PLOS | COMPUTATIONAL BIOLOGY

Protein-Protein Docking with Dynamic Residue Protonation States

Krishna Praneeth Kilambi[1], Kavan Reddy[1], Jeffrey J. Gray[1,2]

1 Department of Chemical & Biomolecular Engineering, Johns Hopkins University, Baltimore, Maryland, United States of America, **2** Program in Molecular Biophysics, Johns Hopkins University, Baltimore, Maryland, United States of America

Abstract

Protein-protein interactions depend on a host of environmental factors. Local pH conditions influence the interactions through the protonation states of the ionizable residues that can change upon binding. In this work, we present a pH-sensitive docking approach, pHDock, that can sample side-chain protonation states of five ionizable residues (Asp, Glu, His, Tyr, Lys) on-the-fly during the docking simulation. pHDock produces successful local docking funnels in approximately half (79/161) the protein complexes, including 19 cases where standard RosettaDock fails. pHDock also performs better than the two control cases comprising docking at pH 7.0 or using fixed, predetermined protonation states. On average, the top-ranked pHDock structures have lower interface RMSDs and recover more native interface residue-residue contacts and hydrogen bonds compared to RosettaDock. Addition of backbone flexibility using a computationally-generated conformational ensemble further improves native contact and hydrogen bond recovery in the top-ranked structures. Although pHDock is designed to improve docking, it also successfully predicts a large pH-dependent binding affinity change in the Fc–FcRn complex, suggesting that it can be exploited to improve affinity predictions. The approaches in the study contribute to the goal of structural simulations of whole-cell protein-protein interactions including all the environmental factors, and they can be further expanded for pH-sensitive protein design.

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are within the paper and its Supporting Information files.

Competing Interests: The authors have declared that no competing interests exist.

* Email: jgray@jhu.edu

This is a *PLOS Computational Biology* Methods article.

Introduction

Through tightly controlled cellular pH, posttranslational modification by protons regulates biological function [1]. Cellular pH can vary from highly-acidic in the lysosomes (∼pH 5) to basic in the peroxisomes (∼pH 8) [2], profoundly influencing biomolecular folding and assembly processes [3,4]. pH effects are especially critical in protein-protein binding, and binding-induced protonation state changes contribute to the association energy of most protein-protein complexes [5,6]. However, computational protein-protein docking algorithms often ignore the pH effects. In this paper, we develop a pH-sensitive protein-protein docking algorithm and demonstrate that it can improve prediction accuracy and recover pH-dependent binding effects.

Computational docking algorithms are playing an increasingly influential role in driving large-scale protein-protein interactions (PPI) surveys [7,8] and genome-wide interactome studies [9], but they need to accommodate sensitivity to local environment pH for improved reliability. Although pH effects on protein-small molecule complex calculations are well studied (e.g., refs. [10–15]), efforts to incorporate pH effects in computational protein-protein complex calculations have just begun. For example,

Spassov *et al.* [16] recently demonstrated a pH-sensitive binding prediction method with an aim to prolong the half-life of therapeutic antibodies. HADDOCK [17] determines the missing protonation state of the histidine residues in the input protein complex using the WHATIF server [18] before the start of the docking simulation. However, in real systems protonation states are affected not only by the solution pH but also the change in the local environment of the ionizable surface residues due to the receptor-ligand interactions during binding. pK_a calculation studies (e.g. [19]) stress the importance of simultaneously evaluating both favorable residue side-chain conformations and their preferred ionization states. Similarly, in docking algorithms, residue pK_a values vary depending on the conformations of the ligand relative to the receptor. Hence dynamic evaluation of the protonation states during docking using pK_a calculation algorithms on-the-fly is more true to the physical process of binding and may improve prediction accuracy.

Current computational pK_a calculation algorithms have been collectively assessed by the scientific community recently to improve their accuracy [20]. One of the primary aims of the pK_a calculation methods is to identify and improve the deficiencies of the energy function, particularly the electrostatics [21]. Despite the deficiencies, pK_a calculations by many algorithms are within a root-mean-square deviation (RMSD) of 1 pH unit from the

Author Summary

Protein-protein interactions are fundamental for biological function and are strongly influenced by their local environment. Cellular pH is tightly controlled and is one of the critical environmental factors that regulates protein-protein interactions. Three-dimensional structures of the protein complexes can help us understand the mechanism of the interactions. Since experimental determination of the structures of protein-protein complexes is expensive and time-consuming, computational docking algorithms are helpful to predict the structures. However, none of the current protein-protein docking algorithms account for the critical environmental pH effects. So we developed a pH-sensitive docking algorithm that can dynamically pick the favorable protonation states of the ionizable amino-acid residues. Compared to our previous standard docking algorithm, the new algorithm improves docking accuracy and generates higher-quality predictions over a large dataset of protein-protein complexes. We also use a case study to demonstrate efficacy of the algorithm in predicting a large pH-dependent binding affinity change that cannot be captured by the other methods that neglect pH effects. In principle, the approaches in the study can be used for rational design of pH-dependent protein inhibitors or industrial enzymes that are active over a wide range of pH values.

experimental pK_a values (except in extreme cases with very large pK_a shifts [22–24]). Hence unless the solution pH is very close to the shifted pK_a values of the ionizable residues, current algorithms can in principle reasonably estimate the relevant pH-sensitive protonation state during docking. Since computational protein-protein docking algorithms typically generate hundreds to several thousand target conformations, effective use of the protonation state data requires pK_a calculations to be fast, accurate and compatible with the docking methodology. Unfortunately, the most rigorous physics-based pK_a calculation methods prohibitively require several minutes to hours to calculate a single pK_a value, and the faster empirical methods are not currently compatible with the docking frameworks.

We previously created Rosetta-pH [25], a fast and efficient pK_a calculation algorithm with a focus on the use of the protonation state data in protein structure prediction and design. After we added a pH-sensitive score term to the standard (pH-independent) Rosetta score function and calibrated the electrostatic and solvation score terms, Rosetta-pH achieved a RMSD of 0.83 pH units from the experimental pK_a values. Since we built Rosetta-pH using the object-oriented Rosetta biomolecular modeling suite [26] which forms the basis for the protein-protein docking algorithm RosettaDock [27,28], we were able to fuse the methods to create, to our knowledge, the first pH-sensitive protein-protein docking algorithm.

In the remainder of this article, we describe our fast pH-sensitive docking algorithm (pHDock) that can sample side-chain protonation states of five ionizable residue types (Asp, Glu, His, Tyr, Lys) on-the-fly during the docking simulation. After combining the Rosetta-pH and RosettaDock frameworks, we recalibrate the pHDock score function to accommodate the new pH-sensitive score term. We use local docking studies to test pHDock's performance on a dataset of protein-protein complexes [29] and compare it to RosettaDock. We also study the effects of incorporating backbone flexibility in pHDock using a backbone conformational ensemble for docking a subset of the complexes. Finally, we explore a case study to investigate the efficacy of

pHDock in the prediction of large pH-dependent binding affinity change in a protein complex [30].

Results

pHDock algorithm

We developed pHDock, a multi-scale Monte Carlo (MC) algorithm based on the RosettaDock framework [27,28] with modifications to allow dynamic sampling of the residue protonation states during simulation. Residue protonation states at the environment pH are constantly updated during multiple side-chain packing steps throughout the protocol by explicitly sampling both protonated and deprotonated versions of the side chains from a discrete rotamer library [31].

The pHDock algorithm is illustrated in Fig. 1. In the first pre-packing step, the protein complex side chains are idealized, and the residue ionization states are equilibrated with the solution pH using Rosetta-pH [25]. Then, following the standard RosettaDock low-resolution stage, the residue side chains are represented by coarse-grained centroid atoms. This stage comprises i) a random initial perturbation of the partners, and ii) rigid-body ligand moves relative to the receptor which are accepted/rejected based on the Metropolis criteria. In the high-resolution stage, the side-chain centroid pseudo-atoms are replaced by the side-chain atoms from the initial unbound conformation. The high-resolution stage involves i) repacking the residue side chains with simultaneous evaluation of the most favorable residue protonation states at the environmental pH, and ii) minimization of the side-chain torsion angles and rigid-body orientation of the ligand relative to the receptor with an accompanying Metropolis criteria check. One thousand candidate structures, or models, are generated for each target and then ranked according to their interface scores, and the top-ranked model is picked as the final prediction.

To test the performance of the algorithm, we use both standard RosettaDock (henceforth referred to as simply 'RosettaDock') and pHDock to generate local docked models starting from a dataset of unbound structures from the curated Docking Benchmark 4.0 [29]. For pHDock, we assume the crystallization pH of the corresponding bound complex as the solution pH. In the following sections, we first illustrate the docking performance analysis of the new algorithm using a sample protein complex. Next, we compare the performance of pHDock to RosettaDock over the complete benchmark dataset using several metrics and inspect a few predictions in greater detail. We later focus on the effects of backbone flexibility on the docking accuracy. Finally we use a case study to demonstrate pHDock's performance in the prediction of pH effects on binding affinities.

Sample docking analysis: Xylanase–TAXI-IA binding at non-standard pH

Performance of structural docking algorithms can be analyzed by studying the distribution plots of the free energies or score function vs. the deviation from the starting native bound complex. The native complex is assumed to be at the free energy minimum, hence structural models generated using the docking algorithm with receptor-ligand orientation close to the native structure are expected to have lower energies compared to the structures farther away. To create a set of models sampling both near-native and non-native conformations, starting positions of the ligand relative to the receptor are perturbed by up to 3 Å translation and 8° rotation around the axis joining the centers of the two partners.

Fig. 2 shows sample plots for the *Triticum aestivum* xylanase inhibitor-I (TAXI-I) in complex with *Bacillus subtilis* xylanase crystallized at a pH of 4.6 (PDB: 2B42 [32]). The *y*-axis represents
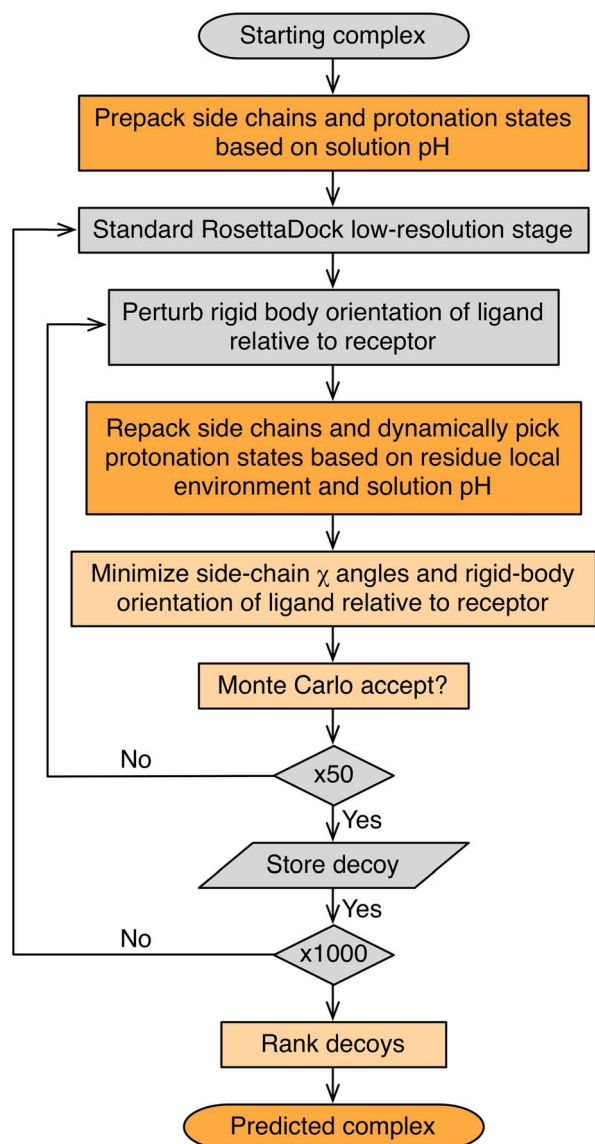
Fig. 1. pHDock flowchart. Each step in the pHDock workflow is colored based on the differences compared to RosettaDock: unmodified steps are colored in grey, and steps with minor (light orange) and major (dark orange) modifications are colored in shades of orange.
doi:10.1371/journal.pcbi.1004018.g001

the interface score (Isc), an approximation of the binding free energy, normalized by the difference between the 5th and 95th percentile scores. The x-axis quantifies deviation from the native complex using interface RMSD (Irmsd). Each point on the plot represents a single docking model and is colored based on the CAPRI structural quality rating [33] (see Methods). The interface of the top-scoring pHDock-generated structure (Fig. 2B) is just 1.7 Å from the native interface, compared to 4.7 Å for the RosettaDock-generated structure (Fig. 2A). While RosettaDock does not generate any structures better than acceptable quality, pHDock produces a structure with higher native residue-residue contact recovery qualifying as medium quality.

We quantified the docking performance using a discrimination score [34] (shown in bottom right in the docking score plot), which captures the extent to which the low-rmsd models have lower energies compared to the high-rmsd (incorrect) models. The

discrimination score is calculated by dividing the x-axis using multiple Irmsd cut-offs and averaging the energy gaps between the lowest scoring structure on the left and right of each cut-off (see Methods). A lower discrimination score is an indicator of better docking performance, with a negative score indicating a successful docking prediction. The additional side-chain protonation state sampling helps pHDock produce a successful and more pronounced docking funnel (discrimination score: -1.19) compared to RosettaDock (discrimination score: 0.16).

Fig. 2C compares the interfaces of the crystal structure and the top-ranked pHDock model for the xylanase–TAXI-IA complex. Experimental studies [32,35] discussed the importance of the strong salt bridge between the positively charged imidazole side chain of TAXI-IA His-374 (spheres) with the negatively charged Asp-37. This ionic interaction is critical for binding, and the pH optimum of the xylanase (determined by the pK_a value of Asp-37) is reported to directly influence the affinity of the enzyme–inhibitor complex, with a lower Asp pK_a value leading to stronger binding. The top-scoring pHDock model not only captures this interaction through precise prediction of the positively charged His-374 side-chain rotamer but also recovers all the xylanase active-site-residue side-chain rotamers. RosettaDock, which assumes a neutral His side chain, fails to capture the interaction. Overall, while the top-scoring RosettaDock model recovers just 13% of the native interface contacts, the pHDock model recovers 49% of all the interface residue-residue contacts.

pHDock improves docking accuracy in a majority of docking targets

For a large-scale docking performance analysis, we tested pHDock over a dataset of diverse protein–protein complexes from the curated Docking Benchmark 4.0 [29]. On average, 25% of the interface residues in the dataset complexes are ionizable (Asp, Glu, His, Tyr, Lys) (S1 Figure). Fig. 3 compares the discrimination scores of the docking funnels generated using pHDock and RosettaDock. pHDock produces successful docking funnels (discrimination score ≤ 0) in approximately half (79/161) the structures from the dataset, including 19 cases where RosettaDock fails to produce a successful prediction. Based on the discrimination score, pHDock outperforms RosettaDock in approximately 60% of the targets (94/161) (Table S2), and the improvements are statistically significant (paired t-test, $p = 0.039$). Additionally, since models are generated stochastically, we performed bootstrap case resampling [36] to quantify the variation of the discrimination scores. The bootstrap mean discrimination scores $\mu(D)$ (S2 Figure) again show that pHDock produces successful funnels [$\mu(D) \leq 0$] in half the targets (79/161) including 17 cases where RosettaDock fails. Hence the results are robust to the stochastic sampling noise. The average standard deviation of the discrimination scores [$\sigma(D)$: 0.07] is approximately 4% of the total observed $\mu(D)$ range.

As pHDock has access to nonstandard residue protonation states unlike RosettaDock, we examined the prevalence of such protonation states and their effect on docking accuracy. In docking funnel plots in S9 Figure, structures with nonstandard residue protonation states are distinguished. pHDock produces models with nonstandard protonation states for all the target complexes (S3 Figure), with a majority of the nonstandard protonation states observed in complexes with docking pH within one pH unit of the residue intrinsic pK_a values (S4 Figure). Overall, pHDock outperforms RosettaDock in 67% (20/30) of the cases where the top-ranked pHDock model recovers a nonstandard protonation state observed in the native bound complex (S5 Figure). pHDock also performs better than RosettaDock in 64% (7/11) of the cases where the top-ranked pHDock produces a nonstandard
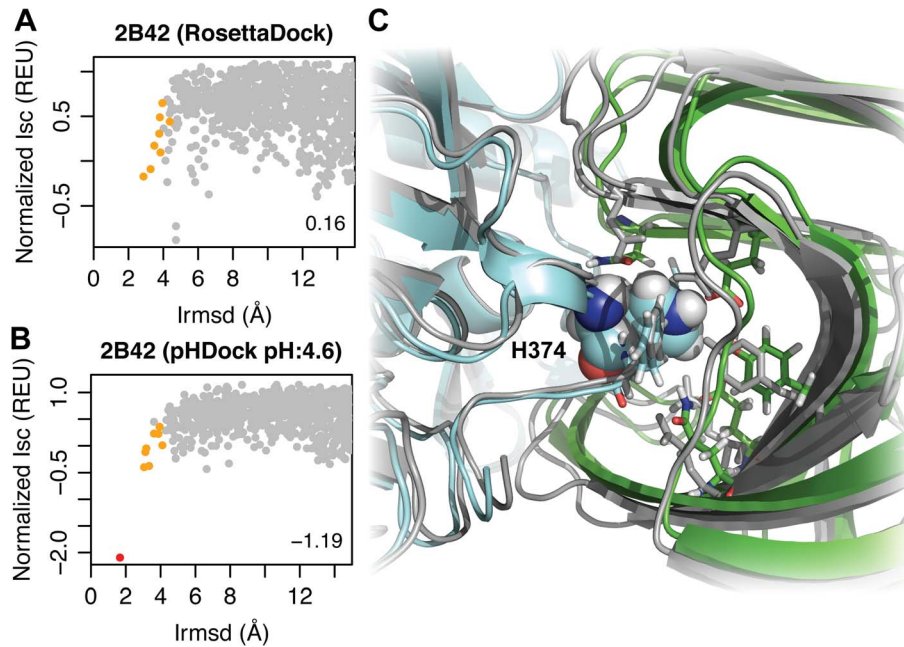
Fig. 2. Docking predictions for xylanase – TAXI-IA complex. Docking plots generated by (A) RosettaDock, and (B) pHDock at pH 4.6. Grey, orange, and red points represent incorrect, acceptable-, and medium- quality predictions, respectively. Discrimination scores are shown in the bottom right corner of the plots. (C) Interface of the top-scoring pHDock prediction (medium accuracy) superimposed on the crystal complex (grey) (2B42 [32]). Predicted orientation of the TAXI-IA inhibitor and xylanase, cyan and green, respectively; critical His-374 residue from TAXI-IA, spheres; xylanase active site and other critical binding site residues, sticks.
doi:10.1371/journal.pcbi.1004018.g002

protonation state different from the one observed in the native bound complex illustrating the importance of dynamic protonation states.
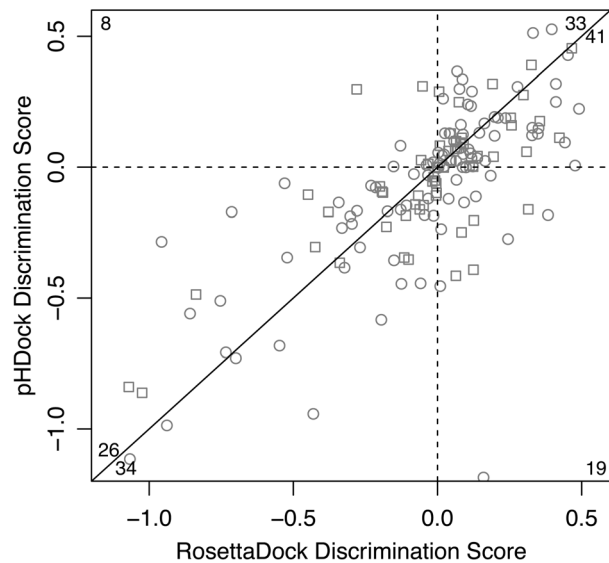

Fig. 3. Summary of pHDock performance. Correlation plot comparing discrimination scores of pHDock and RosettaDock docking predictions for each target in the complete benchmark dataset. Complexes docked at acidic pH (pH≤7.0) and basic pH (pH>7.0) are represented as circles and squares, respectively. The discrimination score cutoffs for a successful prediction ($D<0$) are marked using broken lines. Corner numbers indicate the total predictions in each plot section (edges defined by the broken lines and the solid line at 45°).
doi:10.1371/journal.pcbi.1004018.g003

Since pHDock is a stochastic docking algorithm that generates several candidate models, the performance of the algorithm broadly depends on (i) the quality and diversity of the generated ensemble of models, or 'sampling', and (ii) the ability of the final score function to discriminate native-like models from non-native-like models, or 'scoring'. To test the sampling performance of pHDock, we examined the lowest-Irmsd models for all the complexes in the dataset. The Irmsd distribution for pHDock is similar to RosettaDock (Fig. 4A), and in 92% of the docking targets, it generates at least one model within 4 Å from the native interface. Out of 1000 models generated for each target, pHDock creates on average 1.9, 18.5, and 90.8 high-, medium-, and acceptable-quality models, respectively. In comparison, Rosetta-Dock samples 7–12% fewer medium- and high-quality models (S6 Figure). To test the scoring performance of pHDock, we calculated the Irmsd and f_{nat} distributions of the top-scoring models for each target (Figs. 4B–C). pHDock generates top-ranked models within 4 Å in 57% of the targets (RosettaDock 51%), and 52% of the time these models recover more than 30% of the native residue-residue contacts (RosettaDock 46%).

To further assess the quality of the predicted top-ranked structures, we examined the receptor-ligand interface hydrogen bonds (henceforth referred to as simply 'interface hydrogen bonds'). Previous surveys found 8–13 interface hydrogen bonds in each protein–protein complex [37,38]. Using Rosetta's hydrogen bonding definition, the native crystal complexes in our dataset contain 6.4 ± 3.5 interface hydrogen bonds on average (Fig. 5A). In comparison, the top pHDock models are involved in 5.1 ± 2.5 interface hydrogen bonds, while the top RosettaDock models form only 3.4 ± 2.1 interface hydrogen bonds. As pHDock primarily focuses on ionizable residues, we also calculated the number of interface hydrogen bonds containing such residues as donors or acceptors. The native complexes contain 3.5 ± 2.6 ionizable
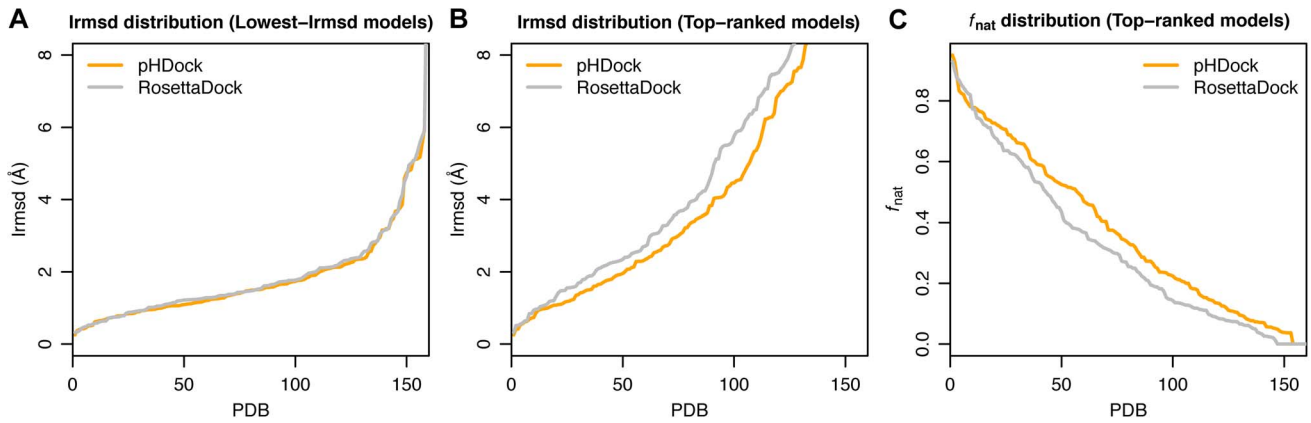
Fig. 4. Distribution curves of interface RMSDs (Irmsd) and fraction of recovered native contacts (f_{nat}) for the docking models. (A) Irmsd distribution curve of the lowest-Irmsd models generated using pHDock (orange) and RosettaDock (grey). (B, C) Irmsd and f_{nat} distribution curve for the top-ranked models according to interface scores (Isc) for each protein complex. The distribution curves are generated after independent sorting of the pHDock and RosettaDock models based on (A, B) increasing Irmsd values and (C) decreasing f_{nat}.
doi:10.1371/journal.pcbi.1004018.g004
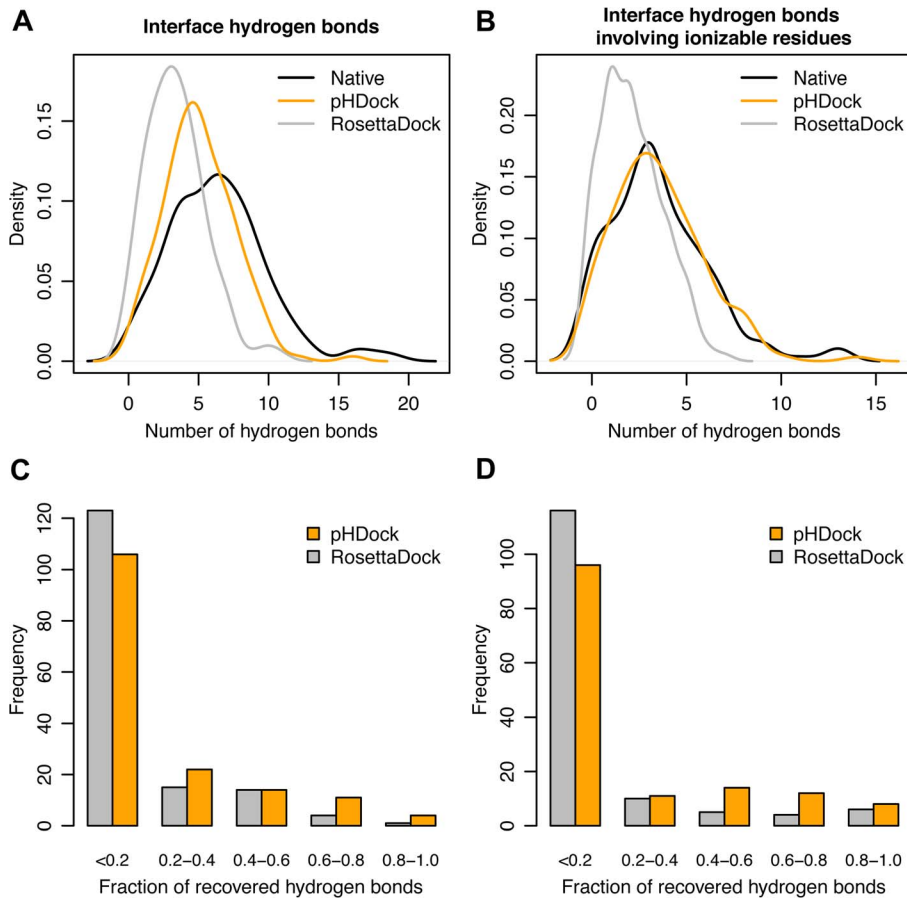
Fig. 5. Distributions of native and model interface hydrogen bonds. Kernel density estimate curves for the number of (A) interface hydrogen bonds and (B) interface hydrogen bonds involving ionizable residues in the top-scoring models generated using pHDock (orange) and RosettaDock (grey), and the native crystal complexes (black) across the complete Docking Benchmark dataset. Frequency histograms of the fraction of (C) recovered interface hydrogen bonds and (D) recovered interface hydrogen bonds involving ionizable residues in the top-scoring models.
doi:10.1371/journal.pcbi.1004018.g005

interface hydrogen bonds (Fig. 5B). Encouragingly, the top pHDock models are found to form an identical 3.5±2.4 ionizable interface hydrogen bonds, while the top RosettaDock models form only 2.1±1.6 hydrogen bonds.

The analysis of the total number of interface hydrogen bonds shows significant pHDock improvements in generating models with a larger receptor-ligand hydrogen bond network. However, such an analysis does not reveal the accuracy of the generated interface hydrogen bonds. So we also examined the fraction of the native interface hydrogen bonds recovered in the top-ranked models. pHDock recovers more than one-fifth of the native interface hydrogen bonds in only 33% of the targets from the dataset, while RosettaDock performs worse, recovering the same fraction in just 22% of the targets (Fig. 5C). The results are similar for the fraction of recovered ionizable interface hydrogen bonds. pHDock recovers more than one-fifth of the ionizable interface hydrogen bonds in 32% of the targets, while the performance of RosettaDock drops further to just 19% of the total targets in the dataset (Fig. 5D). In summary, while pHDock generates more interface hydrogen bonds, only a minor faction of these hydrogen bonds match those seen in the native complex.

Finally, to test the effects of hydrogen bonding accuracy on docking results, we examined a few sample cases in greater detail. The tumor susceptibility gene 101 protein–ubiquitin complex (1S1Q; pH 4.6 [39]) has four native interface hydrogen bonds. The top pHDock model recovers three of them and forms a total five interface hydrogen bonds, while the top RosettaDock model exhibits three interface hydrogen bonds but none of them are native. The docking plots for both pHDock and RosettaDock (discrimination score −0.19 vs. −0.01) show success based on discrimination scores, but the docking funnel is clearly more pronounced in pHDock (Fig. 6A). Although the near-native sampling in both pHDock and RosettaDock is comparable, the additional recovered native hydrogen bonds help pHDock in the final scoring, and the top model interface is only 1.4 Å away from the native interface. The improved performance is likely due to a protonated interface histidine (His-66) in ubiquitin. In a second case, the PPARgamma+ RXRalpha–GW409544+co-activator peptide complex (1K74; pH 7.5 [40]) has five interface hydrogen bonds. The top pHDock model exhibits eight interface hydrogen bonds, three of them being native, while none of the ten hydrogen bonds found in the top RosettaDock model are native (Fig. 6B). In this case, pHDock (discrimination score −0.35) outperforms Roset-taDock (discrimination score −0.12) in both sampling and scoring (Fig. 6B). The top-scoring pHDock model is a high-quality prediction just 0.93 Å from the native interface. In this case, the interface residues are all in their standard protonation states; we infer that the improvement must be due to kinetic effects during the Monte Carlo docking search. The larger number of interface hydrogen bonds in pHDock models do not always translate to improvements in docking predictions. For example, the CDK2 kinase–cell cycle-regulatory protein CksHs1 complex (1BUH; pH 7.5 [41]) has four native hydrogen bonds. Again, the interface residues in the top-ranked pHDock model are predicted to be in their standard protonation states. Neither top pHDock nor RosettaDock models recover any of the native interface hydrogen bonds although they form nine and one interface hydrogen bonds, respectively. As shown in the docking plots in Fig. 6C, pHDock scoring favors a false-positive docking prediction with a large number of interface hydrogen bonds more than 12 Å from the native interface.

Backbone flexibility further improves native contacts and hydrogen bond recovery

Inclusion of backbone flexibility in protein-protein docking is critical to capture the conformational changes during the binding event [42]. Within RosettaDock, backbone flexibility mimicking both conformer selection (CS) and induced fit (IF) binding models increases native contact recovery, although the computational costs are higher and there is a risk of false positive predictions [43]. Thus we tested whether the addition of backbone flexibility further improved native contact recovery in pHDock. We chose a subset of 14 complexes common among the published study and the curated Docking Benchmark 4.0 used for pHDock (Table S3). We then used the RosettaRelax [44,45] protocol to generate an ensemble of unbound backbones. RosettaRelax, an MC algorithm, employs a cycle of small backbone dihedral (φ, ψ) perturbations, residue side-chain packing and score function minimization along the gradient in the torsion space to generate a backbone ensemble typically within 1 Å C_α RMSD of the starting structure. We generated 500 models starting from the ligand unbound coordinates for each of the complexes and picked the ten top-scoring models for docking.

S10 Figure compares the docking funnels generated using RosettaDock, pHDock and ensemble pHDock. The ligand backbone flexibility helps ensemble pHDock generate better docking funnels (based on discrimination score) in 11 targets compared to pHDock. The Irmsd values of the lowest-Irmsd models generated using ensemble pHDock are not significantly better compared to pHDock. However, there is a noticeable improvement in the quality of the receptor-ligand interfaces in the top-ranked models. The top-ranked models generated using ensemble pHDock outperform pHDock in native contact recovery with comparable or better f_{nat} values in 12 targets. Encouragingly, the top-ranked models also recover comparable or more native interface hydrogen bonds in all the targets compared to pHDock and RosettaDock (Table S3). To summarize, the additional backbone flexibility further improves the docking funnel quality in a majority of the targets and generates top-ranked models that recover more native contacts and hydrogen bonds.

pHDock is better at solution pH than pH 7 or using fixed, predetermined protonation states

pHDock simulates the complexes at solution pH and relies on dynamic residue protonation state sampling. To assess the individual contribution of these two components, we performed control docking experiments using a subset of complexes (same 14 complexes used for ensemble pHDock). First, to test the robustness of the docking predictions to changes in the solution pH, we used pHDock at physiological pH (pH 7.0). Second, to test the benefits of employing dynamic residue protonation states, we docked the complexes with fixed residue protonation states obtained from the lowest energy rotamer state of the starting partners at the solution pH (fix-pHDock).

Of the cases where both RosettaDock and pHDock either fail (four targets) or succeed (eight targets), the fix-pHDock and pHDock at pH 7.0 runs perform similarly (see docking funnel plots, S11 Figure), showing, as might be expected, an insensitivity to pH effects. There are two cases in this test set where RosettaDock fails and pHDock produces a successful docking funnel. In the α-chymotrypsin–eglin C complex (1ACB; pH 6.5 [46]), pHDock produces a discrimination score of −0.24 at pH 6.5, and RosettaDock a discrimination score of 0.01. pHDock at pH 7.0 produces a weaker funnel (discrimination score: −0.1) while fix-pHDock fails (discrimination score: 0.09) due to a false
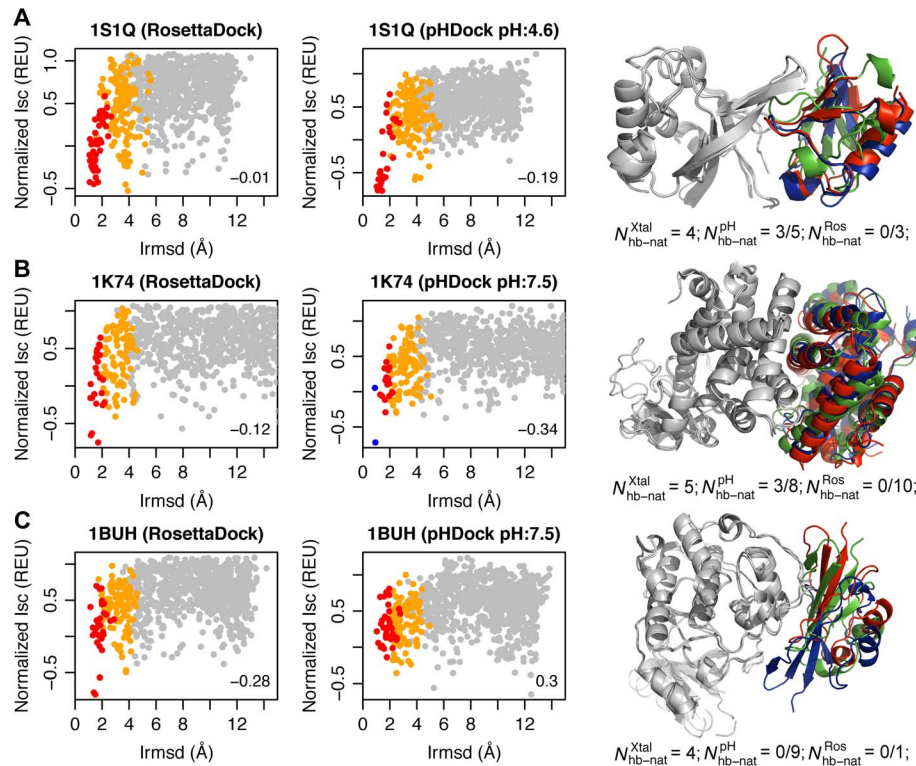
Fig. 6. Hydrogen bonding recovery correlates with docking performance. Docking plots generated using RosettaDock and pHDock for (A) tumor susceptibility gene 101 protein–Ubiquitin complex (1S1Q; pH 4.6), (B) PPARgamma+RXRalpha–GW409544+co-activator peptide complex (1K74; pH 7.5), and (C) CDK2 kinase–cell cycle-regulatory protein CksHs1 complex (1BUH; pH 7.5). Grey, orange, red, and blue points represent incorrect, acceptable-, medium-, and high-quality models, respectively. Discrimination scores are shown in the bottom right corner of the plots. The right panel shows structures of the top pHDock (blue) and RosettaDock (green) models superimposed on the native complex (red). The number of native hydrogen bonds among the total interface hydrogen bonds observed in the bound crystal complex($N_{\text{hb-nat}}^{\text{Xtal}}$), and the top-scoring pHDock($N_{\text{hb-nat}}^{\text{pH}}$)and RosettaDock($N_{\text{hb-nat}}^{\text{Ros}}$)models are also listed.
doi:10.1371/journal.pcbi.1004018.g006

positive model 7 Å Irmsd away from the native complex. Similarly, in the Fab D44.1–lysozyme complex (1MLC; pH 6.0 [47]), pHDock generates a discrimination score of −0.11 while RosettaDock, pHDock at pH 7.0, and fix-pHDock all fail (discrimination scores 0.13, 0.07, 0.33, respectively). Thus, in these two cases where RosettaDock fails, both pHDock at pH 7.0 and fix-pHDock fail to completely capture pHDock's success. These cases suggest that accurate knowledge of the solution pH and the dynamic protonation states are vital for maximum pHDock accuracy.

pHDock captures the large pH-dependent binding affinity change in the Fc–FcRn complex

In the discussion so far, we analyzed pHDock's performance at the solution pH and compared it to RosettaDock (no pH dependence) over a large dataset of protein complexes. However, such an analysis does not test pHDock's performance in predicting effects of subtle environmental pH changes on a single protein-protein complex. In previous work, we and other groups have used RosettaDock interface scores in correlating binding affinities [48,49] and in predicting relative affinities [50]. The neonatal Fc receptor (FcRn) binds maternal immunoglobulin G (IgG) from ingested milk in the gut at acidic pH (pH≤6.5) and releases it in the bloodstream of the newborn at basic pH (pH 7.4) [51]. This process is facilitated through a drastic drop in the binding affinity by more than two orders of magnitude as the pH changes from 6.0–6.5 to 7.0–7.5 [51,52]. The Fc–FcRn system has been

previously used for a pH-dependent binding calculation [16], but to our knowledge, there are no existing pH-sensitive docking studies.

To test the efficacy of pHDock in predicting pH effects on binding affinities, we used the pHDock algorithm to dock the murine Fc–FcRn complex (1I1A [30]) at various environmental pH values. We tested all integral pH values between 3.0 and 11.0, and used a finer interval of 0.25 pH units for the relevant pH range of 6.0–8.0 where the striking binding affinity change is observed. We used the interface scores (I) of the top-scoring pHDock models to approximate the binding affinity at different pH values. Fc–FcRn complex shows a binding minimum at pH 6.25 ($I_{\text{pH6.25}}$: −13.99 Rosetta Energy Units (REU)), and thereafter the affinity rapidly weakens as the environment pH increases to 7.50 ($I_{\text{pH7.50}}$: −11.82 REU) (Fig. 7A). Converting the binding energies to equilibrium constants using the relation $K_d = e^{-\Delta G/k_B T}$,we estimated the ratio of equilibrium constants at pH values 6.25 and 7.50 as

$$\frac{K_{\text{pH6.25}}}{K_{\text{pH7.50}}} = e^{\frac{-1}{k_B T}\left(I_{\text{pH6.25}} - I_{\text{pH7.50}}\right)} = 40,$$

where $K_{\text{pH6.25}}$ and $K_{\text{pH7.50}}$ are the equilibrium binding constants at pH 6.25 and 7.50, respectively, and $k_B T$ is 0.59 kcal/mol at 298K. The equation yields a 40-fold drop in the binding affinity as the pH increases from 6.25 to 7.50, which is similar to the 50 to
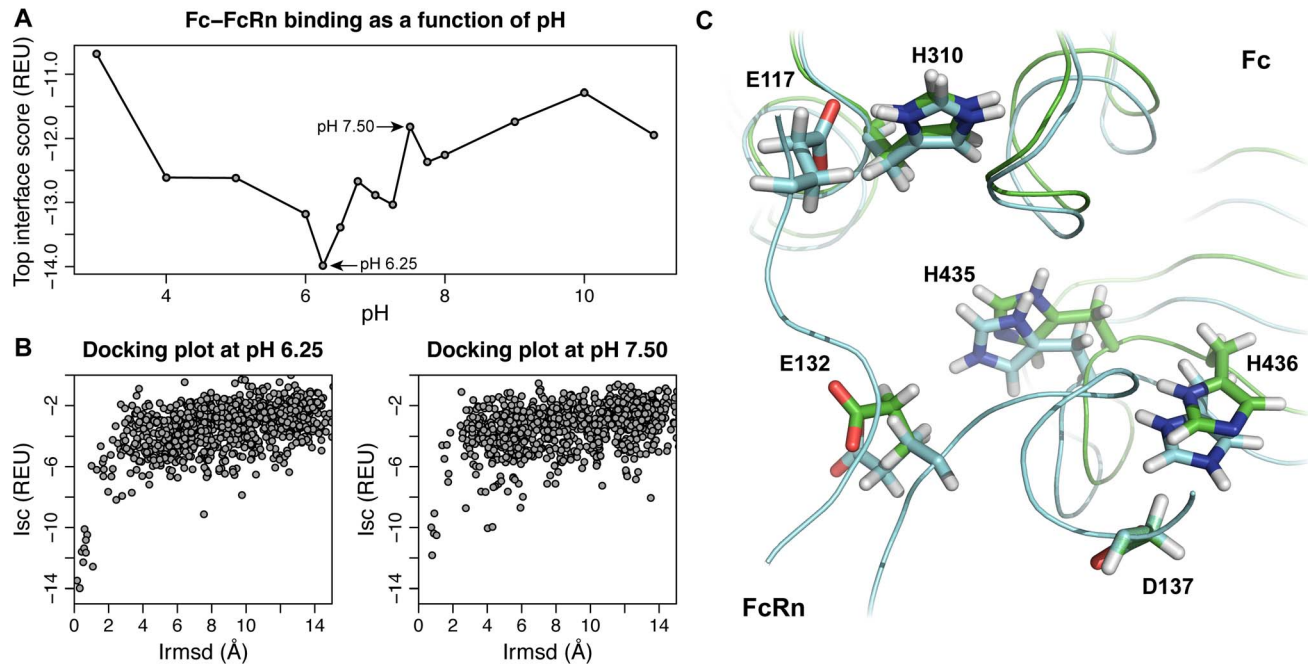
Fig. 7. pH-dependent binding effects in Fc–FcRn complex. (A) Interface score of the top pHDock prediction for the Fc–FcRn complex as a function of the docking pH. (B) Interface score vs Irmsd plots generated using pHDock at pH 6.25 and pH 7.50. (C) Top pHDock models at pH 6.25 (cyan) and pH 7.50 (green) showing the three critical ionic interactions responsible for the large pH-dependent binding affinity change. Note the change in the protonation states of His-435 and His-436.
doi:10.1371/journal.pcbi.1004018.g007

100-fold drop from experiments [52]. Interestingly, the docking plots show successful energy funnels for both pH values (Fig. 7B). However, the energy funnel is more pronounced at pH 6.25 (discrimination score −0.96) than pH 7.50 (discrimination score −0.47), indicating a site-specific binding event at both pH values, but with markedly different affinities.

Previous studies [30,51] attribute the pH-dependence of Fc–FcRn binding to the titration of interface histidine residues with pK_a values in the range of binding affinity transition (6.5≤pH≤ 7.0). The Fc–FcRn interface has three salt bridges with the residues His-310, His-435, and His-436 from Fc interacting with Glu-117, Glu-132, and Asp-137 from FcRn. The proposed mechanism involves titration of all the three histidine residues disrupting the binding as the environment pH increases, but studies have shown two buried titratable salt bridges are sufficient to confer pH dependence. Encouragingly, the top-scoring pHDock-generated models at different pH values successfully capture the titration event. While His-310 remains protonated in the models at both pH values, His-435 and His-436 are protonated at pH 6.25 and deprotonated at pH 7.50 and are involved in salt bridges with Glu-132 and Asp-137, respectively (Fig. 7C). Thus, pHDock not only predicts the relative Fc–FcRn binding affinities at different pH values, but also captures the expected physical mechanisms responsible for the different affinities.

Discussion

We have created pHDock, the first pH-sensitive protein-protein docking algorithm that samples residue protonation states dynamically during the search. The algorithm integrates the Rosetta-pH pK_a calculation method [25] with the RosettaDock framework using the object-oriented design of the Rosetta modeling suite [26]. Local docking studies show that pHDock

outperforms RosettaDock in 60% of the docking targets and also performs better than control cases involving docking at pH 7.0 or using fixed, predetermined protonation states. pHDock also shows encouraging improvements in the quality of the generated candidate predictions. On average, the top-ranked pHDock structures have lower interface RMSDs and recover more native residue-residue contacts and hydrogen bonds. While pHDock is designed to improve docking predictions by accounting for environmental pH effects, the successful prediction of a large pH-dependent binding affinity change in the Fc–FcRn complex suggests that it can be further exploited to improve affinity predictions.

pHDock improves docking primarily by enhancing the scoring in the docking high-resolution stage, as the improved score function finely tuned for pK_a predictions is active only during the high-resolution steps involving dynamic protonation states. Although there are few cases where pHDock samples conformations closer to the native compared to RosettaDock, the similarity of the interface RMSD distributions of the closest-sampled models (to the native complex) shows that its sampling quality is largely unchanged, likely because it retains the RosettaDock low-resolution stage which is largely responsible for model diversity. Over the complete dataset, pHDock generates at least one high-quality model in 25% of the complexes (41 targets), slightly higher than RosettaDock (34 targets). ReplicaDock [53], which uses a set of temperature replicas, overcomes the kinetic barriers and improves sampling in the low-resolution docking stage. Further work can thus focus on combining the principles of ReplicaDock with pHDock to improve the model diversity in the low-resolution centroid phase. Also, availability of even sparse biochemical information [54] can be used as an alternative to constrain the conformational search space and circumvent the sampling concerns in the centroid phase to improve docking accuracy.

Although the top-ranked pHDock models show significant advancements in recovering native contacts, the hydrogen bonding performance is mixed. The geometry of interface hydrogen bonds is less optimal than intra-chain hydrogen bonds, but they are nevertheless critical for protein-protein binding [37]. The top pHDock models exhibit more hydrogen bonds than RosettaDock on average. The increase is especially evident in the case of ionizable residues where the pHDock hydrogen bond distribution matches the native distribution. However, many of the pHDock interface hydrogen bonds are non-native, *i.e.*, they are not observed in the bound crystal complexes. In fact, in two-thirds of the targets, pHDock fails to recover more than one-fifth of the native interface hydrogen bonds, a shocking number revealing the limitations still present in the hydrogen bonding model.

There are a few possible explanations for the poor hydrogen bond performance. First, pHDock uses an implicit solvation model and thus fails to capture the water-mediated interface hydrogen bonds. Although the water-mediated hydrogen bonds are excluded from native hydrogen bond calculations, ignoring the water molecules during docking can result in the compensation of unsatisfied hydrogen bond donors/acceptors through formation of non-native hydrogen bonds. Second, pHDock ignores protein backbone flexibility and uses the unbound coordinates of the protein partners for docking, hence any resulting backbone inaccuracies can shift the hydrogen bond network. Accounting for backbone flexibility using a conformational ensemble for a small subset of complexes improves hydrogen bond recovery compared to pHDock, but the top-ranked models still recover just a quarter of the native interface hydrogen bonds. Further studies to improve hydrogen bond recovery can focus on calibrating the score function using the bound coordinates of the complex to minimize the errors introduced due to the rigid backbone assumption and the inaccuracies in the receptor-ligand orientation in the docking models. However, work will be needed to reconcile the changes with the docking score function that is tuned for recovering native-like structures.

We tested pHDock's ability to capture the large pH-dependent binding affinity change in the Fc–FcRn complex. Since the binding changes are a result of protonation state shifts in the interface histidine residues, any docking algorithm ignoring environment pH will fail to capture the effect. pHDock predicts a 40-fold drop in the binding affinity due to the increase in the environment pH, and the top-scoring model captures the resulting disrupted salt bridges at the Fc–FcRn complex interface. The accuracy of the affinity prediction suggests that pHDock can be expanded to power computational protein design studies such as those that recently began to exploit the pH-dependence for regulating protein binding activity [55]. Previously during the CAPRI rounds 20–27 [56], we used pHDock for the blind prediction of the g-type lysozyme–PliG inhibitor complex [50]. Lysozyme operates in a low pH environment [57] and hence provided an opportunity to test pHDock's performance. Docking the complex at pH 6.2 (crystallization pH of the unbound lysozyme) generated a medium-quality prediction just 2.0 Å from the interface of the native complex. The encouraging performance of pHDock proves that it can be effective in capturing environment-pH effects on both docking and binding.

Recent efforts have begun to capture structural details of protein interactions in complete cellular environments [58–60]. There is tremendous scope for computational docking algorithms to power such studies, but the methods must be versatile and include the effects of environmental conditions. Since intracellular pH is strictly regulated across multiple eukaryotic cellular compartments and is critical for protein interactions [61], accounting for pH effects can boost prediction accuracy. The results in this paper contribute to the community effort to simulate protein-protein interactions in the complete cell with all environmental factors.

Methods

Benchmark dataset

The Protein-Protein Docking Benchmark 4.0 by Hwang *et al.* [29] is a set of 176 non-redundant protein-protein complexes with both bound and corresponding unbound crystal coordinates from the Protein Data Bank [62]. The dataset comprises 121 'rigid-body', 30 'medium', and 25 'difficult' targets based on the interface backbone conformation variation between bound and unbound coordinates [63].

We curated the benchmark dataset in multiple stages. First, we removed water and all non-peptide molecules containing heteroatoms from the complex structures. Since Rosetta pH does not currently predict protonation states of non-peptide molecules, we excluded complexes with such molecules at the interface. We also eliminated structures in which Rosetta was unable to resolve the steric clashes in the starting atomic coordinates due to the conformational changes between bound and unbound complexes, leaving 161 test complexes for the study. Second, we truncated both the unbound and bound structures to the same amino-acid sequences for Rosetta scoring consistency. Third, we collected the crystallization pH values in the PDB coordinate file for each bound complex to determine the docking environment pH. For structures missing pH information in the PDB files, we used the pH value from the corresponding original research article if available. For the remaining structures, we assumed a physiological pH of 7.0 (Table S2).

Rosetta-pH

Rosetta-pH [25] is a Metropolis Monte Carlo algorithm in which the protonation state of the lowest energy conformation is evaluated using the Rosetta-pH score function at intervals of pH to estimate pK_a values. The Rosetta-pH score function is based on the standard Rosetta score function with additional terms including:

i) Protonation potential based on the probability of protonation of individual amino acid residues at a given pH. The probability of protonation (f_{prot}) of an amino acid is

$$f_{prot} = \frac{1}{10^{pH - IpK_a} + 1},$$

and the protonation potential (E_{pH}) is

$$E_{pH} = \begin{cases} -k_B T \ln f_{prot} & \text{if protonated} \\ -k_B T \ln(1 - f_{prot}) & \text{if deprotonated,} \end{cases}$$

where pH is defined by the environment, and IpK_a is the unperturbed intrinsic pK_a value of the model compound in solution (4.0 for Asp, 4.4 for Glu, 6.3 for His, 10.0 for Tyr and 10.4 for Lys). $k_B T$ is assigned a value of 0.59 kcal/mol, corresponding to $T = 298K$. Cys protonation state changes (intrinsic pK_a 8.5) are ignored due to the complications of coupling between pK_a and redox equilibrium [64].

ii) Coulomb electrostatic potential with a distance-dependent dielectric ($\varepsilon = 10r$) for gradual shielding at increasing interatomic distances [65], and

iii) Recalibrated solvation reference energies (ΔG_i^{ref}) for the non-standard protonation variants in the Lazaridis–Karplus implicit model for solvation [66] (See [25] for details).

pHDock development

Rosetta pHDock uses the object-oriented design of the Rosetta biomolecular modeling suite [26] to implement the environment pH effects in the RosettaDock protocol. The pHDock development workflow can be broadly classified into three stages:

i) In the first stage, we incorporated explicit protonation state sampling from Rosetta-pH [25] into the RosettaDock algorithm. RosettaDock accounts for residue side chain flexibility in the prepacking step and the later high-resolution stage with full-atom side chains. The sampling of the side-chain χ-angles is discrete based on a backbone-dependent rotamer library [31]. Rosetta pHDock augments the sampling by allowing variable residue ionization states to be simultaneously sampled during every side-chain packing step and picking the most favorable residue protonation state based on the residue's local interactions and the solution pH. For neutral His, both possible tautomers (with proton on either $N_{\delta 1}$ or $N_{\varepsilon 2}$ atoms) are sampled. The conformational degeneracy in the protonated variants of Asp and Glu (with H atoms on either of the terminal O_δ and O_ε atoms, respectively) is also explicitly incorporated by accommodating both possible protonated versions for the residues during sampling.

ii) In the second stage, we generated a dataset of structures and evaluated the contributions of the individual score terms (including e_pH) to the total interface score. We first generated 1000 models (for each complex) using the standard RosettaDock local docking routine [28] on a subset of 60 randomly-selected bound complexes ($\sim 1/3$ of the total docking benchmark set). We then repacked each model (sampling both side chains and protonation states) at the crystal pH of the bound complex and calculated the interface contribution of each score term (E_i) as

$$E_i = E_i^{\text{complex}} - \sum_{j = \text{partners}} E_i^{j,\text{res(int}\cap\text{ion)}}$$

where E_i^{complex} is the contribution of the score term i in the repacked complex, and $E_i^{j,\text{res(int}\cap\text{ion)}}$ is the score term contribution in each separate binding partner j after repacking the ionizable interface residues at the crystal pH of the bound complex. Repacking the ionizable residues is required for accurate score term estimation, as separation of the binding partners exposes the previously-buried interface residues to the solvent affecting their preferential protonation state.

iii) In the third stage, we parameterized the pHDock score function. Reweighting is mandatory since the original RosettaDock score function had a minimal weight on electrostatics, and the new electrostatic weight and pH reference term must be rebalanced against the hydrogen bonding and solvation contributions. Similar to prior parameterization of the RosettaDock score function [27], we sought to maximize the free energy gap between 'near-native' and 'non-native' models. Models in the top 5% based

on CAPRI rating [33] (high, medium and acceptable-quality in that order) with repulsive van der Waals scores lower than the 80th percentile are classified as near-native models. Models with the same CAPRI rating are ordered based on the f_{nat} values (higher f_{nat} is better). We classified the remaining models as non-native models. We then derived the score term weights using a generalized linear regression to maximize the free energy gap between the near-native and non-native model clusters. The free energy gap (ΔE) is

$$\Delta E = \sum_{\substack{\text{targets}}} \sum_{\substack{\text{score} \\ \text{terms}}} w_i \Delta E_i$$

where w_i is the weight for score term E_i and $\Delta E_i = E_i^{\text{non-native}} - E_i^{\text{near-native}}$. The score terms include an attractive van der Waals score (E_{atr}), a repulsive van der Waals score (E_{rep}), an implicit solvation score (E_{sol}) [66], a hydrogen bonding score (E_{hb}) [67], rotamer probability term (E_{dun}) [31], a statistical residue pair term for ion-ion interactions (E_{pair}) [68], a Coulomb electrostatic term (E_{elec}), and a term for the pH effects (E_{pH}) [25].

Table S1 compares the optimized pHDock weights to the RosettaDock weights. The new pHDock weights for the dominant score terms E_{atr}, E_{sol}, and E_{hb} show small deviations compared to RosettaDock (0.377, 0.225, and 0.249 *versus* 0.338, 0.242, and 0.245). Besides the new addition of pH-sensitive score term E_{pH} (weight 0.21), the major changes in the score function are in the score term weights for E_{pair}, E_{elec}, E_{dun}, and E_{rep}. The E_{pair} term is completely absent and is balanced by the increased E_{elec} weight (0.319 compared to 0.026 in RosettaDock). While the E_{dun} weight also increases (0.036 to 0.080), the E_{rep} weight drastically drops from 0.044 to 0.005 demonstrating that the repulsive van der Waals score does not aid in docking model discrimination. The exceptionally small E_{rep} weight however creates two issues. First, the algorithm produces structures with steric clashes during the rigid-body minimization step in the docking high-resolution stage (Fig. 1). RosettaDock [27] addresses this issue by increasing the E_{rep} weight during minimization using a multiplier. We followed the same strategy and raised the E_{rep} weight to match the RosettaDock weight during minimization. Second, some structures with unfavorable sterics are ranked higher during the final model discrimination. To address this, we eliminated the worst 5% percent of the pHDock structures sorted by their E_{rep} scores. For a balanced comparison, we also omitted the worst 5% of the RosettaDock structures sorted by their interface scores.

Docking starting conformation generation

In local docking, the input complex consists of unbound partners (orientation determined by superimposing on the coordinates of the bound complex) and the starting positions are generated by randomly perturbing the ligand relative to the receptor by up to 3 Å translation and 8° rotation around the axis joining the centers of the two partners. Both pHDock and RosettaDock use local docking to generate a diverse set of models sampling both near-native (Irmsd <4 Å) and non-native (Irmsd> 4 Å) conformations around the binding site.

Docking metrics

The CAPRI structural quality rating [33] classifies docking predictions as incorrect, acceptable-, medium-, or high-quality based on a combination of the metrics Lrmsd, Irmsd, and f_{nat}.

L_rmsd is defined as the root-mean-square deviation (RMSD) of the ligand C_α atoms after superposition of the receptor chains of the predicted and the native bound complexes. Irmsd is the C_α-atom RMSD after superposition of the interface residues (residues <4.0 Å from the binding partner) with coordinates from the bound complex. f_{nat} is the fraction of the residue-residue contacts (<5.0 Å all-atom distance) in the native bound complex that are recovered in the predicted complex. CAPRI ratings depend on multiple criteria, but models are considered to be at least acceptable quality if they are within 4 Å from the native interface and recover at least 30% of the native contacts (f_{nat}) [33].

Docking funnel metrics

A 'docking funnel' derives its name from the funnel-like appearance of the target score vs RMSD plots where the near-native models have better scores than non-native models. It is often used as a measure to determine the success of a docking simulation. We used two different metrics to quantify docking funnels.

i) N_5: As defined by Chaudhury et al. [28], N_5 is the number of models with an Irmsd of at most 4.0 Å among the five top-scoring structures based on interface score. A docking result is considered a success if $N_5 \geq 3$. We performed bootstrap case resampling (1000 models per target with replacement) to compare correlation between the mean $\mu(N_5)$ and calculated N_5, and to quantify the inherent noise within set of models using the standard deviation $\sigma(N_5)$ (S7 Figure).

ii) Discrimination score (D): Applying the formulation by Conway et al. [34] to docking, we first normalize the model interface scores ($\hat{I}$) using the 5th and 95th percentile scores as the reference by assigning them values of 0 and 1, respectively. The models are then divided into clusters based on Irmsd with cut-offs from $\mathbb{R} = \{1.0, 1.5, 2.0, 2.5, 3.0, 4.0, 6.0\}$ in Ångstroms. Discrimination score (D) is defined as the normalized interface score difference of the lowest-energy model below and above each cut-off $r \in \mathbb{R}$, averaged over the number of cut-offs (N_r):

$$D = \frac{1}{N_r} \sum_{r \in \mathbb{R}} \min_{i \ni RMS(i) \leq r} \hat{I}_i - \min_{i \ni RMS(i) > r} \hat{I}_i$$

A docking result is considered a success if $D \leq 0$. We performed bootstrap case resampling (1000 models per target with replacement) to quantify the inherent noise within the set of models using the standard deviation $\sigma(D)$ (S2 Figure).

Algorithm availability

pHDock is part of the Rosetta biomolecular modeling suite (www.rosettacommons.org) which is freely available for academic and non-profit use. The Supporting Information includes the complete list of structures from the docking benchmark dataset with the corresponding pH values and the command-line syntax for using pHDock method in Rosetta. Component methods and objects are also available in the PyRosetta libraries (www.pyrosetta.org) [69].

Supporting Information

S1 Figure Distribution of ionizable residues at docking interfaces. Frequency histogram of the number of dataset complexes with various fractions of ionizable interface residues (Asp, Glu, His, Tyr, Lys).
(TIF)

S2 Figure Discrimination score (*D*) distributions for RosettaDock and pHDock algorithms. Mean D ($\mu(D)$) values obtained from bootstrap case resampling of the docking models (1000 models per target with replacement) for pHDock (orange) and RosettaDock (grey). Standard deviations ($\sigma(D)$) are represented as error margins. The average $\mu(D)$ value for pHDock (-0.05) is lower than RosettaDock (-0.02) over the complete dataset. The average $\sigma(D)$ values for pHDock (0.07) and RosettaDock (0.07) are similar, approximately 4% of the observed $\mu(D)$ value range. The distribution curves are generated after independent sorting of the pHDock and RosettaDock targets based on increasing D values.
(TIF)

S3 Figure pHDock models containing nonstandard residue protonation states. Number of near-native (Irmsd <4 Å) (black) and non-native (Irmsd>4 Å) (grey) pHDock models containing nonstandard residue protonation states (protonated Asp, Glu, His; deprotonated Tyr, Lys) for each target complex in the curated docking benchmark dataset. For almost all pHDock target complexes (160/161), at least one non-native model exhibits a nonstandard protonation state, while for approximately 4/5 of the complexes (127/161), at least one near-native model has nonstandard residue protonation states. The complexes are sorted based on the crystallization pH.
(TIF)

S4 Figure Nonstandard residue protonation states in pHDock models. (A) Number of ionizable residues exhibiting nonstandard protonation states in pHDock models for each target complex. The number of recovered nonstandard residue proton-ation states (compared to the protonation state in the native bound complex) in (B) near-native and (C) non-native pHDock models are also shown. The complexes are sorted based on the crystallization pH. A majority of the nonstandard residue protonation states are observed in complexes with docking pH within one pH unit of the residue intrinsic pK_a values (Asp 50%, Glu 78%, His 59%, Tyr 53%, Lys 70%). Only a small fraction of all the pHDock-generated nonstandard protonation states (Asp 17%, Glu 30%, His 70%, Tyr 34%, Lys 66%) are recovered nonstandard residue protonation states that are also observed in the native bound complex.
(TIF)

S5 Figure Summary of pHDock performance highlight-ing cases with nonstandard protonation states. Correla-tion plot comparing discrimination scores of pHDock and RosettaDock docking predictions for each target in the complete benchmark dataset. This plot is the same as Fig. 3 in main manuscript. However, here, grey, orange and red points represent complexes where top-ranked pHDock models contain no nonstandard protonation states, recovered nonstandard protonation states found in the native bound complex, and nonstandard protonation states not observed in the bound complex, respectively. Complexes docked at acidic pH (pH≤ 7.0) and basic pH (pH>7.0) are represented as circles and squares, respectively. The discrimination score cutoffs for a successful prediction (*D*<0) are marked using broken lines. Corner numbers indicate the total predictions in each plot section (edges defined by the broken lines and the solid line at 45°). Overall, pHDock outperforms RosettaDock in 67% (20/30) of the cases where the top-ranked pHDock model recovers a

nonstandard protonation state observed in the native bound complex. pHDock also performs better than RosettaDock in 64% (7/11) of the cases where the top-ranked pHDock produces a nonstandard protonation state different from the one observed in the native bound complex illustrating the importance of dynamic protonation states.
(TIF)

S6 Figure Quality of models sampled during docking. Kernel density estimate curves of the number of high-, medium-, and acceptable-quality models sampled by pHDock and Rosetta-Dock during a docking run generating 1000 models. Numbers in the parentheses in the legends are the average number of the various quality models sampled by the docking algorithms.
(TIF)

S7 Figure N_5 distributions for RosettaDock and pHDock algorithms. Mean N_5 ($\mu(N_5)$) values obtained from bootstrap case resampling of the docking models (1000 models per target with replacement) for pHDock (orange) and RosettaDock (grey). Standard deviations ($\sigma(N_5)$) are represented as error margins. The average $\mu(N_5)$ values for pHDock (2.60) and RosettaDock (2.55) are similar over the complete dataset. The average $\sigma(N_5)$ values are high for both pHDock (0.65) and RosettaDock (0.65), approximately 13% of the observed $\mu(N_5)$ range, indicating significant inherent noise using the N_5 metric for the set of models. The distribution curves are generated after independent sorting of the pHDock and RosettaDock targets based on decreasing N_5 values.
(TIF)

S8 Figure Docking plots for pHDock and RosettaDock. Grey, orange, red, and blue points represent incorrect, acceptable-, medium-, and high-quality predictions, respectively. Discrimination scores are shown in the bottom right corner of the plots.
(PDF)

S9 Figure Docking plots for pHDock and RosettaDock highlighting models with nonstandard residue protonation states. Grey, orange and red points represent models containing no nonstandard protonation states, recovered nonstandard protonation states found in the bound complex, and nonstandard protonation states not observed in the bound complex, respectively. Discrimination scores are shown in the bottom right corner of the plots.
(PDF)

S10 Figure Docking plots comparing ensemble pHDock to pHDock and RosettaDock. Grey, orange, and red points represent incorrect, acceptable-, and medium- quality predictions,

respectively. Discrimination scores are shown in the bottom right corner of the plots.
(PDF)

S11 Figure Docking plots comparing RosettaDock, pHDock at crystallization pH, pHDock at pH 7.0 and FixpHDock. Grey, orange and red points represent models containing no nonstandard residue protonation states, recovered nonstandard residue protonation states found in the bound complex, and nonstandard residue protonation states not observed in the bound complex, respectively. In FixpHDock, the protonation states found in the starting unbound complex are held constant during docking. Discrimination scores are shown in the bottom right corner of the plots.
(PDF)

S1 Table Score functions used for the study. Weights for the score terms used during residue side-chain and protonation state sampling, receptor-ligand minimization, and for ranking final docking models.
(PDF)

S2 Table Docking performance summary. PDB IDs and pH values of the benchmark dataset used for the study. Discrimination scores, N_5 values, Irmsd and f_{nat} of the lowest-Irmsd and top-ranked models generated using pHDock and RosettaDock are also listed.
(PDF)

S3 Table Ensemble pHDock performance summary. PDB IDs and pH values of the benchmark subset used for ensemble pHDock. Discrimination scores, Irmsd, f_{nat}, and number of recovered native interface hydrogen bonds in the top-ranked models generated using ensemble pHDock are compared to pHDock and RosettaDock.
(PDF)

S1 Text Command lines for pHDock, RosettaDock and ensemble pHDock.
(PDF)

Author Contributions

Conceived and designed the experiments: KPK JJG. Performed the experiments: KPK KR. Analyzed the data: KPK KR JJG. Contributed reagents/materials/analysis tools: JJG. Wrote the paper: KPK JJG.

References

1. Schönichen A, Webb BA, Jacobson MP, Barber DL (2013) Considering Protonation as a Posttranslational Modification Regulating Protein Structure and Function. Annu Rev Biophys 42: 289–314. doi: 10.1146/annurev-biophys-050511-102349.
2. Garcia-Moreno B (2009) Adaptations of proteins to cellular and subcellular pH. J Biol 8: 1–4. doi: 10.1186/jbiol199.
3. Sheinerman FB, Norel R, Honig B (2000) Electrostatic aspects of protein-protein interactions. Curr Opin Struct Biol 10: 153–159. doi: 10.1016/S0959-440X(00)00065-8.
4. Warshel A, Dryga A (2011) Simulating electrostatic energies in proteins: Perspectives and some recent studies of pKas, redox, and other crucial functional properties. Proteins 79: 3469–3484. doi: 10.1002/prot.23125.
5. Mitra RC, Zhang Z, Alexov E (2011) In silico modeling of pH-optimum of protein–protein binding. Proteins 79: 925–936.
6. Onufriev AV, Alexov E (2013) Protonation and pK changes in protein–ligand binding. Q Rev Biophys 46: 181–209. doi: 10.1017/S0033583513000024.
7. Wass MN, Fuentes G, Pons C, Pazos F, Valencia A (2011) Towards the prediction of protein interaction partners using physical docking. Mol Syst Biol 7: 469. doi: 10.1038/msb.2011.3.
8. Lopes A, Sacquin-Mora S, Dimitrova V, Laine E, Ponty Y, et al. (2013) Protein-Protein Interactions in a Crowded Environment: An Analysis via Cross-Docking Simulations and Evolutionary Information. PLoS Comput Biol 9: e1003369. doi: 10.1371/journal.pcbi.1003369.
9. Szilagyi A, Zhang Y (2014) Template-based structure modeling of protein–protein interactions. Curr Opin Struct Biol 24: 10–23. doi: 10.1016/j.sbi.2013.11.005.
10. Warren GL, Andrews CW, Capelli A-M, Clarke B, LaLonde J, et al. (2006) A Critical Assessment of Docking Programs and Scoring Functions. J Med Chem 49: 5912–5931. doi: 10.1021/jm050362n.
11. Polgár T, Magyar C, Simon I, Keserü GM (2007) Impact of Ligand Protonation on Virtual Screening against β-Secretase (BACE1). J Chem Inf Model 47: 2366–2373. doi: 10.1021/ci700223p.
12. Bas DC, Rogers DM, Jensen JH (2008) Very fast prediction and rationalization of pKa values for protein–ligand complexes. Proteins 73: 765–783.
13. Kalliokoski T, Salo HS, Lahtela-Kakkonen M, Poso A (2009) The Effect of Ligand-Based Tautomer and Protomer Prediction on Structure-Based Virtual Screening. J Chem Inf Model 49: 2742–2748. doi: 10.1021/ci900364w.

14. Ten Brink T, Exner TE (2009) Influence of Protonation, Tautomeric, and Stereoisomeric States on Protein−Ligand Docking Results. J Chem Inf Model 49: 1535–1546. doi: 10.1021/ci800420z.

15. Park M-S, Gao C, Stern HA (2011) Estimating binding affinities by docking/scoring methods using variable protonation states. Proteins 79: 304–314. doi: 10.1002/prot.22883.

16. Spassov VZ, Yan L (2013) pH-selective mutagenesis of protein–protein interfaces: In silico design of therapeutic antibodies with prolonged half-life. Proteins 81: 704–714. doi: 10.1002/prot.24230.

17. De Vries SJ, van Dijk M, Bonvin AMJJ (2010) The HADDOCK web server for data-driven biomolecular docking. Nat Protocols 5: 883–897. doi: 10.1038/nprot.2010.32.

18. Vriend G (1990) WHAT IF: A molecular modeling and drug design program. J Mol Graph 8: 52–56. doi: 10.1016/0263-7855(90)80070-V.

19. Georgescu RE, Alexov EG, Gunner MR (2002) Combining Conformational Flexibility and Continuum Electrostatics for Calculating pKas in Proteins. Biophys J 83: 1731–1748.

20. Nielsen JE, Gunner MR, García-Moreno E B (2011) The pKa Cooperative: A collaborative effort to advance structure-based calculations of pKa values and electrostatic effects in proteins. Proteins 79: 3249–3259. doi: 10.1002/prot.23194.

21. Dong F, Olsen B, Baker NA (2008) Computational Methods for Biomolecular Electrostatics. Methods Cell Biol 84: 843–870. doi: 10.1016/S0091-679X(07)84026-X.

22. Castañeda CA, Fitch CA, Majumdar A, Khangulov V, Schlessman JL, et al. (2009) Molecular determinants of the pKa values of Asp and Glu residues in staphylococcal nuclease. Proteins 77: 570–588.

23. Karp DA, Stahley MR, García-Moreno E B (2010) Conformational Consequences of Ionization of Lys, Asp, and Glu Buried at Position 66 in Staphylococcal Nuclease. Biochemistry 49: 4138–4146. doi: 10.1021/bi902114m.

24. Isom DG, Castañeda CA, Cannon BR, García-Moreno E B (2011) Large shifts in pKa values of lysine residues buried inside a protein. Proc Natl Acad Sci USA 108: 5260–5265. doi: 10.1073/pnas.1010750108.

25. Kilambi KP, Gray JJ (2012) Rapid calculation of protein pKa values using Rosetta. Biophys J 103: 587–595. doi: 10.1016/j.bpj.2012.06.044.

26. Leaver-Fay A, Tyka M, Lewis S, Lange O, Thompson J, et al. (2011) ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. Meth Enz 487: 545–574.

27. Gray JJ, Moughon S, Wang C, Schueler-Furman O, Kuhlman B, et al. (2003) Protein-Protein Docking with Simultaneous Optimization of Rigid-body Displacement and Side-chain Conformations. J Mol Biol 331: 281–299. doi: 10.1016/S0022-2836(03)00670-3.

28. Chaudhury S, Berrondo M, Weitzner BD, Muthu P, Bergman H, et al. (2011) Benchmarking and Analysis of Protein Docking Performance in Rosetta v3.2. PLoS ONE 6: e22477. doi: 10.1371/journal.pone.0022477.

29. Hwang H, Vreven T, Janin J, Weng Z (2010) Protein–protein docking benchmark version 4.0. Proteins 78: 3111–3114. doi: 10.1002/prot.22830.

30. Martin WL, West Jr AP, Gan L, Bjorkman PJ (2001) Crystal Structure at 2.8 Å of an FcRn/Heterodimeric Fc Complex: Mechanism of pH-Dependent Binding. Molecular Cell 7: 867–877. doi: 10.1016/S1097-2765(01)00230-1.

31. Dunbrack RL, Cohen FE (1997) Bayesian statistical analysis of protein side-chain rotamer preferences. Protein Science 6: 1661–1681.

32. Pollet A, Sansen S, Raedschelders G, Gebruers K, Rabijns A, et al. (2009) Identification of structural determinants for inhibition strength and specificity of wheat xylanase inhibitors TAXI-IA and TAXI-IIA. FEBS J 276: 3916–3927. doi: 10.1111/j.1742-4658.2009.07105.x.

33. Méndez R, Leplae R, Lensink MF, Wodak SJ (2005) Assessment of CAPRI predictions in rounds 3–5 shows progress in docking procedures. Proteins 60: 150–169. doi: 10.1002/prot.20551.

34. Conway P, Tyka MD, DiMaio F, Konerding DE, Baker D (2014) Relaxation of backbone bond geometry improves protein energy landscape modeling. Protein Science 23: 47–55. doi: 10.1002/pro.2389.

35. Fierens K, Gils A, Sansen S, Brijs K, Courtin CM, et al. (2005) His374 of wheat endoxylanase inhibitor TAXI-I stabilizes complex formation with glycoside hydrolase family 11 endoxylanases. FEBS J 272: 5872–5882. doi: 10.1111/j.1742-4658.2005.04987.x.

36. Efron B (1979) Bootstrap Methods: Another Look at the Jackknife. Ann Statist 7: 1–26. doi: 10.1214/aos/1176344552.

37. Xu D, Tsai CJ, Nussinov R (1997) Hydrogen bonds and salt bridges across protein-protein interfaces. Protein Eng 10: 999–1012. doi: 10.1093/protein/10.9.999.

38. Janin J, Chothia C (1990) The structure of protein-protein recognition sites. J Biol Chem 265: 16027–16030.

39. Sundquist WI, Schubert HL, Kelly BN, Hill GC, Holton JM, et al. (2004) Ubiquitin Recognition by the Human TSG101 Protein. Molecular Cell 13: 783–789. doi: 10.1016/S1097-2765(04)00129-7.

40. Xu HE, Lambert MH, Montana VG, Plunket KD, Moore LB, et al. (2001) Structural determinants of ligand binding selectivity between the peroxisome proliferator-activated receptors. Proc Natl Acad Sci USA 98: 13919–13924. doi: 10.1073/pnas.241410198.

41. Bourne Y, Watson MH, Hickey MJ, Holmes W, Rocque W, et al. (1996) Crystal Structure and Mutational Analysis of the Human CDK2 Kinase Complex with Cell Cycle–Regulatory Protein CksHs1. Cell 84: 863–874. doi: 10.1016/S0092-8674(00)81065-X.

42. Bonvin AM (2006) Flexible protein–protein docking. Curr Opin Struct Biol 16: 194–200. doi: 10.1016/j.sbi.2006.02.002.

43. Chaudhury S, Gray JJ (2008) Conformer Selection and Induced Fit in Flexible Backbone Protein-Protein Docking Using Computational and NMR Ensembles. J Mol Biol 381: 1068–1087. doi: 10.1016/j.jmb.2008.05.042.

44. Bradley P, Misura KMS, Baker D (2005) Toward High-Resolution de Novo Structure Prediction for Small Proteins. Science 309: 1868–1871.

45. Misura KMS, Baker D (2005) Progress and challenges in high-resolution refinement of protein structure models. Proteins 59: 15–29. doi: 10.1002/prot.20376.

46. Frigerio F, Coda A, Pugliese L, Lionetti C, Menegatti E, et al. (1992) Crystal and molecular structure of the bovine α-chymotrypsin-eglin c complex at 2.0 Å resolution. J Mol Biol 225: 107–123. doi: 10.1016/0022-2836(92)91029-O.

47. Braden BC, Souchon H, Eiselé J-L, Bentley GA, Bhat TN, et al. (1994) Three-dimensional structures of the free and the antigen-complexed Fab from monoclonal anti-lysozyme antibody D44.1. J Mol Biol 243: 767–781. doi: 10.1016/0022-2836(94)90046-9.

48. Kastritis PL, Bonvin AMJJ (2010) Are Scoring Functions in Protein−Protein Docking Ready To Predict Interactomes? Clues from a Novel Binding Affinity Benchmark. J Proteome Res 9: 2216–2225. doi: 10.1021/pr9009854.

49. Moal IH, Agius R, Bates PA (2011) Protein–protein binding affinity prediction on a diverse set of structures. Bioinformatics 27: 3002–3009. doi: 10.1093/bioinformatics/btr513.

50. Kilambi KP, Pacella MS, Xu J, Labonte JW, Porter JR, et al. (2013) Extending RosettaDock with water, sugar, and pH for prediction of complex structures and affinities for CAPRI rounds 20–27. Proteins 81: 2201–2209. doi: 10.1002/prot.24425.

51. Ghetie V, Ward ES (2000) Multiple Roles for the Major Histocompatibility Complex Class I– Related Receptor FcRn. Annu Rev Immunol 18: 739–766. doi: 10.1146/annurev.immunol.18.1.739.

52. Raghavan M, Bonagura VR, Morrison SL, Bjorkman PJ (1995) Analysis of the pH Dependence of the Neonatal Fc Receptor/Immunoglobulin G Interaction Using Antibody and Receptor Variants. Biochemistry 34: 14649–14657. doi: 10.1021/bi00045a005.

53. Zhang Z, Lange OF (2013) Replica Exchange Improves Sampling in Low-Resolution Docking Stage of RosettaDock. PLoS ONE 8: e72096. doi: 10.1371/journal.pone.0072096.

54. Kaufmann KW, Lemmon GH, DeLuca SL, Sheehan JH, Meiler J (2010) Practically Useful: What the Rosetta Protein Modeling Suite Can Do for You. Biochemistry 49: 2987–2998. doi: 10.1021/bi902153g.

55. Strauch E-M, Fleishman SJ, Baker D (2013) Computational design of a pH-sensitive IgG binding protein. Proc Natl Acad Sci USA: 201313605. doi: 10.1073/pnas.1313605111.

56. Janin J (2013) The targets of CAPRI rounds 20–27. Proteins 81: 2075–2081. doi: 10.1002/prot.24375.

57. Anderson DE, Becktel WJ, Dahlquist FW (1990) pH-Induced denaturation of proteins: a single salt bridge contributes 3–5 kcal/mol to the free energy of folding of T4 lysozyme. Biochemistry 29: 2403–2408. doi: 10.1021/bi00461a025.

58. Mosca R, Céol A, Aloy P (2013) Interactome3D: adding structural details to protein networks. Nat Meth 10: 47–53. doi: 10.1038/nmeth.2289.

59. Vakser IA (2013) Low-resolution structural modeling of protein interactome. Curr Opin Struct Biol 23: 198–205. doi: 10.1016/j.sbi.2012.12.003.

60. Zhang QC, Petrey D, Deng L, Qiang L, Shi Y, et al. (2012) Structure-based prediction of protein-protein interactions on a genome-wide scale. Nature 490: 556–560. doi: 10.1038/nature11503.

61. Casey JR, Grinstein S, Orlowski J (2010) Sensors and regulators of intracellular pH. Nat Rev Mol Cell Biol 11: 50–61. doi: 10.1038/nrm2820.

62. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, et al. (2000) The Protein Data Bank. Nucleic Acids Res 28: 235–242. doi: 10.1093/nar/28.1.235.

63. Mintseris J, Wiehe K, Pierce B, Anderson R, Chen R, et al. (2005) Protein–protein docking benchmark 2.0: An update. Proteins 60: 214–216. doi: 10.1002/prot.20560.

64. Moutevelis E, Warwicker J (2004) Prediction of pKa and redox properties in the thioredoxin superfamily. Protein Science 13: 2744–2752. doi: 10.1110/ps.04804504.

65. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, et al. (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 4: 187–217. doi: 10.1002/jcc.540040211.

66. Lazaridis T, Karplus M (1999) Effective energy function for proteins in solution. Proteins 35: 133–152.

67. Kortemme T, Morozov AV, Baker D (2003) An Orientation-dependent Hydrogen Bonding Potential Improves Prediction of Specificity and Structure for Proteins and Protein-Protein Complexes. J Mol Biol 326: 1239–1259. doi: 10.1016/S0022-2836(03)00021-4.

68. Simons KT, Ruczinski I, Kooperberg C, Fox BA, Bystroff C, et al. (1999) Improved recognition of native-like protein structures using a combination of sequence-dependent and sequence-independent features of proteins. Proteins 34: 82–95.

69. Chaudhury S, Lyskov S, Gray JJ (2010) PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. Bioinformatics 26: 689–691. doi: 10.1093/bioinformatics/btq007.