, 20140241 (2015).
- [66] F. Horns, C. Vollmers, C. L. Dekker, and S. R. Quake, Signatures of selection in the human antibody repertoire: Selective sweeps, competing subclones, and neutral drift, *Proc. Natl. Acad. Sci. USA* **116**, 1261 (2019).
- [67] K. B. Hoehn *et al.*, Repertoire-wide phylogenetic models of B cell molecular evolution reveal evolutionary signatures of aging and vaccination, *Proc. Natl. Acad. Sci. USA* **116**, 22664 (2019).
- [68] K. M. Roskin *et al.*, Aberrant B cell repertoire selection associated with HIV neutralizing antibody breadth, *Nat. Immunol.* **21**, 199 (2020).
- [69] C. Kreer *et al.*, Longitudinal isolation of potent near-germline SARS-CoV-2-neutralizing antibodies from COVID-19 patients, *Cell* **182**, 843 (2020).
- [70] S. C. Nielsen *et al.*, Human B cell clonal expansion and convergent antibody responses to SARS-CoV-2, *Cell Host & Microbe* **28**, 516 (2020).
- [71] D. K. Ralph and F. A. M. Iv, Likelihood-based inference of B cell clonal families, *PLoS Comput. Biol.* **12**, e1005086 (2016).
- [72] N. Nouri and S. H. Kleinstejn, Somatic hypermutation analysis for improved identification of B cell clonal families from next-generation sequencing data, *PLoS Comput. Biol.* **16**, e1007977 (2020).
- [73] N. Spisak, T. Dupic, T. Mora, and A. M. Walczak, Combining mutation and recombination statistics to infer clonal families in antibody repertoires, *arXiv:2212.11997* (2022).
- [74] G. Yaari, M. Uduman, and S. H. Kleinstejn, Quantifying selection in high-throughput Immunoglobulin sequencing data sets, *Nucl. Acids Res.* **40**, e134 (2012).
- [75] C. O. McCoy *et al.*, Quantifying evolutionary constraints on B-cell affinity maturation, *Philos. Trans. R. Soc. B* **370**, 20140244 (2015).
- [76] G. Yaari, J. I. C. Benichou, J. A. Vander Heiden, S. H. Kleinstejn, and Y. Louzoun, The mutation patterns in B-cell immunoglobulin receptors reflect the influence of selection acting at multiple time-scales, *Philos. Trans. R. Soc. B* **370**, 20140242 (2015).
- [77] A. Nourmohammad, J. Otwinowski, M. Łuksza, T. Mora, and A. M. Walczak, Fierce selection and interference in B-cell repertoire response to chronic HIV-1, *Mol. Biol. Evol.* **36**, 2184 (2019).
- [78] B. Briney, A. Inderbitzin, C. Joyce, and D. R. Burton, Commonality despite exceptional diversity in the baseline human antibody repertoire, *Nature (London)* **566**, 393 (2019).
- [79] F. Horns, C. L. Dekker, and S. R. Quake, Memory B cell activation, broad anti-influenza antibodies, and bystander activation revealed by single-cell transcriptomics, *Cell Rep.* **30**, 905 (2020).

- [80] K. B. Hoehn *et al.*, Human B cell lineages associated with germinal centers following influenza vaccination are measurably evolving, *eLife* **10**, e70873 (2021).
- [81] N. Strelkova and M. Lässig, Clonal interference in the evolution of influenza, *Genetics* **192**, 671 (2012).
- [82] M. Ruiz Ortega, N. Spisak, T. Mora, and A. M. Walczak, Modeling and predicting the overlap of B- and T-cell receptor repertoires in healthy and SARS-CoV-2 infected individuals, *PLOS Gen.* **19**, e1010652 (2023).
- [83] N. Strauli *et al.*, The genetic interaction between HIV and the antibody repertoire, *bioRxiv* (2019).
- [84] A. Mazzolini, T. Mora, and A. M. Walczak, Inspecting the interaction between human immunodeficiency virus and the immune system through genetic turnover, *Philos. Trans. R. Soc. B* **378**, 20220056 (2023).
- [85] A. Nourmohammad, J. Otwinowski, and J. B. Plotkin, Host-pathogen coevolution and the emergence of broadly neutralizing antibodies in chronic infections, *PLOS Gen.* **12**, e1006171 (2016).
- [86] J. P. Julien *et al.*, Crystal structure of a soluble cleaved HIV-1 envelope trimer, *Science* **342**, 1477 (2013).
- [87] H. X. Liao *et al.*, Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus, *Nature (London)* **496**, 469 (2013).
- [88] C. Wang *et al.*, B-cell repertoire responses to varicella-zoster vaccination in human identical twins, *Proc. Natl. Acad. Sci. USA* **112**, 500 (2015).
- [89] S. Luo and A. S. Perelson, Competitive exclusion by autologous antibodies can prevent broad HIV-1 antibodies from arising, *Proc. Natl. Acad. Sci. USA* **112**, 11654 (2015).
- [90] D. D. Richman, T. Wrin, S. J. Little, and C. J. Petropoulos, Rapid evolution of the neutralizing antibody response to HIV type 1 infection, *Proc. Natl. Acad. Sci. USA* **100**, 4144 (2003).
- [91] S. Wang, Optimal sequential immunization can focus antibody responses against diversity loss and distraction, *PLoS Comput. Biol.* **13**, e1005336 (2017).
- [92] K. G. Sprenger, J. E. Louveau, P. M. Murugan, and A. K. Chakraborty, Optimizing immunization protocols to elicit broadly neutralizing antibodies, *Proc. Natl. Acad. Sci. USA* **117**, 20077 (2020).
- [93] R. S. Ganti and A. K. Chakraborty, Mechanisms underlying vaccination protocols that may optimally elicit broadly neutralizing antibodies against highly mutable pathogens, *Phys. Rev. E* **103**, 052408 (2021).
- [94] L. Yang *et al.*, Antigen presentation dynamics shape the antibody response to variants like SARS-CoV-2 Omicron after multiple vaccinations with the original strain, *Cell Rep.* **42**, 112256 (2023).
- [95] A. Moulana, T. Dupic, A. M. Phillips, and M. M. Desai, Genotype—phenotype landscapes for immune—pathogen coevolution, *Trends Immunol.* **44**, 384 (2023).
- [96] R. M. Adams, T. Mora, A. M. Walczak, and J. B. Kinney, Measuring the sequence-affinity landscape of antibodies with massively parallel titration curves, *eLife* **5**, e23156 (2016).
- [97] B. Dadonaite *et al.*, A pseudovirus system enables deep mutational scanning of the full SARS-CoV-2 spike, *Cell* **186**, 1263 (2023).
- [98] G. J. Xu *et al.*, Comprehensive serological profiling of human populations using a synthetic human virome, *Science* **348**, aaa0698 (2015).
- [99] E. T. Boder, K. S. Midelfort, and K. D. Wittrup, Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity, *Proc. Natl. Acad. Sci. USA* **97**, 10701 (2000).
- [100] J. Zahradnik *et al.*, SARS-CoV-2 variant prediction and antiviral drug design are enabled by RBD in vitro evolution, *Nat. Microbiol.* **6**, 1188 (2021).
- [101] T. N. Starr *et al.*, Deep Mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding, *Cell* **182**, 1295 (2020).
- [102] A. Moulana *et al.*, Compensatory epistasis maintains ACE2 affinity in SARS-CoV-2 Omicron BA.1, *Nat. Commun.* **13**, 7011 (2022).
- [103] T. N. Starr *et al.*, SARS-CoV-2 RBD antibodies that maximize breadth and resistance to escape, *Nature (London)* **597**, 97 (2021).
- [104] I. A. Wilson and N. J. Cox, Structural basis of immune recognition of influenza virus hemagglutinin, *Annu. Rev. Immunol.* **8**, 737 (1990).
- [105] J. Dunbar *et al.*, SAbDab: The structural antibody database, *Nucl. Acids Res.* **42**, D1140 (2014).
- [106] S. Shan *et al.*, Deep learning guided optimization of human antibody against SARS-CoV-2 variants with broad neutralization, *Proc. Natl. Acad. Sci. USA* **119**, e2122954119 (2022).
- [107] J. M. Taft *et al.*, Deep mutational learning predicts ACE2 binding and antibody escape to combinatorial mutations in the SARS-CoV-2 receptor-binding domain, *Cell* **185**, 4008 (2022).
- [108] B. Meynard-Piganeau, C. Feinauer, M. Weigt, A. Walczak, and T. Mora, TULIP—a transformer based unsupervised language model for interacting peptides and T-cell receptors that generalizes to unseen epitopes, *bioRxiv* (2023).
- [109] V. Mustonen and M. Lässig, From fitness landscapes to seascape: Non-equilibrium dynamics of selection and adaptation, *Trends Genet.* **25**, 111 (2009).