

Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing

Timothy A Whitehead, Aaron Chevalier, Yifan Song, Cyrille Dreyfus, Sarel J Fleishman, Cecilia De Mattos, Chris A Myers, Hetunandan Kamisetty, Patrick Blair, Ian A Wilson & David Baker

Nature Biotechnology (2012) doi:10.1038/nbt.2214

Received 27 September 2011 Accepted 12 April 2012 Published online 27 May 2012

Abstract

[Abstract](#) [Introduction](#) [Results](#) [Discussion](#) [Methods](#) [Accession Codes](#) [References](#) [Acknowledgments](#)

[Author Information](#) [Supplementary Information](#)

We show that comprehensive sequence-function maps obtained by deep sequencing can be used to reprogram interaction specificity and to leapfrog over bottlenecks in affinity maturation by combining many individually small contributions not detectable in conventional approaches. We use this approach to optimize two computationally designed inhibitors against H1N1 influenza hemagglutinin and, in both cases, obtain variants with subnanomolar binding affinity. The most potent of these, a 51-residue protein, is broadly cross-reactive against all influenza group 1 hemagglutinins, including human H2, and neutralizes H1N1 viruses with a potency that rivals that of several human monoclonal antibodies, demonstrating that computational design followed by comprehensive energy landscape mapping can generate proteins with potential therapeutic utility.

Introduction

[Abstract](#) [Introduction](#) [Results](#) [Discussion](#) [Methods](#) [Accession Codes](#) [References](#) [Acknowledgments](#)

[Author Information](#) [Supplementary Information](#)

Influenza is a serious public health concern, and new therapeutics that protect against this highly adaptable virus are urgently needed. We recently reported the *de novo* design of two proteins that, after affinity maturation using error-prone PCR, bound with nanomolar affinity to influenza hemagglutinin at a conserved stem epitope that is the target of broadly neutralizing antibodies¹. One of these designed binders, HB80.3, inhibited the pH-induced conformational change necessary for influenza virus infectivity and so was a promising candidate for generating a broad-spectrum antiviral agent against influenza, but additional screening failed to isolate higher-affinity variants. We hypothesized that further improvement of activity could require a combination of multiple small contributions from mutations that might individually be difficult to identify. To identify such sequence variants and obtain a complete map of their contributions to binding in these designed proteins, we extended a recently described approach for mapping binding interfaces using deep sequencing^{2,3} to encompass much larger sets of positions (from 25 to 50 positions, large enough to encompass the entire HB80.3 protein). We generated libraries containing ~1,000 unique single-point mutant variants, and used deep sequencing to determine the frequencies of each point mutant before and after selection for binding. Comprehensive sequence-function landscapes for both designed proteins were generated based on these data, and used to guide the improvement of the design force field and the creation of subtype-specific binders. Combinations of substitutions favored in the binding landscapes yielded high-affinity ($K_D = \sim 1$ nM) variants that bind most group 1 influenza viruses and neutralize H1N1 viruses in cell culture

experiments.

Results

Abstract Introduction Results Discussion Methods Accession Codes References Acknowledgments

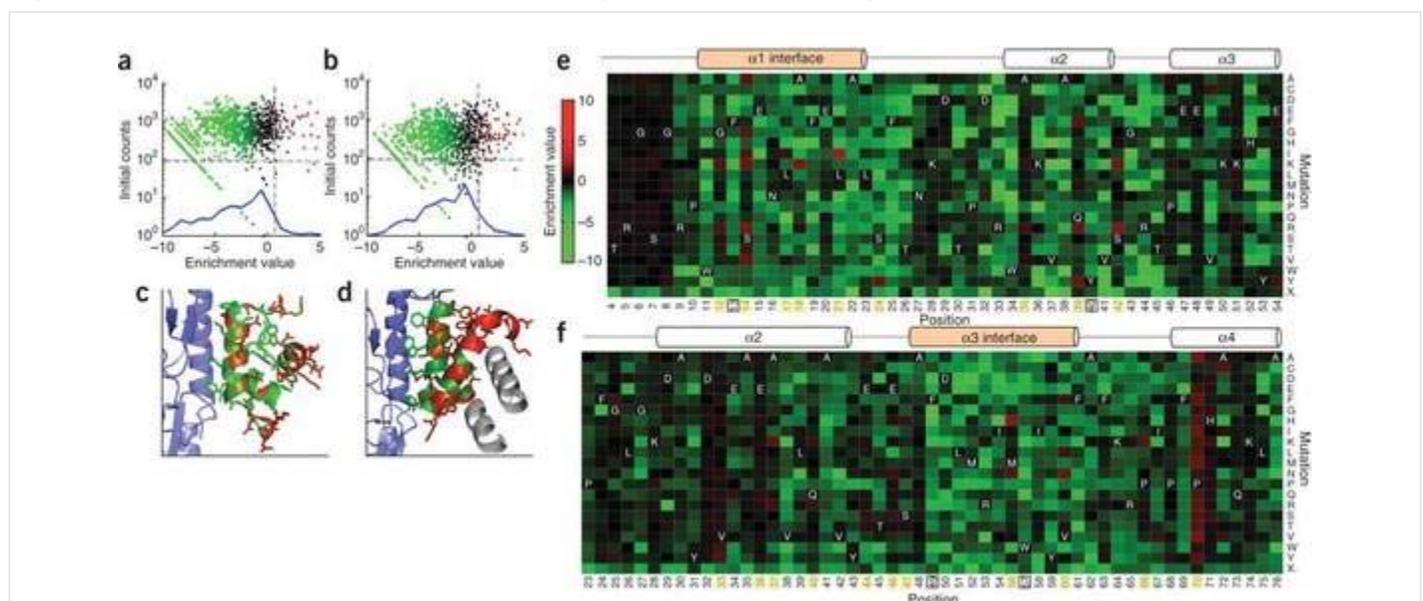
Author Information Supplementary Information

Binding energy landscape mapping

We investigated the contributions to binding of all 51 positions in HB80.3 and of 53 positions (out of 93 possible) surrounding the experimentally determined binding surface in the designed binder HB36.4 (Supplementary Table 1 and Supplementary Fig. 1). To ensure adequate statistics with such a large number of positions and to compensate for short sequencing read lengths, which allow coverage of only a subset of the interrogated positions, we used libraries in which each member contained a single substitution. A complete set of amino-acid variants was generated at each position, and the individual position libraries were then combined. Using yeast display⁴ and fluorescence-activated cell sorting (FACS), we collected populations from each library that bound to either SC1918/H1 (H1) or VN2004/H5 (H5) hemagglutinin subtypes under sorting conditions of varying stringency (details are in Supplementary Fig. 2 and Supplementary Tables 2–3). From each selected population, plasmid DNA was extracted and the mutant genes PCR amplified and then sequenced in two segments using Illumina GA-II 76-bp paired-end deep sequencing.

Analysis of the unselected libraries showed that near-complete sequence coverage was achieved: the HB36.4 library contained 1,053 of the possible 1,061 single amino-acid substitutions, and the HB80.3 library, 1,013 of the 1,021 possibilities. In each selected population, the ~1,000 unique amino-acid sequence variants were sampled with a median depth of coverage of >300 per variant and little sequencing error (Fig. 1a–c, Supplementary Figs. 3–5 and Supplementary Tables 2,3). The median number of DNA reads per population was 1,534,424, and the minimum 1,049,035. In libraries sorted solely for display on the yeast surface, the variant frequencies were surprisingly similar to those in the unselected population, suggesting that even aberrantly folded proteins make it to the surface despite the yeast secretion quality control system, perhaps due to the small size of the displayed proteins (Supplementary Fig. 6).

Figure 1: Sequence-function landscapes of designed influenza-binding proteins.



(a,b) Deep sequencing yields large numbers of independent observations to robustly determine enrichment values in stringent binding selections to the H1 hemagglutinin subtype. Mutations that are heavily depleted are shown in green, whereas beneficial mutations are indicated in red. Horizontal dashed lines indicate 100 sequence counts for unique nonsynonymous substitutions in the library, whereas vertical dashed lines demarcate the enrichment ratio of the starting sequence, showing that most substitutions

are neutral to deleterious. (a, HB80.3 library; b, HB36.4 library). (c,d) Model of H1 hemagglutinin (shown in blue ribbons) bound to HB80.3 (c) and HB36.4 (d). The designed binding proteins are colored by positional Shannon entropy with green indicating positions of low entropy and red those of high entropy. Gray ribbons on HB36.4 indicate positions without deep sequencing data. (e,f) Heat maps representing H1 hemagglutinin-binding enrichment values under stringent binding selection for all possible single mutations in all 51 positions of HB80.3 (e) and in 53/93 positions of HB36.4 (f). Starting residue identities are shown in white font, and the central helix paratope for the design variants is colored in orange in the secondary structure diagrams above the heat maps. Positions with enrichment greater than fourfold are colored yellow and were included in the subsequent designed library and black boxes around positions indicate hot-spot residues in the original designs.

The ratio of the frequencies of a single substitution variant in the selected versus unselected population provides a measure of the effect of the substitution on binding. We refer to the base 2 logarithm of this frequency ratio as the “enrichment value” in the remainder of the text. Under ideal conditions (e.g., free equilibration of fluorescently labeled hemagglutinin among the different clones, equal growth rates of all clones), this measure would be directly proportional to the change in free energy of binding resulting from the substitution. These conditions are not likely to be perfectly met in the experiment, but several lines of evidence suggest that the measure is a reasonable proxy. The enrichment values are nearly identical for synonymous mutations (Supplementary Fig. 7) and correlate with independent affinity measurements on individual variants using yeast surface display titrations (Supplementary Table 4). In experiments in which clones with widely ranging *in vitro* affinities were mixed and then subjected to yeast display selection, the highest-affinity clone rapidly took over the population (Supplementary Fig. 8). Finally, as noted below, the enrichment ratio is broadly consistent with the structures of the designed complexes.

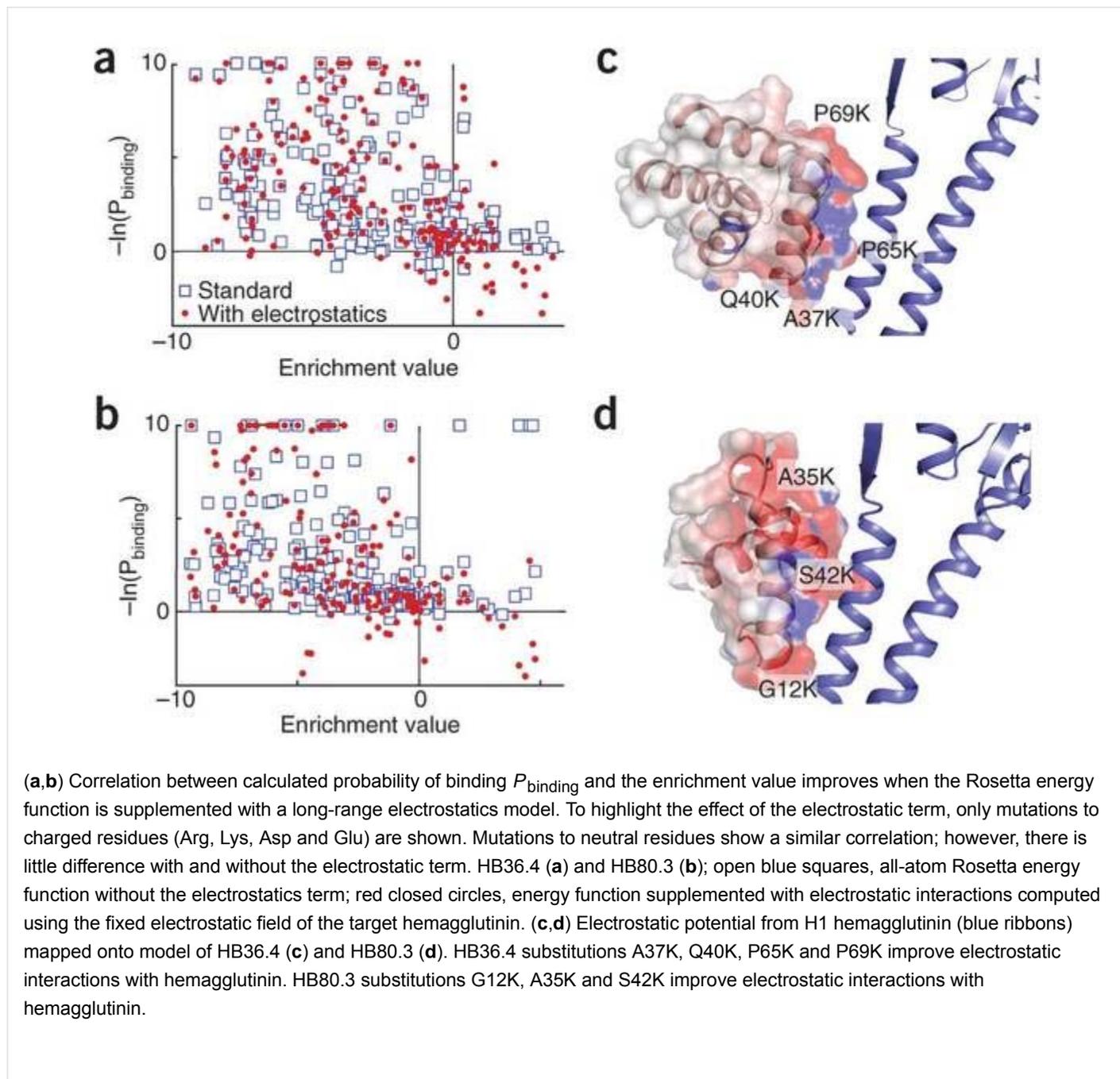
Maps of the enrichment values for H1 hemagglutinin binding of each of the ~1,000 single amino-acid substitutions in HB36.4 and HB80.3 suggest that most substitutions are neutral or deleterious (Fig. 1a,b); the computationally designed interfaces in this respect are similar to naturally occurring interfaces as found in previous large-scale mapping experiments of protein sequence/function^{5, 6, 7, 8}. The positions where very little sequence variation is tolerated are either in the core of the protein or directly at the designed interface (Fig. 1c,d) with the starting designed amino acid being almost always favored (Fig. 1e,f). In HB36.4, few substitutions were tolerated for the binding hotspots Phe49 and Trp57, and, in HB80.3, the hotspot residues Phe13 and Tyr40 are also strongly conserved. Overall, the enrichment values are consistent with the design models of both interfaces and the crystal structure of the HB36.3 interface¹.

Energy function improvement

More detailed analysis of the enrichment values provides a comprehensive view of the binding energy landscapes of computationally designed interfaces, which differ from naturally evolved interfaces in not being optimized by countless generations of natural selection. These data provide an unprecedented opportunity to identify and remedy the shortcomings in the computational model that underlies the design calculations. We tested the energy function used in the design calculations by attempting to recapitulate computationally the experimental maps using a simple model that accounts for the effects of mutations on the free energy of both folding and binding ($P_{\text{binding}} = \text{probability_of_folding} * \text{probability_of_binding_if_folded}$; see Fig. 2 and Online Methods)^{9, 10}. Although the model partially discriminates deleterious substitutions from neutral ones, it does not identify beneficial substitutions (Fig. 2a,b); this result is expected as any substitutions that are favorable according to the design model would have been incorporated in the original design. Many of the newly identified beneficial mutations likely increase electrostatic complementarity at the interface periphery, including substitutions to basic residues in the vicinity of acidic patches on the hemagglutinin surface (e.g., P66K/R on HB36.4 and G12K/R on HB80.3) (Fig. 2c,d). Long-range electrostatics were not modeled in the original design calculations because of difficulties in computationally efficient and accurate modeling of these interactions, and hence these beneficial substitutions were missed. To remedy this shortcoming, we incorporated into the energy function used in the calculations a rapidly computable static Poisson-Boltzmann electrostatics model, which results in improved recapitulation of the beneficial

electrostatic substitutions (Fig. 2a,b) and better overall recapitulation of the experimental results (Supplementary Table 5). The model also improves recapitulation of the free energy changes brought about by mutation in the completely independent Barnase-Barstar complex (Supplementary Fig. 9).

Figure 2: Improvement of computational model by incorporation of long-range electrostatics.

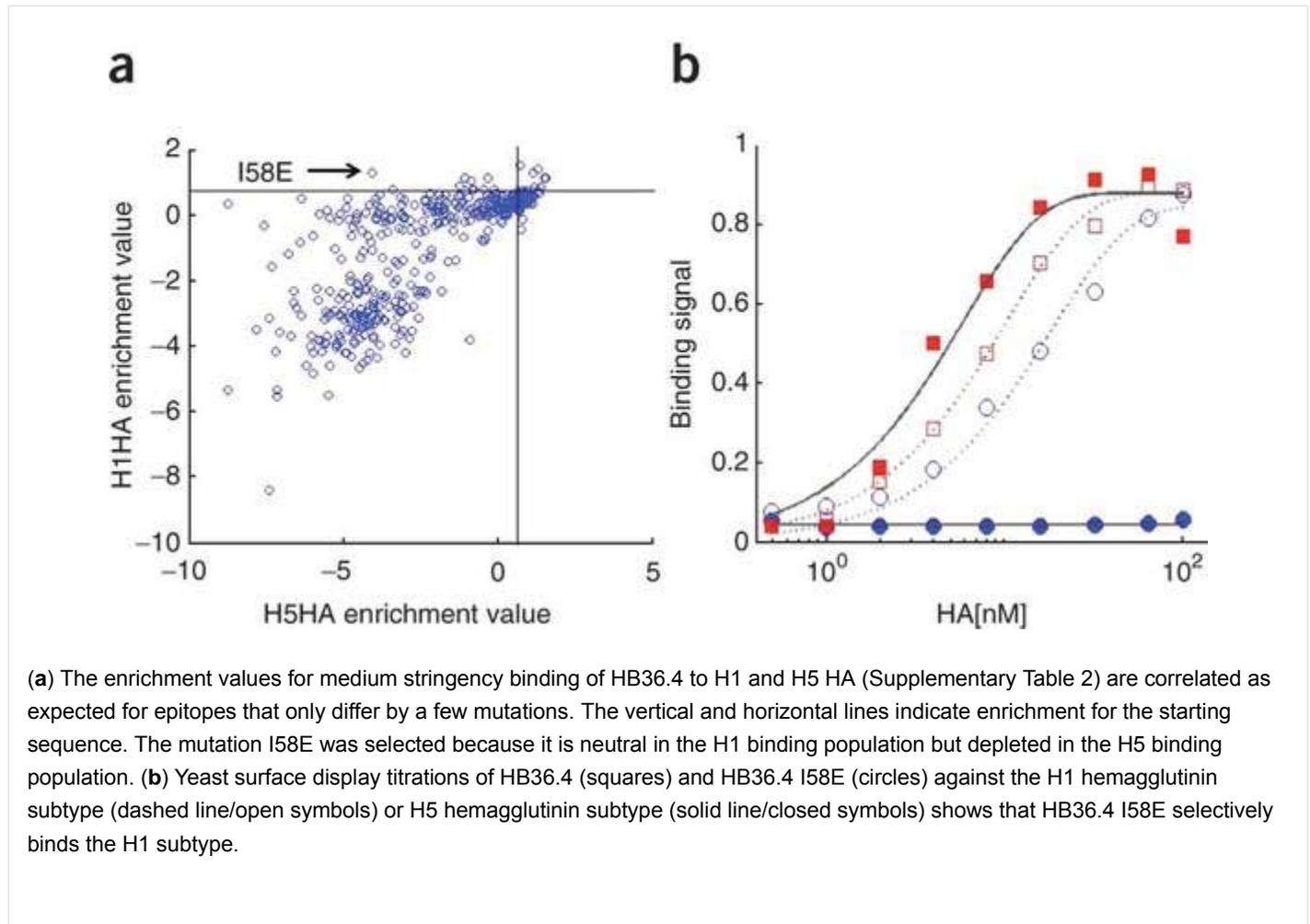


Energy landscape-guided specificity switch

Achieving binding specificity among structurally related ligands has proven challenging in protein engineering; this is typically approached by alternating negative selection steps with positive selection, but negative selection can be problematic, and the iteration can make the approach laborious¹¹. The energetic differences revealed by the experimental maps can be exploited to achieve binding specificity by identifying substitutions that are neutral or enriched in one population and depleted in another. The SC1918/H1 (H1) or VN2004/H5 (H5) hemagglutinin subtypes differ only by a handful of conservative substitutions at the target surface, making engineering for specificity quite challenging.

Comparative analysis of the HB36.4 H1 and H5 hemagglutinin medium-stringency binding maps (Fig. 3a) uncovered the single substitution I58E, which is completely depleted in the H5 binding population, but not at all depleted in the H1 binding population (in the bound complex, position 58 binds close to a region in which H1 and H5 differ; see Supplementary Fig. 10). HB36.4 I58E bound H1 hemagglutinin, but showed no binding of H5 hemagglutinin at the maximum concentration tested, where the net change in specificity is over 30-fold (Fig. 3b; compare open and closed circles). Comparison of the energy landscapes mapped by deep sequencing thus allows reprogramming of interaction specificity, in this case providing a route to the development of subtype-specific influenza binders for clinical diagnosis.

Figure 3: Exploitation of sequence-function landscapes to produce a subtype-specific hemagglutinin binder.



Combining enriched substitutions yields high-affinity binders

The enrichment landscapes also provide a route forward to obtain higher-affinity variants by combining individually small beneficial effects that may not be detectable by conventional directed evolution selections. To investigate whether the substitutions that were enriched in the selections for hemagglutinin binding can be combined to produce higher-affinity binders and whether the contributions of the individual substitutions are additive, we created libraries consisting of 12 variable positions and 4,600,000 unique variants for HB36.4 and 9 variable positions with a total diversity of 300,000 unique variants for HB80.3 by allowing, at each position, the starting residue type and the beneficial substitutions with more than fourfold enrichment (Supplementary Table 6). We carried out Illumina sequencing of the HB80.3 library before and after selection for H1 hemagglutinin binding, and compared the enrichments of each pair of substitutions at the 9 variable positions to those expected if the mutational effects were purely additive. A strong overall correlation was observed between the experimentally determined enrichment of pairs and the prediction based on the effects of the individual mutations (Supplementary Fig. 11), but a statistical model that distinguishes between direct (positions i and j covary) and

indirect (positions i and k covary because both covary with j) covariance using a maximum-likelihood approach found statistically significant covariances between several positions (Supplementary Fig. 12)¹². Because the effects were not strictly additive, we carried out four additional yeast display sorts for increased H1 hemagglutinin binding affinity and slower off-rates and determined the sequences of selected clones in the enriched population. The likelihood of these selected sequences using the maximum likelihood model based on the round 1 deep sequencing data increased when the observed co-variances were included (Supplementary Fig. 13); we anticipate that deep sequencing of more complex libraries followed by model fitting including covariances will allow creation of more active variants in situations where the size of the library makes exhaustive experimental characterization impossible.

A subset of the enriched HB80.3 and HB36.4 variants (Supplementary Tables 7–9) were expressed in *Escherichia coli* with an N-terminal FLAG tag and a C-terminal His tag and purified by affinity chromatography. The binding affinities for hemagglutinin of six of the variants that were soluble and monomeric were determined by surface plasmon resonance. The highest affinity of the HB36 variants, F-HB36.5 (F- denotes an N-terminal FLAG tag), differs at eight positions from the starting sequence and binds SC1918/H1 hemagglutinin with a binding dissociation constant (K_D) of 890 pM, 28-fold lower than HB36.4, and a reduced off-rate (k_{off}) of 0.0015 s^{-1} . The best of the HB80.3 variants, F-HB80.4, which harbors 5 mutations compared to HB80.3 (Supplementary Fig. 14), has a K_D of 600 pM, 25-fold lower than that of HB80.3, and a k_{off} of 0.0022 s^{-1} , tenfold slower than F-HB80.3 (Table 1). Three of the five substitutions in HB80.4 likely improve long-range electrostatics (G12R, A35R, S42R). Incorporation of these three substitutions alone (construct F-HB80.4.1) yields a K_D of 1.2 nM and a k_{off} of 0.0056 s^{-1} (Supplementary Fig. 15), showing that much, but not all, of the binding improvements are due to the contributions from charge-charge interactions.

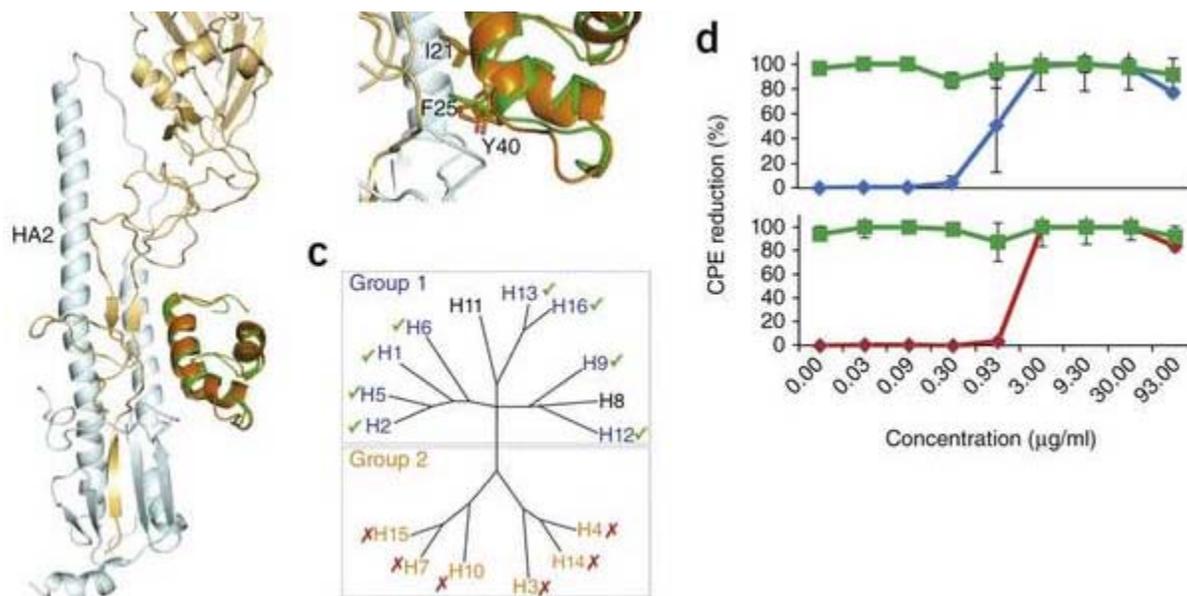
Table 1: Binding affinity and kinetics of selected design variants

Structure determination

To investigate the molecular determinants of recognition of the improved design variant, we determined the X-ray structure of F-HB80.4 in complex with the SC1918/H1 hemagglutinin ectodomain at 2.7 Å resolution. After molecular replacement using only the SC1918/H1 hemagglutinin structure as the search model (PDB 3GBN)¹³, clear electron density was observed for the inhibitor. F-HB80.4 binds the target hemagglutinin region in the orientation predicted by the designed model, with the main recognition helix packed in the hydrophobic groove between helix A and the N-terminal segment of HA1 (Fig. 4a,b). The overall backbone conformation of F-HB80.4 agrees well with the electron density maps, but atomic displacement parameters (B-values) are elevated and a few features, such as some side chains, are not apparent for residues that are distant from the F-HB80.4-HA interface, presumably due to conformational plasticity in F-HB80.4 or some heterogeneity in binding (Supplementary Figs. 16–18 and Supplementary Table 10). However, the main contact helix on F-HB80.4 is well ordered and, after refinement, electron density was apparent for most of the key contact residues on F-HB80.4, including Phe13, Ile17, Ile21, Phe25 and Tyr40. Taken together, the crystal structure of F-HB80.4, as well as that of the previously solved HB36.3, are in excellent agreement with the designed interface, with no significant deviations at any of the contact positions. This agreement between the design model and the crystal structure is quite encouraging given that *de novo* protein interface design is at an early stage. F-HB80.4 not only interacts with the hydrophobic cleft in hemagglutinin recognized by HB36 (ref. 1) but also interacts with the A helix and N-terminal segment of HA1 through the designed hotspot residue Tyr40, which recapitulates the similar interaction of Tyr98 in CR6261 and Tyr102 in the broadly neutralizing antibody F10 (ref. 14).

Figure 4: Structure and functional analysis of F-HB80.4.





Binding and neutralization

Evaluation of the binding affinity of F-HB80.4 against a panel of group 1 hemagglutinins by biolayer interferometry showed that it is more cross-reactive than the starting HB80.3 and many neutralizing antibodies targeting the same surface on hemagglutinin, such as CR6261 (Fig. 4c and Table 2). In addition to binding all of the group 1 hemagglutinins recognized by antibody CR6261 (H1, H2, H5, H6, H9, H13 and H16), F-HB80.4 also binds to H12 hemagglutinin, which neither CR6261 nor HB80.3 do^{1, 13}. Most notably, F-HB80.4 binds human H2 hemagglutinins with high affinity.

Table 2: Binding specificity of HB80.4 and CR6261 for different hemagglutinin subtypes

Given its high-affinity, heterosubtypic binding and inhibitory activity in biochemical assays (Supplementary Fig. 19)¹, we tested the neutralization potential of F-HB80.4 against the recent A/California/04/2009 H1N1 virus, which was responsible for the 2009 H1N1 pandemic and is currently established as the predominant circulating strain, as well as the seasonal human flu virus A/H1N1/Hawaii/31/2007. F-HB80.4 showed 50% effective concentrations (EC₅₀s) of 170 nM (1.6 μg/ml) and 98 nM (0.9 μg/ml) against 25 TCID₅₀ (50% tissue culture infective dose) of these viruses (Fig. 4d).

Discussion

Abstract **Introduction** **Results** **Discussion** **Methods** **Accession Codes** **References** **Acknowledgments**

Author Information **Supplementary Information**

Deep sequencing of populations undergoing nonpurifying selection has been used to experimentally determine fitness landscapes for a heat shock protein¹⁵ and an RNA enzyme¹⁶, and to map interactions for protein-DNA^{17, 18}, protein-peptide² and HIV-1 antibody-antigen complexes¹⁹. These approaches probed sequence changes within a single segment no larger than the ~80 bp that can be covered in an Illumina sequencing run. Our approach using single-site mutagenesis libraries and multiple-segment Illumina sequencing has the advantage of being able to interrogate large stretches of sequence and still allow enrichment values to be associated with specific substitutions. Furthermore, our use of single-site mutagenesis libraries allowed complete probing of an extended region (150 bp) with relatively small starting libraries, which resulted in extensive sampling and robust statistics for the vast majority of the substitutions investigated; as in previous approaches, normalization to the starting pools corrected for any initial library bias (from either codon usage or synthesis). Beyond these technical advances, because we applied the method to computationally designed, rather than evolutionarily optimized native proteins, our landscapes differ from those observed in previous studies in that there are positions where substitutions provide significant enrichment over the initial starting sequence.

The HB36.4 and HB80.3 results both show that landscapes mapped by deep sequencing can be used to rapidly obtain large increases in binding affinity after conventional directed evolution by PCR mutagenesis has plateaued by combining large numbers of individually small, favorable effects. The specific combination of mutations contained within these variants would be very difficult to find by conventional affinity maturation approaches. For example, identification of the F-HB80.4 variant with 5 amino-acid mutations (8 DNA sequence changes) using unbiased libraries would have required screening all five amino-acid mutant combinations—a diversity of $7.5E+12$ —whereas the total diversity of the landscape-guided library was 10^7 -fold lower. The traditional approach of carrying out multiple rounds of selection and then using conventional sequencing to identify the few best clones would also not have arrived at the high-affinity variants; only one of the substitutions found in the highest-affinity variant was among the most heavily enriched in the population and, therefore, combining the few top mutations found after conventional selection and sequencing would not have led to the best combined variant. The results also illustrate how the landscapes can be exploited to reprogram interaction specificity for closely related targets (H1 and H5 hemagglutinin) by examining not just beneficial mutations but also neutral and deleterious ones.

Our results show how the landscapes generated by deep sequencing can provide a comprehensive view of the shortcomings in computational protein design and can guide the development of more accurate force fields and more powerful design methods. The incorporation of long-range electrostatics into the design force field considerably improved recapitulation of the energy landscape data. Continuum electrostatics calculations have been applied to modeling protein-protein interactions previously^{20, 21}; our implementation is particularly well suited to calculations on large numbers of mutations because it employs a single full Poisson-Boltzmann solution for the potential of the fixed target in all calculations, which makes computations rapid and reduces noise due to changing boundary conditions. The large number (~2,000) of experimental data points generated by the approach was invaluable for guiding robust improvement of the force field; the much smaller data sets generated by conventional methods can be all too readily overfit.

Antivirals with more potent and cross-reactive activity against the H2 subtype, such as F-HB80.4, could be key components of a comprehensive therapy for influenza. H2N2 viruses were responsible for the deaths of ~1 million people during the 1957 pandemic, and these viruses continued to circulate in humans until 1968. Given their proven capacity for sustained replication and transmission in humans and the lack of widespread immunity to H2N2 viruses in the general population (that is, people born after 1968 have never been exposed to H2 viruses and immunity among individuals infected more than 40 years ago may have declined), the reservoir of H2N2 viruses in birds is a possible source for a future pandemic. The

Ile45Phe substitution in the HA2 subunit found in all human H2 viruses strongly reduces the binding of CR6261 and other V_{H1-69} -related antibodies²². Consequently, CR6261 neutralization of H2 is restricted to avian viruses (with Ile45), and only the recently described FI6v3 antibody has been reported to neutralize all virus subtypes, including human H2 viruses²³. Despite targeting the same surface recognized by neutralizing antibodies, the high-affinity interaction of F-HB80.4 with human H2 hemagglutinin underscores a potential advantage of *de novo*-designed binders, as they are likely to bind the target differently than an antibody (e.g., using a helix rather than the antibody CDR loops) and can, in some cases, circumvent barriers that have posed some problems for antibodies, such as that for V_{H1-69} antibodies binding H2 viruses.

The levels of neutralization activity attained with F-HB80.4 are nearly equivalent to those of neutralizing antibodies, which have a 50% inhibitory concentration (IC_{50}) range of 0.1–100 $\mu\text{g/ml}$ IgG (e.g., the IC_{50} for CR6261 IgG against H1 hemagglutinin is 9 $\mu\text{g/ml}$ (~ 120 nM))²². Although the therapeutic potential of small binding proteins remains to be proven in humans, F-HB80.4 either alone, as a fusion with an antibody Fc, or as a high-avidity oligomer is a promising lead candidate for the next generation of antiviral therapeutics.

More generally, integration of deep sequencing with computational protein design provides, in principle, a powerful route to inhibitors or binders for any surface patch on any desired target of interest. Given a newly arising pathogen, for example, following structure determination and identification of sites of interaction with the host, hot-spot-based protein interface design can be used to generate diverse small proteins predicted to block the host interaction surface. With modern oligonucleotide assembly methods, genes for large numbers of designs can be rapidly built and displayed on yeast, where the functional designs can be readily identified by flow cytometry. Complete single-site saturation mutagenesis libraries can then be generated for functional designs and subjected to deep sequencing before and after one round of selection for increased binding activity. The enriched substitutions can be combined in a final library, and optimized high-affinity variants selected from this pool. We anticipate that this combined approach will be widely useful in generating high-affinity and high-specificity binders to a broad range of targets for use in therapeutics, diagnostics and targeting.

Methods

[Abstract](#) [Introduction](#) [Results](#) [Discussion](#) [Methods](#) [Accession Codes](#) [References](#) [Acknowledgments](#)

[Author Information](#) [Supplementary Information](#)

Library creation.

Single-site saturation mutagenesis (SSM) libraries for HB36.4 and HB80.3 were constructed from synthetic DNA by Genscript. Parental DNA sequences are listed in Supplementary Table 1 with the mutagenic region highlighted in red. Yeast EBY100 cells were transformed with library DNA and linearized pETCON¹ using an established protocol⁴³, yielding 1.4e6 and 3.3e6 transformants for the HB36.4 and HB80.3 SSM libraries, respectively. After transformation, cells were grown overnight in SDCAA media in 30 ml cultures at 30 °C, passaged once, and stored in 20 mM HEPES 150 mM NaCl pH 7.5, 20% (w/v) glycerol in 1e7 aliquots at –80 °C.

Yeast display selections and titrations.

Cell aliquots were thawed on ice, centrifuged at 13,000 r.p.m. for 30 s, resuspended in 1e7 cells per ml of SDCAA media and grown at 30 °C for 6 h. Cells were then centrifuged for 13,000 r.p.m. and resuspended at 1e7 cells per ml SGCAA media and induced at 22 °C for 16–24 h. Cells were labeled with either biotinylated Viet/2004/H5 hemagglutinin or SC1918/H1 hemagglutinin, washed, secondary labeled with SAPE (Invitrogen) and anti-cmyc FITC (Miltenyi Biotec), and sorted by fluorescent gates (Supplementary Tables 2 and 3 and Supplementary Fig. 2). Biotinylated hemagglutinin was produced as previously described¹. Cells were recovered overnight at 2.5e5 collected cells per ml SDCAA media, whereupon at least 1e7 cells were spun down at 13,000 r.p.m. for 1 min and stored as cell pellets at –80 °C before library prep for deep sequencing. Plasmid DNA for individual clones was produced according to the method of Kunkel⁴⁴ and yeast display titration was done as previously reported^{1, 43}.

Library prep and sequencing.

Between $1e7$ and $4e7$ yeast cells were resuspended in Solution I (Zymo Research yeast plasmid miniprep II kit) with 25 U zymolase and incubated at $37\text{ }^{\circ}\text{C}$ for 4 h. Cells were frozen/thawed using a dry ice/ethanol bath and a $42\text{ }^{\circ}\text{C}$ incubator. Afterwards, plasmid was recovered using a Zymo Research yeast plasmid miniprep II kit (Zymo Research, Irvine, CA) into a final volume of $30\text{ }\mu\text{L}$ 10 mM Tris-HCl pH 8.0. Contaminant genomic DNA was processed (per $20\text{ }\mu\text{L}$ rxn) using $2\text{ }\mu\text{L}$ Exol exonuclease (NEB), $1\text{ }\mu\text{L}$ lambda exonuclease (NEB) and $2\text{ }\mu\text{L}$ lambda buffer at $30\text{ }^{\circ}\text{C}$ for 90 min followed by heat inactivation of the enzymes at $80\text{ }^{\circ}\text{C}$ for 20 min. Plasmid DNA was separated from the reaction mixture using a Qiagen PCR cleanup kit (Qiagen). Next, 18 cycles of PCR ($98\text{ }^{\circ}\text{C}$ 10 s, $68\text{ }^{\circ}\text{C}$ 30s, $72\text{ }^{\circ}\text{C}$ 10 s) using Phusion high fidelity polymerase (NEB, Waltham, MA) were used to amplify the template and add the Illumina adaptor sections. Primers used were population-specific and are listed in Supplementary Table 11. The PCR reaction was purified using an Agencourt AMPure XP kit (Agencourt, Danvers, MA) according to the manufacturer's specifications. Samples were quantified using Qubit dsDNA HS kit (Invitrogen) for a final yield of 1–4 ng/ μL . Samples were combined in an equimolar ratio; from this pool, 0.32 fmol of total DNA was loaded on two separate lanes and sequenced using a Genome Analyzer IIx (Illumina) with appropriate sequencing primers (Supplementary Table 11).

Sequencing analysis.

Alignment and quality filtering of the sequencing data from raw Illumina reads were treated essentially as described². Sequencing reads were assigned to the correct pool on the basis of a unique 8 bp barcode identifier (Supplementary Table 11). All pools were treated identically in sequence analysis and quality filtration. Custom scripts were used to align all paired-end reads with both reads above an average Phred quality score equal or above 20. Paired-end reads were aligned using a global Needleman-Wunsch algorithm, reads without gaps were merged into a single sequence and differences between sequences resolved using the higher quality score for the read.

To investigate amino-acid sequence covariance, two-body analysis was performed whereby the enrichment ratio for pairs of mutations was compared to the predicted enrichment ratio based on the individual component mutations. The individual enrichment value was calculated as the overall normalized probability of finding the mutation in the selected pool, the predicted enrichment for a pair of mutations was the sum of the component mutations enrichment values, and the actual enrichment ratio was calculated as the overall normalized probability of finding that pair of mutations in a selected pool. A more rigorous analysis was performed to rank each mutational variant found in the deep sequenced library using a statistical model based on the method of Balakrishnan¹². In brief, the method constructs a maximum entropy statistical model of the following functional form:

$$P(s) \propto \exp\left(\sum_i f_i(s_i) + \sum_{(i,j) \in E} f_{ij}(s_i, s_j)\right)$$

where s is a particular 9-mer from the sort1 set, s_i and s_j are the amino acids at the i th/ j th positions of this sequence, E is the set of interacting pairs of positions identified by the model and f_i , f_{ij} are model parameters that can be thought of as 1 and 2 body (negative) statistical energies, respectively. Thus, each f_i can be thought of as a vector that stores the statistical energies for the possible amino acids at that position, whereas f_{ij} is, analogously, a matrix that stores the statistical energies for the amino acid pairs at positions i and j . These parameters are learned from the data using a maximum likelihood procedure based on LASSO²⁴. A baseline model that does not capture sequence covariation (that is, a model with all f_{ij} s set to zero) was also learnt from the data. Note that, as expected, the probability of an entire sequence can then be written as the product of probabilities of the amino-acid compositions at each position; that is, each position of the 9 mer is treated independently under the baseline model.

Affinity maturation and specificity.

Beneficial mutations predicted to result in higher affinity for SC1918/H1 hemagglutinin were combined into single libraries for both HB80.3 and HB36.4. The DNA library for each design was constructed from assembly PCR using an Ultramer

oligonucleotide (Integrated DNA Technologies, CA) to encode the variable region. Primers and sequences are listed in Supplementary Table 11, whereas the DNA sequence for the libraries is listed in Supplementary Table 6. The total library size was 3×10^5 for HB80.3 and 4×10^6 for HB36.4, and was transformed into yeast²⁵, yielding 8×10^6 and 1.5×10^7 transformants, respectively. These libraries went through five sorts of yeast display selection with increasing stringency against HA1–2 as specified in Supplementary Table 11. Promising constructs were subcloned into a custom pET-29-based plasmid (NdeI/XhoI) with an N-terminal FLAG tag and a C-terminal His₆ tag and transformed into *E. coli* Rosetta (DE3) chemically competent cells for expression.

Solubility screening.

HB80.3 clones selected from the affinity maturation library were screened by solubility in an *E. coli* expression system using a dot-blot assay. Cells were grown from colonies in deep well plates overnight and diluted 25-fold into deep well plates at 37 °C for 3 h, followed by IPTG induction (1 mM) for 4 h at 37 °C. Following induction, cells were separated from spent media by centrifugation at $3,000 \times g$ for 15 min at 4 °C and stored as pellets overnight at –20 °C. The next morning, plates were thawed on ice for at least 15 min and 200 μ L binding buffer (200 mM HEPES, 150 mM NaCl, pH 7.5) was added to each well. The plate was sonicated using the Ultrasonic Processor 96-well sonicator for 3 min at 70% pulsing power and lysate centrifuged for 4,000 r.p.m. for 30 min at 4 °C. Supernatant at 100-fold dilution was transferred to a dot blot manifold Minifold I (Whatman) and dried onto nitrocellulose membrane for 5 min. The membrane was then labeled with an anti-FLAG HRP conjugated mouse antibody (Sigma, St. Louis, MO) and visualized with DAB substrate (Pierce).

Protein production and purification.

Protein expression was induced using the autoinduction method of Studier²⁶. Cells were harvested by centrifugation, resuspended into buffer HBS (20 mM Hepes, 150 mM NaCl pH 7.4) and sonicated to release cell lysate. Following clarification by centrifugation, supernatant was applied to a Talon resin column for purification. Proteins were eluted by step elution at 400 mM imidazole in HBS. Size exclusion chromatography on a Superdex75 column was used as a finishing purification step for HB80.3 variants. Proteins were stored at 4 °C for short-term analysis or flash frozen in liquid nitrogen.

Binding analysis.

All surface plasmon resonance data were recorded on a Biacore model T100 (Biacore, Uppsala, Sweden). A Biotin CAPture chip (Biacore) was coated with 500 response units (RU) of biotinylated SC1918/H1 HA1-2 ectodomain. All proteins were in buffer HBS-EP with 3 mM EDTA and 0.005% (v/v) P20 surfactant. 238 μ L of designed protein was applied at a flow rate of 100 μ L/min for 2 min and a dissociation time of 300s with full chip regeneration between each trace. At least five varying concentrations of protein were used to determine kinetic and equilibrium fits. Binding kinetics were determined using a 1:1 Langmuir binding model with Biacore T100 evaluation software and double background-subtracted values.

Bi-layer interferometry using an Octet Red (ForteBio, Menlo Park, CA) was used to determine subtype-specific binding for HB80.4 and CR6261. Biotinylated hemagglutinins, purified as described¹, were used for these measurements (Supplementary Table 13). Briefly hemagglutinins at ~10–50 μ g/ml in 1x kinetics buffer (1x PBS, pH 7.4, 0.01% BSA, and 0.002% Tween 20) were loaded onto streptavidin-coated biosensors and incubated with varying concentrations of HB80.4 in solution. All binding data were collected at 30 °C. The experiments comprised 5 steps: 1. Baseline acquisition (60 s); 2. Hemagglutinin loading onto sensor (300 s); 3. Second baseline acquisition (180 s); 4. Association of HB80.4 for the measurement of k_{on} (180 s); and 5. Dissociation of HB80.4 for the measurement of k_{off} (180 s). Five concentrations of HB80.4 were used, with the highest concentration varying, depending on the hemagglutinin affinity from 50 to 200 nM. Baseline and dissociation steps were carried out in buffer only. Binding kinetics were determined using a 1:1 Langmuir binding model in kinetics data analysis mode using the Fortebio data processing software. The sequences of all biotinylated hemagglutinins used in this work are available in Fasta format in Supplementary Table 12.

Protease susceptibility assays.

Protease susceptibility assays were done as described¹. For A/South Carolina/1/1918 (H1N1) hemagglutinin, each reaction contained ~2.5 µg hemagglutinin or ~2.5 µg hemagglutinin and a fivefold molar excess of F-HB80.4. Significant inhibition was detected with a high ratio of binder to hemagglutinin, presumably due to the stringency of our assay (1 h at 37 °C at low pH). Little protection was observed when the reaction contained approximated 1 binder per hemagglutinin protomer.

Computational methods.

The Rosetta all atom energy function and design methodology was used to calculate the predicted effect of every possible point mutation in the designed proteins on the free energies of folding and binding using

$$-\ln(P_{\text{binding}}) = -\ln \left[\frac{e^{-\frac{-(\Delta\Delta G_{\text{folding}} + \Delta G_0)}{kT}}}{1 + e^{-\frac{-(\Delta\Delta G_{\text{folding}} + \Delta G_0)}{kT}}} \right] + \Delta\Delta G_{\text{binding}}$$

where $\Delta\Delta G_{\text{folding}}$ is the computed change in stability²⁷, $\Delta\Delta G_{\text{binding}}$ is the computed change in binding free energy and ΔG_0 is the free energy of folding, taken to be 1.0 in the units used here. The first term accounts for the reduction in the population of the folded state brought about by mutation, the second term, the direct effect of the mutation on the binding interaction.

Starting from models of the HB36.4 and HB80.3 complexes that came for the experimentally determined structures for HB36.3 and F-HB80.4 (ref. 1), each position was singly mutated to all 20 amino-acid identities and for each mutation the structure was optimized by combinatorial repacking of side chains and gradient-based steepest-descent minimization of degrees of freedom on side chain of both sides of the complex and backbone of the designed protein. The complex binding affinity and the unbound stability of the designed monomer were both analyzed using an all-atom energy function dominated by van-der-Waals interactions, hydrogen bonding and solvation²⁸. In binding-affinity calculations, the monomers were repacked in the unbound state but backbone degrees of freedom were kept fixed. For monomer stability calculations, a Coulombic model using distance dependent dielectric constant ($\epsilon = r$) is added to account for intra-molecular electrostatic interactions. The PARSE charges²⁹ are used for all residues. The $\Delta\Delta G$ of protein stability and binding energy upon mutation is calculated with both standard van-der-Waals parameters and a reduced repulsive term²⁷. Earlier benchmarks showed that this is an efficient approach to identify mutations that introduce van-der-Waals clashes but can be tolerated given more structural flexibility. If $\Delta\Delta G$ decreases by over 5 R.e.u. (Rosetta energy units), an additional step of structure optimization is added with standard van-der-Waals parameters, allowing freedom on the rigid body movement between the proteins and side chain and the backbone of both sides of the complex. This additional optimization step leads to more small to large mutations favored in the calculations, decreasing the number of false negatives, but increasing the number of false positives for predicting the favored mutations. This is a desirable behavior for the protocol, as it leads to more favorable mutations that can be tested. This procedure was implemented using the Rosetta macromolecular software package¹⁰. To model long-range electrostatics efficiently and with minimal noise, we calculated the electrostatic potential in the vicinity of the designed proteins due to hemagglutinin on a grid by solving the PB equation with charges on the atoms in hemagglutinin, but with all atoms in the designed proteins neutral. The Poisson-Boltzmann equation was solved using APBS²⁶ with PARSE charges and radii^{29, 30} for hemagglutinin atoms, but no charges for HB atoms and the electrostatic potential generated by hemagglutinin was calculated on a grid with 0.5 Å. The protein is modeled in the low dielectric constant of 4. The solvent is modeled implicitly with high dielectric constant of 80 and salt concentration of 0.15 M. The PARSE charges are assigned to hemagglutinin³⁰ and the HB design variant is neutral. The PARSE radii are assigned to both hemagglutinin and HB. The dielectric boundary is defined by the solvent exclusion surface using a probe with a radius of 1.4 Å³¹. The electrostatic interaction energy caused by each point mutation was computed using $E = \sum q_i * f$, where f is the electrostatic potential from the grid and q_i are the charges of the atoms on the introduced residues. The energy term is converted to the Rosetta score function term by 1 kT = 1 R.e.u. Detailed RosettaScripts⁹ for all computational analyses are

available in Supplementary Scripts. Source code is freely available to academic users through the Rosetta Commons agreement (<http://www.rosettacommons.org/>).

Isolation of F-HB80.4-SC1918/H1 hemagglutinin complex for crystallization.

Following Ni-NTA purification, SC1918 hemagglutinin was digested with trypsin (New England Biolabs, 5mU trypsin per mg hemagglutinin, 16 h at 17 °C) to produce uniformly cleaved (HA1/HA2), and to remove the trimerization domain and His-tag. After quenching the digests with 2 mM PMSF, the digested material was purified by anion exchange chromatography (10 mM Tris, pH 8.0, 0.05–1M NaCl) and size exclusion chromatography (10 mM Tris, pH 8.0, 150 mM NaCl), essentially as previously described for other hemagglutinins¹.

To prepare the F-HB80.4/SC1918 complex for crystallization, 1.5 molar excess of F-HB80.4 was mixed with purified SC1918 hemagglutinin in 10 mM Tris pH 8.0, 150 mM NaCl at ~2 mg/ml. The mixtures were incubated overnight at 4 °C to allow complex formation. Saturated complexes were then purified from unbound F-HB80.4 by gel filtration.

Crystallization and structure determination of F-HB80.4-SC1918/H1 hemagglutinin complex.

Gel filtration fractions containing the F-HB80.4/SC1918 complex were concentrated to ~10 mg/ml in 10 mM Tris, pH 8.0 and 50 mM NaCl. Initial crystallization trials were set up using the automated Rigaku CrystalMation robotic system at the Joint Center for Structural Genomics (<http://www.jcsg.org/>). Several hits were obtained, with the most promising candidates grown in ~15% PEG3350 around pH 7. Optimization of these conditions resulted in diffraction quality crystals. The crystals used for data collection were grown by the sitting drop vapor diffusion method with a reservoir solution (100 μ L) containing 16% PEG3350, and 100 mM Tris pH 7.5. Drops consisting of 100 nL protein + 100 nL precipitant were set up at 4 °C, and crystals appeared after 3 days. The resulting crystals were cryoprotected by soaking in well solution supplemented with increasing concentrations of ethylene glycol (5% steps, 5 min/step), to a final concentration of 25%, then flash cooled and stored in liquid nitrogen until data collection.

Diffraction data for the F-HB80.4-SC1918/H1 complex were collected at the Advanced Photon Source (APS) General Medicine/Cancer Institutes-Collaborative Access Team (GM/CA-CAT) beamline 23ID-D at the Argonne National Laboratory. The data were indexed in $P2_12_12_1$, integrated using HKL2000 (HKL Research) and scaled using Xprep (Bruker). The structure was solved by molecular replacement to 2.5 Å resolution using Phaser³². An unpublished, in house, high-resolution structure of the 1918 hemagglutinin was used as the initial search model. Examination of the maps at this stage revealed clear positive electron density around the membrane distal end of hemagglutinin consistent with the expected location and orientation of F-HB80.4. As for HB36.3 (ref. 1), attempts to place F-HB80.4 by molecular replacement using Phaser were unsuccessful. However, phasing using the hemagglutinin only yielded maps with continuous density for F-HB80.4, including key side-chain features. This phasing model allowed F-HB80.4 to be fitted into the maps manually and unambiguously. Rigid-body refinement, torsion-angle simulated annealing and restrained refinement (including TLS refinement, with one group for HA1, one for HA2 and one for F-HB80.4) was carried out in Phenix³³. Between rounds of refinement, the model was rebuilt and adjusted using Coot³⁴. Although we report the structure to a final resolution of 2.7 Å, the crystals diffracted anisotropically to 2.4 Å (along a), 2.5 Å (along b), 2.8 Å (along c) as determined by the diffraction anisotropy server³⁵. Data that were truncated and scaled by this server were used for model building. The electron density maps from these 2.7 Å data were of better quality and slightly easier to interpret than those at a higher resolution of 2.5Å. Data collection statistics are reported for data with the ellipsoidal truncation applied before merging of reflections. The final round of refinement was carried out with data that were ellipsoidally truncated, but with no negative isotropic B-value applied to the data. For the inhibitor F-HB80.4, residues distant from the F-HB80.4-hemagglutinin interface lacking side-chain electron density were modeled as alanine. The hemagglutinin head region is well ordered with lower B-values, which increase toward the stem and the inhibitor where there are fewer to no crystal lattice contacts. Final refinement statistics can be found in Supplementary Table 10.

Structural analyses.

Hydrogen bonds and van-der-Waals contacts between F-HB80.4 and SC1918/H1 hemagglutinin were calculated using HBPLUS³⁶ and CONTACTSYM³⁷, respectively. MacPyMol (DeLano Scientific)³⁸ was used to render structure figures and for general manipulations. The final coordinates were validated using the JCSG quality control server (v2.7), which includes MolProbity³⁹.

Neutralization assay viruses.

A/California/04/2009 (pdmH1N1) and A/Hawaii/31/2007 (H1N1) were propagated in Madin-Darby canine Kidney (MDCK) cells (American Type Culture Collection, Manassas, VA) to produce working viral stocks.

Cell culture.

MDCK cells were grown in minimum essential medium (MEM) with Earle's Balanced Salts supplemented with 5% FBS (Hyclone Laboratories, Logan, UT). Virus amplification for virus stock production was carried out in MEM containing gentamicin (50 µg/ml), porcine trypsin (10 units/ml) and EDTA (1 µg/ml)⁴⁰. The antiviral testing was performed in MEM supplemented only with gentamicin (50 µg/ml).

Viral inhibition assays.

To calculate the F-HB80.4 concentration-response curve, the peptides were half log diluted in MEM from 10 µM to 0.00032 µM and incubated with 25 TCID₅₀ of virus at 37 °C with 5% CO₂ for 1 h. After incubation, the reaction mixture of each concentration was added to three wells of MDCK cells (8 × 10⁴ cells/well) prepared in 96 well plates. Cell controls (uninfected and untreated cells), virus controls (infected and untreated cells) and F-HB80.4 toxicity controls (infected and untreated cells) were included in each test plate. The test was read at day 6 post-inoculation when virus control wells showed 100% cytopathic effect (CPE). The CPE was evaluated via cell viability through the cellular intake of neutral red (NR) (Thermo Fisher Scientific Inc., Pittsburg, PA)⁴¹. The NR was used at 0.011% diluted in MEM, the cells were incubated at 37 °C with 5% CO₂ for 2 h and the plates were read spectrophotometrically.

The EC₅₀ for the peptides were obtained by the standardization of the NR results for each of the peptide concentration repetitions against the cell controls (100% viability) and virus controls (100% cell death). A plot of the obtained data as percentage of cell viability and percentage of CPE reduction against the peptide concentration was constructed using Excel, 2007. The curve points were also fitted using Excel, 2007⁴².

Accession code.

The X-ray crystallographic coordinates have been deposited in the Protein Data Bank with accession ID 4EEF.

Accession codes

[Abstract](#) [Introduction](#) [Results](#) [Discussion](#) [Methods](#) [Accession Codes](#) [References](#) [Acknowledgments](#)
[Author Information](#) [Supplementary Information](#)

Primary accessions

Protein Data Bank

4EEF 4EEF

Referenced accessions

Protein Data Bank

3GBN 3GBN

References

[Abstract](#) [Introduction](#) [Results](#) [Discussion](#) [Methods](#) [Accession Codes](#) [References](#) [Acknowledgments](#)

Author Information Supplementary Information

1. Fleishman, S.J. *et al.* Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science* **332**, 816–821 (2011).
Show context Article PubMed ISI ADS CAS
2. Fowler, D.M. *et al.* High-resolution mapping of protein sequence-function relationships. *Nat. Methods* **7**, 741–746 (2010).
Show context Article PubMed ISI CAS
3. Araya, C.L. & Fowler, D.M. Deep mutational scanning: assessing protein function on a massive scale. *Trends Biotechnol.* 435–442 (2011).
Show context
4. Chao, G. *et al.* Isolating and engineering human antibodies using yeast surface display. *Nat. Protoc.* **1**, 755–768 (2006).
Show context Article PubMed ISI CAS
5. Cunningham, B.C. & Wells, J.A. High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science* **244**, 1081–1085 (1989).
Show context Article PubMed ISI CAS
6. Bowie, J.U., Reidhaar-Olson, J.F., Lim, W.A. & Sauer, R.T. Deciphering the message in protein sequences: tolerance to amino acid substitutions. *Science* **247**, 1306–1310 (1990).
Show context Article PubMed ISI ADS CAS
7. Pal, G., Kouadio, J.L., Artis, D.R., Kossiakoff, A.A. & Sidhu, S.S. Comprehensive and quantitative mapping of energy landscapes for protein-protein interactions by rapid combinatorial scanning. *J. Biol. Chem.* **281**, 22378–22385 (2006).
Show context Article PubMed CAS
8. Bershtein, S., Segal, M., Bekerman, R., Tokuriki, N. & Tawfik, D.S. Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature* **444**, 929–932 (2006).
Show context Article PubMed ISI ADS CAS
9. Fleishman, S.J. *et al.* RosettaScripts: a scripting language interface to the rosetta macromolecular modeling suite. *PLoS ONE* **6**, e20161 (2011).
Show context Article PubMed CAS
10. Leaver-Fay, A. *et al.* ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **487**, 545–574 (2011).
Show context PubMed CAS

11. Dutta, S. *et al.* Determinants of BH3 binding specificity for Mcl-1 versus Bcl-xL. *J. Mol. Biol.* **398**, 747–762 (2010).
Show context Article PubMed CAS
12. Balakrishnan, S., Kamisetty, H., Carbonell, J.G., Lee, S.I. & Langmead, C.J. Learning generative models for protein fold families. *Proteins* **79**, 1061–1078 (2011).
Show context Article PubMed CAS
13. Ekiert, D.C. *et al.* Antibody recognition of a highly conserved influenza virus epitope. *Science* **324**, 246–251 (2009).
Show context Article PubMed ISI ADS CAS
14. Sui, J. *et al.* Structural and functional bases for broad-spectrum neutralization of avian and human influenza A viruses. *Nat. Struct. Mol. Biol.* **16**, 265–273 (2009).
Show context Article PubMed ISI CAS
15. Hietpas, R.T., Jensen, J.D. & Bolon, D.N. Experimental illumination of a fitness landscape. *Proc. Natl. Acad. Sci. USA* **108**, 7896–7901 (2011).
Show context Article PubMed ADS CAS
16. Pitt, J.N. & Ferre-D'Amare, A.R. Rapid construction of empirical RNA fitness landscapes. *Science* **330**, 376–379 (2010).
Show context Article PubMed ADS CAS
17. Patwardhan, R.P. *et al.* High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. *Nat. Biotechnol.* **27**, 1173–1175 (2009).
Show context Article PubMed ISI CAS
18. Shultzaberger, R.K., Malashock, D.S., Kirsch, J.F. & Eisen, M.B. The fitness landscapes of cis-acting binding sites in different promoter and environmental contexts. *PLoS Genet.* **6**, e1001042 (2010).
Show context Article PubMed CAS
19. Wu, X. *et al.* Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing. *Science* **333**, 1593–1602 (2011).
Show context Article PubMed ISI ADS CAS
20. Joughin, B.A., Green, D.F. & Tidor, B. Action-at-a-distance interactions enhance protein binding affinity. *Protein Sci.* **14**, 1363–1369 (2005).
Show context Article PubMed ISI CAS
21. Marshall, S.A., Vizcarra, C.L. & Mayo, S.L. One- and two-body decomposable Poisson-Boltzmann methods for protein design calculations. *Protein Sci.* **14**, 1293–1304 (2005).
Show context Article PubMed CAS

22. Throsby, M. *et al.* Heterosubtypic neutralizing monoclonal antibodies cross-protective against H5N1 and H1N1 recovered from human IgM+ memory B cells. *PLoS ONE* **3**, e3942 (2008).
- Show context Article PubMed ADS CAS
23. Corti, D. *et al.* A neutralizing antibody selected from plasma cells that binds to group 1 and group 2 influenza A hemagglutinins. *Science* **333**, 850–856 (2011).
- Show context Article PubMed ISI ADS CAS
24. Efron, B., Hastie, T., Johnstone, I. & Tibshirani, R. Least angle regression. *Ann. Stat.* **32**, 407–499 (2002).
- Show context
25. Benatuil, L., Perez, J.M., Belk, J. & Hsieh, C.M. An improved yeast transformation method for the generation of very large human antibody libraries. *Protein Eng. Des. Sel.* **23**, 155–159 (2010).
- Show context Article PubMed CAS
26. Studier, F.W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).
- Show context Article PubMed ISI CAS
27. Kellogg, E.H., Leaver-Fay, A. & Baker, D. Role of conformational sampling in computing mutation-induced changes in protein structure and stability. *Proteins* **79**, 830–838 (2011).
- Show context Article PubMed ISI CAS
28. Rohl, C.A., Strauss, C.E., Misura, K.M. & Baker, D. Protein structure prediction using Rosetta. *Methods Enzymol.* **383**, 66–93 (2004).
- Show context PubMed ISI CAS
29. Sitkoff, D., BenTal, N. & Honig, B. Calculation of alkane to water solvation free energies using continuum solvent models. *J. Phys. Chem.* **100**, 2744–2752 (1996).
- Show context Article CAS
30. Sitkoff, D., Sharp, K.A. & Honig, B. Accurate calculation of hydration free-energies using macroscopic solvent models. *J. Phys. Chem.* **98**, 1978–1988 (1994).
- Show context Article ISI CAS
31. Richards, F.M. Areas, volumes, packing, and protein-structure. *Annu. Rev. Biophys. Bioeng.* **6**, 151–176 (1977).
- Show context Article PubMed ISI CAS
32. McCoy, A.J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
- Show context Article PubMed ISI CAS
33. Adams, P.D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta*

Crystallogr. D Biol. Crystallogr. **66**, 213–221 (2010).

Show context

Article PubMed ISI CAS

34. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).

Show context

Article PubMed ISI CAS

35. Strong, M. *et al.* Toward the structural genomics of complexes: Crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. USA* **103**, 8060–8065 (2006).

Show context

Article PubMed ADS CAS

36. McDonald, I.K. & Thornton, J.M. Satisfying hydrogen-bonding potential in proteins. *J. Mol. Biol.* **238**, 777–793 (1994).

Show context

Article PubMed ISI CAS

37. Sheriff, S., Hendrickson, W.A. & Smith, J.L. Structure of myohemerythrin in the azidomet state at 1.7/1.3-Å resolution. *J. Mol. Biol.* **197**, 273–296 (1987).

Show context

Article PubMed ISI CAS

38. The PyMOL Molecular Graphics System, Version 1.5.0.1 Schrödinger, LLC.

Show context

39. Chen, V.B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).

Show context

Article PubMed ISI CAS

40. Nguyen, J.T. *et al.* Triple combination of oseltamivir, amantadine, and ribavirin displays synergistic activity against multiple influenza virus strains *in vitro*. *Antimicrob. Agents Chemother.* **53**, 4115–4126 (2009).

Show context

Article PubMed CAS

41. Smee, D.F., Huffman, J.H., Morrison, A.C., Barnard, D.L. & Sidwell, R.W. Cyclopentane neuraminidase inhibitors with potent *in vitro* anti-influenza virus activities. *Antimicrob. Agents Chemother.* **45**, 743–748 (2001).

Show context

Article PubMed ISI CAS

42. Nguyen, J.T. *et al.* Triple combination of amantadine, ribavirin, and oseltamivir is highly active and synergistic against drug resistant influenza virus strains *in vitro*. *PLoS ONE* **5**, e9332 (2010).

Show context

Article PubMed ADS CAS

43. Chao, G., Cochran, J.R. & Wittrup, K.D. Fine epitope mapping of anti-epidermal growth factor receptor antibodies through random mutagenesis and yeast surface display. *J. Mol. Biol.* **342**, 539–550 (2004).

Show context

Article PubMed ISI CAS

44. Kunkel, T.A. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl. Acad. Sci. USA* **82**, 488–492 (1985).

Show context

Article PubMed ADS CAS

Download references

Acknowledgments

[Abstract](#) [Introduction](#) [Results](#) [Discussion](#) [Methods](#) [Accession Codes](#) [References](#) [Acknowledgments](#)

[Author Information](#) [Supplementary Information](#)

We thank D. Fowler and S. Fields for helpful discussions and use of their in-house software to process sequencing data, C. Lee, J. Shendure and M. Dunham for experimental expertise in DNA prep and sequencing, C. Sitz and C. Santiago for technical help and the Joint Center for Structural Genomics for crystallization using the JCSG/IAVI/TSRI Rigaku Crystallization system. This work was funded by Defense Advanced Research Projects Agency (DARPA) and the Defense Threat Reduction Agency (DTRA), and US National Institutes of Health, National Institute of Allergy and Infectious Diseases and National Institute of General Medical Sciences. The GM/CA CAT 23-ID-B beamline has been funded in whole or in part with federal funds from National Cancer Institute (Y1-CO-1020) and NIGMS (Y1-GM-1104). Use of the Advanced Photon Source (APS) was supported by the US Department of Energy, Basic Energy Sciences, Office of Science, under contract no. DE-AC02-06CH11357. The content is solely the responsibility of the authors and does not necessarily represent the official views of NIGMS or the NIH.

Author information

[Abstract](#) [Introduction](#) [Results](#) [Discussion](#) [Methods](#) [Accession Codes](#) [References](#) [Acknowledgments](#)

[Author Information](#) [Supplementary Information](#)

These authors contributed equally to this work.

Timothy A Whitehead & Aaron Chevalier

Affiliations

Department of Biochemistry, University of Washington, Seattle, Washington, USA.

Timothy A Whitehead, Aaron Chevalier, Yifan Song, Sarel J Fleishman, Hetunandan Kamisetty & David Baker

Department of Molecular Biology and the Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, California, USA.

Cyrille Dreyfus & Ian A Wilson

Naval Health Research Center, San Diego, California, USA.

Cecilia De Mattos, Chris A Myers & Patrick Blair

Howard Hughes Medical Institute, University of Washington, Seattle, Washington, USA.

David Baker

Present addresses: Department of Chemical Engineering and Materials Science, Michigan State University, East Lansing, Michigan, USA (T.A.W.) and Department of Biological Chemistry, Weizmann Institute of Science, Rehovot, Israel (S.J.F.).

Timothy A Whitehead & Sarel J Fleishman

Contributions

T.A.W. and A.C. conceived the idea, performed yeast display selections, analyzed deep sequencing data, performed hemagglutinin binding experiments, and performed computational modeling. Y.S. developed the electrostatics model and ran computational modeling code. C.D. expressed and purified hemagglutinin proteins, determined and analyzed the crystal structures with the guidance of I.A.W., and performed hemagglutinin binding experiments. S.J.F. assisted with structural analysis and developed the computational modeling code. C.D.M. performed the viral neutralization experiments under the guidance of C.A.M. and P.B. H.K. carried out covariance analysis on deep sequencing data. D.B. conceived the idea, analyzed deep sequencing data, and developed the electrostatics model. All authors discussed the results and wrote the manuscript.

Competing financial interests

T.A.W, S.J.F and D.B. have a patent application protecting proteins specified in this manuscript for use as potential influenza therapeutics.

Corresponding author

Correspondence to: David Baker

Supplementary information

[Abstract](#) [Introduction](#) [Results](#) [Discussion](#) [Methods](#) [Accession Codes](#) [References](#) [Acknowledgments](#)

[Author Information](#) [Supplementary Information](#)

PDF files

1. Supplementary Text and Figures (8M)
Supplementary Figures 1-19, Supplementary Tables 1-13 and Supplementary Scripts

Nature Biotechnology ISSN 1087-0156 EISSN 1546-1696

© 2012 Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.
partner of AGORA, HINARI, OARE, INASP, ORCID, CrossRef and COUNTER