JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

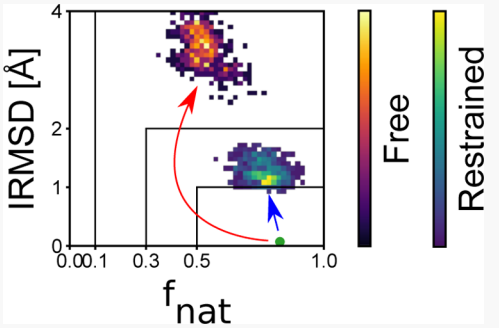
Challenges and Advantages of Accounting for Backbone Flexibility in Prediction of Protein−Protein Complexes

Nabil F. Faruk, Xiangda Peng, Karl F. Freed, Benoît Roux,* and Tobin R. Sosnick*

Cite This: *J. Chem. Theory Comput.* 2022, 18, 2016−2032 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Predicting protein binding is a core problem of computational biophysics. That this objective can be partly achieved with some amount of success using docking algorithms based on rigid protein models is remarkable, although going further requires allowing for protein flexibility. However, accurately capturing the conformational changes upon binding remains an enduring challenge for docking algorithms. Here, we adapt our *Upside* folding model, where side chains are represented as multi-position beads, to explore how flexibility may impact predictions of protein−protein complexes. Specifically, the *Upside* model is used to investigate where backbone flexibility helps, which types of interactions are important, and what is the impact of coarse graining. These efforts also shed light on the relative challenges posed by folding and docking. After training the *Upside* energy function for docking, the model is competitive with the established all-atom methods. However, allowing for backbone flexibility during docking is generally detrimental, as the presence of comparatively minor (3−5 Å) deviations relative to the docked structure has a large negative effect on performance. While this issue appears to be inherent to current forcefield-guided flexible docking methods, systems involving the co-folding of flexible loops such as antibody−antigen complexes represent an interesting exception. In this case, binding is improved when backbone flexibility is allowed using the *Upside* model.



Downloaded via UNIV OF CHICAGO on June 15, 2022 at 19:35:15 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

INTRODUCTION

Because of the central role protein−protein interactions play in many biological processes ranging from cell signaling pathways to antigen recognition, the characterization and prediction of these interactions remain an important challenge for computational biophysics. In this paper, we focus on the conformational aspect of protein−protein interactions, including tools and limitations for flexibly docking proteins and the scoring of docked poses.

Over the past 30 years, the two general approaches to protein docking have been template-based, which tends to be the most successful if homologous complexes can be identified, and free docking, often using fast Fourier transform grid representations or basis function expansions to accelerate the generation of docking poses.[1,2] The top "human" performer in the joint round 46 critical assessment of predicted interactions (CAPRI) and the 13th critical assessment of protein structure prediction (CASP) experiment for protein docking used a hybrid pipeline based on the quality of templates but noted that manual intervention using prior knowledge of interface residues during modeling and scoring plays an important part in their success.[3,4] The incorporation of evolutionary information also contributed to the recent improved performance in CAPRI.[5,6]

In a comparison of free docking algorithms, the ones that incorporated protein flexibility were the best performers on the Docking Benchmark Version 5 (BM5).[7] Two methods for including protein flexibility are normal mode deformations (e.g., SwarmDock) and molecular dynamics (MD) refinement (e.g., HADDOCK).[7−9] The HADDOCK approach also made use of bioinformatic predictions of interface residues and antibody loops to guide docking.[7]

Two recent forcefield-guided (pseudo-)dynamic protein docking methods also warrant mention. At one end of the scale of molecular details and computational resources is the all-atom explicit solvent replica-exchange MD approach of Pan et al. run on the Anton supercomputer.[10] At the other end is the CABS CG model, consisting of Cα, Cβ, united side chain atom placed at the side chain center of mass, and the peptide bond center, with a knowledge-based statistical potential that drives replica-exchange Monte Carlo pseudo-dynamics.[11,12]

AlphaFold 2[13] and RoseTTAFold[14] are recent neural network AI approaches that combine evolutionary and structural features for protein structure prediction, with abilities to predict complexes. AlphaFold 2 achieved leading performance by a wide margin in the latest CASP14

Received: December 13, 2021
Published: February 25, 2022

ACS Publications © 2022 American Chemical Society

https://doi.org/10.1021/acs.jctc.1c01255
J. Chem. Theory Comput. 2022, 18, 2016−2032

experiment, while RoseTTAFold was developed afterward and comes in second place in a postevaluation of CASP14 targets. The authors of RoseTTAFold report some surprising success in predicting the structure of complexes despite its training on single protein chains, with backbone flexibility in the docking intrinsically built into the method due to its construction for protein folding.

The scientific community is exploring protein complex structure prediction with modified AlphaFold 2 protocols upon its open source release with great success. One study using extended multiple sequence alignments obtained an accuracy of up to ∼60% according to certain metrics in a test where traditional docking methods achieved only 22% accuracy.[15] Another study by the AlphaFold team with an AlphaFold model trained on multimers achieved 67% accuracy in predicting at least acceptable-quality heteromeric interfaces.[16]

As a result of these advances, traditional docking approaches, including template-based and free docking with statistical or physical potentials, may soon become obsolete for the sole purpose of structure prediction. However, the molecular determinants responsible for a specific physical outcome are not easily extracted from deep neural networks.[17] Furthermore, RoseTTAFold and AlphaFold 2 rely heavily on coevolutionary information, which obscures a clear interpretation in terms of physical interactions. Therefore, physics-based methods to docking are still important to understand the physical principles behind protein−protein interactions, such as the balance of energies required to make accurate structure prediction. It is particularly important to improve MD approaches for studying the thermodynamics and kinetics of protein association and the associated conformational changes.

The CAPRI experiment[18] has encouraged the development of docking algorithms through a blind prediction challenge open to the scientific community, with 47 rounds held since its inception. Reflections on recent rounds by CAPRI organizers and participants have highlighted the continued difficulty of accounting for conformational changes during docking.[3,5,18,19] Round 46 was run jointly between the CAPRI and the CASP experiment and emphasized that considerable challenges remain. In this CASP-CAPRI round, only sequence information was provided, whereas in regular CAPRI rounds the unbound structures of the subunits were available. The hard targets that were poorly predicted did not have high quality templates for homology modeling and so predictors often had only subunits with a large root-mean-squared deviation (RMSD) to their native bound states in their docking pipeline. It is thus attractive to consider leveraging a method with protein folding capability to capture the flexibility required to obtain the bound conformations in such situations.

Here, we also examine the recent information-driven antibody−antigen (Ab−Ag) docking study by the HADDOCK developers.[20] They compared four different docking algorithms on Ab−Ag complexes from the BM5[7] with different levels of information about the antibody hypervariable loops, also known as complementarity-determining regions (CDRs), and epitope to bias the search. Their own algorithm, HADDOCK, performs the best in part due to the information being incorporated as a restraining potential during the search as opposed to a simple filtering mechanism as used in the other algorithms and due to their flexible refinement procedure. However, they have mixed performance for predicting the conformation of loop H3, the most variable antibody loop.

We recently developed the coarse-grained (CG) *Upside* model for protein folding simulations and now consider its suitability for docking prediction.[21,22] *Upside* is a physics-based MD algorithm that can fold proteins 10^3 to 10^4-fold faster than all-atom methods with comparable accuracy. *Upside*'s speed arises from explicitly accounting only for the backbone N, Cα, and C atoms during the dynamics portion, while during force calculation it infers the position of amide hydrogens, carbonyl oxygens, and Cβ atoms and places the multiposition united atom beads to represent the side chains. Their optimal positions are obtained via a global free energy calculation of the possible rotameric combinations and is calculated between each dynamics step. This calculation helps avoid side chain friction and kinetic locking, which results in a smoother energy surface as compared to all-atom representations that can be slowed by these effects.[21]

Upside can be slotted into a multitude of docking pipelines to make use of the extra information, as mentioned earlier in the context of other docking approaches, and below we discuss the case of information driven Ab−Ag docking. However, the primary goal of this paper is to assess the *Upside* model's suitability for predicting protein−protein interactions taking advantage of its backbone folding capability and rapid side chain sampling. To transition from folding, we extend *Upside* to protein docking by training new binding-specific energy terms with a maximum likelihood approach using the BM5 set of non-redundant complexes.[7]

Our updated model is compared to other docking algorithms using a subset of the benchmark set according to the widely used CAPRI experiment criteria.[18] We identify the penalty for coarse graining side chains and examine the impact of flexibility, including for other forcefield-guided dynamics methods. We then assess flexibility and the performance enhancements of extra information in the case of antibody docking, following the reasonable assumption that *Upside* may perform relatively well for this class of complexes due to the inclusion of backbone dynamics enabling co-folding of the flexible CDR loops during binding. Finally, we conclude with a broader discussion of the different challenges faced by protein docking versus protein folding, and the merit of MD approaches in light of the recent performance of neural network methods.

METHODS

Data Sets for Training and Testing. The BM5 provided 230 nonredundant binary complexes, of which 175 from the previous version were used for training and the 55 new complexes were used for testing.[7] The benchmark set spans a diverse set of enzyme containing Ab−Ag and other types of complexes. The set also contains both bound and unbound forms of the subunits at high sequence identity; the unbound forms are required to compare against other docking algorithms according to the CAPRI methodology. FRODOCK v3,[2] a rigid body docking algorithm based on spherical harmonics, was used to generate 1000 decoys per complex based on the bound conformation of the subunits. FRODOCK has various energy terms that can be used for decoy generation and ranking. Here, we used the defaults for the van der Waals, electrostatics, and SOAP all-atom statistical potential[23] but omitted the desolvation term for efficiency because it would require expensive computation of electrostatic potential maps.

In addition, Ab−Ag complexes were considered to evaluate the impact of backbone flexibility, focusing on the flexible

https://doi.org/10.1021/acs.jctc.1c01255
J. Chem. Theory Comput. 2022, 18, 2016−2032

antibody hypervariable CDR loops that impart specificity. Ambrosetti et al. recently tested four docking algorithms for Ab−Ag docking using various levels of external information about the CDR loops and epitope to bias the results.[20] Their evaluation set was the 16 new Ab−Ag complexes added to the BM5, and so as to enable comparison we also used these complexes. Although these complexes are already included in the general "diverse" set, we docked them anew for two cases of extra information to take advantage of prior knowledge of the CDR loops and how a rough estimate of the epitope can be sourced from the experiment and other predictive tools. Whereas Ambrosetti et al. used the exact Ab loop residues and defined their coarse epitope by those residues within 9 Å of the loops in the native bound structures, for simplicity we defined the loops as residues involved in native interface contacts (Cα distances < 10 Å) plus a zone of up to three residues on either side and same for the coarse epitope. While our definitions and use of the loop and epitope information are different from Ambrosetti et al., we think they are sufficiently similar to allow for a meaningful comparison between the two methods.

For the modeling using only Ab loop information, decoys were again generated with FRODOCK, except beginning with up to 20,000 decoys and filtering down to 1000 around the interface by requiring a minimal number of loop contacts. For the Ab loop information plus coarse epitope information case, 200,000 decoys were first filtered down with the frodockconstraints program in the FRODOCK suite to keep the two furthest residues of the loops within 55 Å of the two furthest residues on the coarse epitope. These decoys were further filtered down to 1000 or less by selecting only those that had contacts between at least one-third of the residues of the loops and epitope.

New Energy Terms for Docking. Side chain−side chain (SC−SC) interactions and burial (desolvation) provide important contributions to protein−protein interactions and for this reason deserve special attention. *Upside*'s basic potential represents side chains by a single, directional bead that may be in up to six different states (positions and orientations) to mimic the diversity in the side chain rotamers.[21] The interaction between beads is given by a pairwise potential composed of radial and angular terms using cubic splines that offer flexibility in the form of the potential. This 2-body SC−SC interaction potential is used to determine the SC state probabilities, and in conjunction with intrinsic (1-body) individual χ_1 rotamer probabilities, give rise to side chain free energies. These free energies are solved in a self-consistent iterative procedure during each MD step using belief propagation (method of inference on graphical models). Side chain−backbone hydrogen bonding and side chain−backbone main atom interactions are combined with the 1-body rotamer probabilities during the free energy solution. The forces obtained from these free energies are then back propagated onto the backbone atoms.

In the present treatment, additional corrective terms to the basic *Upside* potential were introduced. An interprotein SC−SC term, $V_{\text{inter\_rot}}$, copying the functional form of the original SC−SC term was added but it acts only between SC beads on different proteins. This was done such that the additional term would not affect the internal folding behavior of either proteins. The SC−BB components are excluded in this new term for simplicity. The cutoff was extended from the 7 Å of the base rotameric term to 10.5 Å to better account for possible long-range interactions of electrostatic residues that are more prevalent at protein interfaces.[24]

Upside also has a many-body environment term to capture the effects of burial and desolvation. With this term, the number of side chain beads are counted within a hemisphere above the C_β, weighted by their residue types. This count is then coupled to a residue-specific energy composed of cubic splines. A new interprotein environment term, $V_{\text{inter\_env}}$, is added, again copying most of the functional form of the original implementation. However, this new interfacial term requires at least one bead from the opposite protein within the hemisphere for its activation.

The new potential is then given by $V = V_{\text{orig}} + V_{\text{inter\_rot}}(r,\theta_1,\theta_2) + V_{\text{inter\_env}}(N;w)$, where r is the distance between beads, θ_1,θ_2 are the angles between the bead orientation vectors and the displacement vector between the beads, and N is the bead count weighted the residue type weights w. We use the *Upside* folding forcefield v1.5 for our V_{orig}, which was developed with more diverse training ensembles and longer training cycles for better results than the first *Upside* version.[25]

Training of the Potential. To train and optimize the potential for protein docking, we initially used our original contrastive divergence machine learning methodology that we previously developed for protein folding studies.[22] This approach was based on populations (free energies) and minimizing the difference between the approximate distribution of states generated by *Upside* and the "true" distribution of crystal structures found through experiments. However, it proved challenging to achieve the thermodynamic sampling required for this method to correct for a myriad possible misbound poses.

Therefore, we used a simpler objective of minimizing the native poses' potential energies compared to the decoy poses and therefore maximize their Boltzmann probability. In this new strategy, we consider the average potential energy after short simulations for a set of i, ..., N protein complexes starting in k, ..., m poses

$$E_k^i, \qquad \begin{array}{l} k = 0 \text{ is native} \\ k \neq 0 \text{ is decoy} \end{array}$$

We desire to maximize the Boltzman weight (population fraction) of the correct docking poses for the training set by minimizing the negative logarithm of the fraction

$$\max_\alpha \frac{1}{N}\sum_i^N \left\{ \frac{e^{-\beta E_0^i}}{\sum_{k=0}^m e^{-\beta E_k^i}} \right\}$$

$$\Rightarrow \min_\alpha F = \frac{1}{N}\sum_i^N -\ln\left(\frac{e^{-\beta E_0^i}}{\sum_{k=0}^m e^{-\beta E_k^i}} \right)$$

where α denotes the energy term parameters. Rearranging and adding a term for regularization gives the objective function

$$F = \frac{1}{N}\sum_i^N \left(E_0^i + 1/\beta \ln \sum_{k=0}^m e^{-\beta E_k^i} \right) + \frac{\lambda}{2}\alpha_2^2$$

Taking the gradient of this expression with respect to the parameters α yields

https://doi.org/10.1021/acs.jctc.1c01255
J. Chem. Theory Comput. 2022, 18, 2016−2032

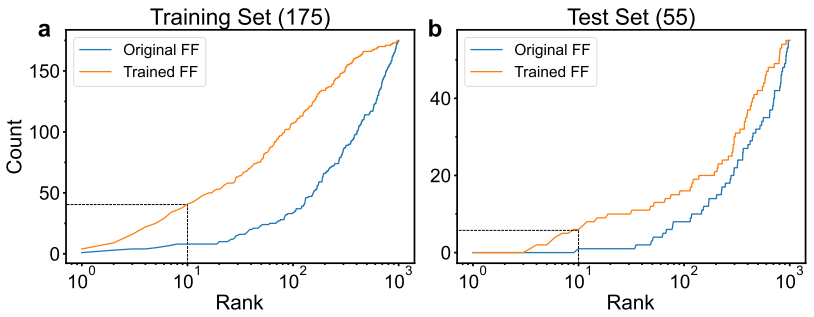


Figure 1. Cumulative counts of native pose rank for the new forcefield trained for docking compared to original forcefield trained for folding. The plots are read as the number of complexes in the given set that has the native pose at a certain rank or better. (a) Training set performance, 175 complexes total. (b) Test set performance, 55 complexes total.

$$\frac{\partial F}{\partial \alpha_l} \cong \frac{1}{N_{c>5}} \sum_{i,c>5}^{N_{c>5}} \left(\frac{\partial E^i_{0,j}}{\partial \alpha_l}_{\,j} - \sum_{k=0}^{m} \frac{\partial E^i_{k,j}}{\partial \alpha_l}_{\,j} \frac{e^{-\beta E^i_{k,j}/s}}{\sum_{k=0}^{m} e^{-\beta E^i_{k,j}/s}} \right) + \lambda \alpha_l$$

where the derivatives are averaged over j frames, s is a temperature scale factor for numerical stability (s = 100 in practice), and $c > 5$ is a condition to exclude well-performing complexes. Such complexes were excluded from contributing to the parameter update if their native pose was in the top five of all poses because the information content of this pair largely had been extracted. The parameters are then updated for the subsequent training cycle according to

$$\alpha_{t+1} = \alpha_t - r\frac{\partial F}{\partial \alpha_t}$$

where t denotes the current cycle. In practice, the training set is divided into five minibatches that are cycled, and each training cycle involves 500 *Upside* time units ≈5−50 ns of simulation to relax the poses of each complex up to 1000 residues. The frame output interval is 2 *Upside* time units. For larger complexes, the simulation time is scaled down by the number of residues according to $t'_{\mathrm{sim}} = t_{\mathrm{sim}}/t_{\mathrm{int}} \times \frac{1}{25} t_{\mathrm{int}} \left(\frac{1000}{n_{\mathrm{res}}}\right)^{3/2}$ for performance reasons.

Cubic spline parameters for the new protein−protein terms are initialized to low values, $\alpha < 1$. We trained on the top 100 decoys of the bound subunit forms ranked according to *Upside* energy from an initial relaxation run. We did not find much benefit in conducting multiple training rounds, where decoys were reordered according to their new energies and a new top 100 selected for training.

Testing and Evaluation Using the Optimized Potential. We ran with the two cases of restraints applied to the backbones, a semirigid case and a fully flexible case. In the semirigid case, Cα atoms were kept within ∼3 Å of their initial positions with spherical flat-bottom quadratic potentials. The flexible case involves no restraints. Run duration was relatively short, 1000 *Upside* time units for $n_{\mathrm{res}} \leq 1000$ and scaled down for larger complexes. The first half of each trajectory was discarded as equilibration, and centroid structures and their respective potential energies were selected as representatives for CAPRI criteria evaluation and ranking.

We also checked whether there is a performance benefit with full side chains. The side chains of the *Upside* structures from the semirigid restraint case are rebuilt using the SCWRL4 algorithm, which minimizes the energy from an atomic interaction model in conjunction with observed backbone-dependent rotamer frequencies to find the most likely rotamer states.[26] With the full side chains rebuilt, the poses are rescored with SOAP-PP,[23] an atomistic statistical potential used in the consensus scoring method of a former top CAPRI experiment group's docking server[6] and as a component in the scoring model of FRODOCK v3.[2]

For the Ab−Ag antibody set, we examined docking with both fully flexible and semirigid loops. For the flexible case, antibody residues involved in the native interface along with up to five residues on either side were kept flexible, in essence allowing the CDR loop residues to remain flexible, while the rest of the Ab fold was restrained with flat-bottom potentials. The semirigid case used the same restraints as the semirigid case in the full diverse set (see above). Antigens were held within ∼2 Å C_α-RMSD with harmonic restraints and able to move as a rigid body up to 10 Å of their starting positions.

To mimic Ambrosetti et al.'s "Scheme 2", whereby a restraining potential between the CDR loops and a coarsely defined epitope is used to bias the results, we apply C_β sigmoid contact potentials of $-0.5 k_B T$ *Upside* energy units between all combinations of antibody interface residues and epitope residues. A note about *Upside* energy units is warranted: the correspondence to physical temperature is not well established for the new training, so we simply provide thermal energies in units of "$k_B T$". Unlike the Ambiguous Information Restraints used by the HADDOCK algorithm in Ambrosetti et al. that offers a level of smoothness and uniformity, our approach is simply pairwise additive.

Trajectories for each decoy pose are were clustered according to interfacial root mean squared deviation (IRMSD) for up to three clusters of frames within 1 Å IRMSD of the minimum energy structure of each cluster. The minimum energy structures of all clusters for all poses were sorted according to their energies and the top 100 are selected for further CAPRI Criteria evaluation.

RESULTS

Training and Testing of the Potential. *Upside*'s forcefield parameters were originally trained for folding on a set of single domain proteins.[22] A preliminary analysis of protein−protein docking performance with this energy function yielded many misbound poses that were energetically favored over the native pose. In this work, we first sought to correct for these deficiencies by training new energy terms for protein docking using the BM5 set of non-redundant complexes.[7] These new terms are similar in form to existing ones but only apply at the interface. We next explored whether

https://doi.org/10.1021/acs.jctc.1c01255
J. Chem. Theory Comput. 2022, 18, 2016−2032

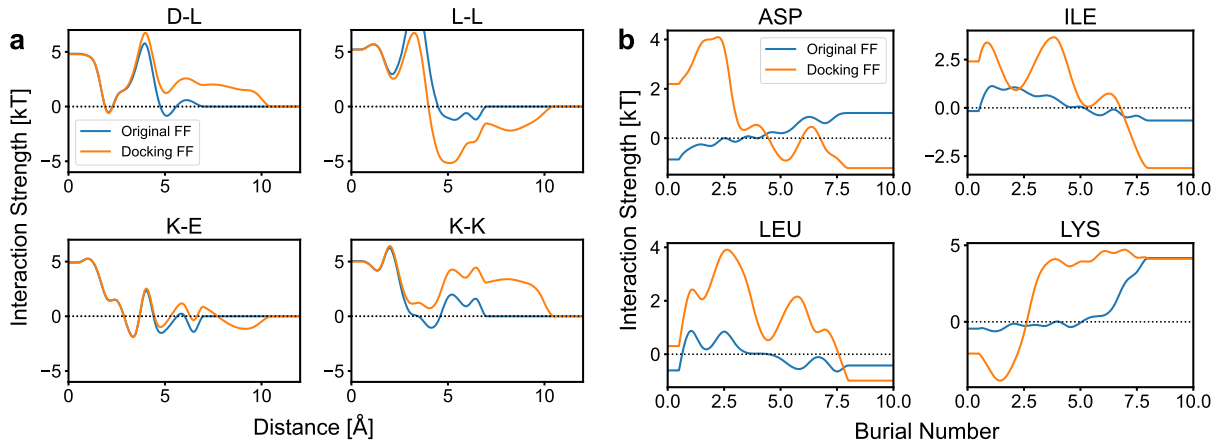


Figure 2. Changes in side chain (SC) interaction terms upon training for binding. (a) Radial part of pairwise SC−SC potentials. The "Docking FF" plots are of the original potential along with the corrective contribution of the trained protein−protein potential. (b) Many body environmental potential. The "original FF" and "docking FF" plots are of the separate contributions of the original and new docking terms.

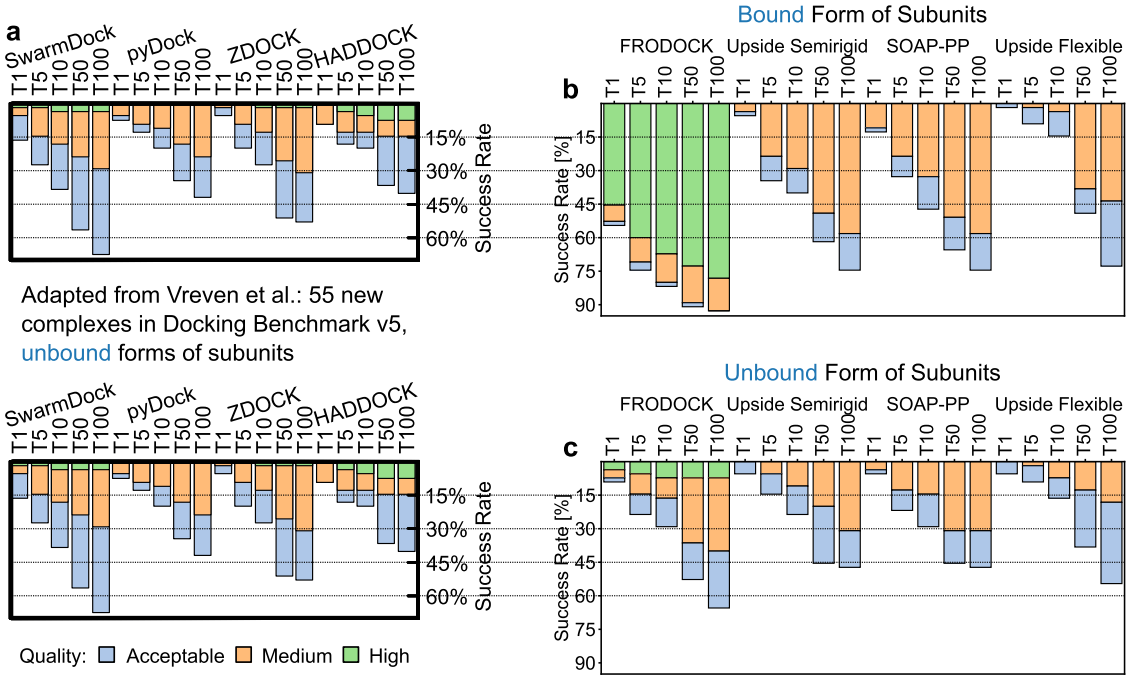


Figure 3. CAPRI criteria evaluation. (a) Performance of four docking algorithms (adapted with permission from ref 7); note that the plots are duplicated for comparison purposes with *Upside* results in (b,c). Their assessment was done on the same set of complexes as our test set. Notation: "T*X*", native pose found within the top *X* predictions. (b) *Upside* results using bound forms of the subunits. From left to right are FRODOCK: ranking of FRODOCK poses according to the FRODOCK scores; *Upside* semirigid: results using *Upside* relaxed poses using spherical flat-bottom restraints for the backbone (with added side chains) ranked according to the *Upside* energy function; SOAP-PP: results for representative semirigid *Upside* structures scored according to the SOAP-PP atomic statistical potential; *Upside* flexible: results for complexes allowing for full backbone movement. (c) *Upside* results using unbound forms of the subunits, with same types of subplots as in panel (b).

more drastic changes to the single side chain bead model were required, that is, if we were limited by our coarse graining, and, finally, explore the differences between protein folding and docking that gave rise to our differing performance on capturing the two tasks.

We followed a force field training protocol of maximizing the probability of the native pose (see the Methods section). The objective function decreased after training the new energy terms, leveling off after 15 cycles (Figure S1). This training procedure significantly improved our ability to distinguish the native from the decoy poses (Figure 1). The cumulative counts of finding the native pose at a higher rank (1 being the highest) increased compared to the decoys for all complexes after relaxing the poses with *Upside* simulations with spherical flat-bottom restraints to keep each partner in its starting conformation. For the training set (Figure 1a), the original forcefield predicted the native pose ranking in the top 10 for only about 8 of the 175 complexes (4.6%), whereas the force field trained for docking did so for about 42 complexes (24%), a substantial improvement (dashed lines in figure). The improvement was smaller for the test set, 1.8 → 12.7% (Figure 1b). We later discuss the possible role of overfitting in the section "Determining Features That Contribute to Performance".

https://doi.org/10.1021/acs.jctc.1c01255
J. Chem. Theory Comput. 2022, 18, 2016−2032

The contributing factors to the improved scoring are hinted at by the representative plots of the new residue specific potentials (Figures 2, S2a). Pairwise charge−charge interactions became more pronounced, particularly the repulsive ones (e.g., Lys−Lys), with a signal extending beyond the original 7 Å cutoff distance (Figure 2a). Surprisingly, some hydrophobic interactions (e.g., Leu−Leu) are among the most strengthened. Although the original training procedure utilized ample information for hydrophobic side chain interactions (e.g., hydrophobic cores), here at the interface, they play a different role in the balance of energies because backbone hydrogen bonding does not play as large a role in binding as it does in folding.

The environmental term potentials (Figure 2b) are more difficult to parse. This is due in part to our burial number (a measure of desolvation) being determined by a summation of the burial weights for the different residues located within the r = 7 Å hemisphere of the buried residue. The burial weights of the new potentials do not appear to have a strong correlation with the size of the neighboring residues. However, the new potential allows some charged residues to have some favorable burial, as shown for Asp at high burial numbers shown in Figure 2b, which supports the formation interfaces with charged residues. Another complicating factor for interpretation of the environmental potential is that a compensation occurs with the pairwise side chain potential.

Evaluation according to CAPRI Criteria. We investigated various protocols using different combinations of subunit starting structures (either bound or unbound conformations) and restraints (either backbone fixed or free). These options were assessed with the full CAPRI criteria where the quality level of a prediction is judged according to three metrics: IRMSD, ligand root mean squared deviation (LRMSD), fraction of native contacts (f_{nat}).[18,27] Given imperfections in our scoring, there may be other poses acceptably close to native-like that score better than the native pose. An evaluation with CAPRI criteria is more generous than just noting the cumulative native ranks because decoys that are native-like according to CARPI and score well are considered successes.

Figure 3 shows CAPRI criteria performance of our newly trained docking forcefield in comparison with other docking algorithms featured in Vreven et al.[7] Figure 3a is adapted from ref 7 in order to compare the benefit of using the restrained bound structure ("semirigid") as compared to the unrestrained situation ("flexible"). The vertical bars indicate how many complexes have predicted native-like structures of a particular quality at different threshold levels of ranking/scoring, with T1 being the top prediction and T100 meaning that they are found in the top 100 complexes.

Figure 3b highlights the results from the *Upside* pipeline starting from the bound forms of the subunits from the Dockground testing set. From left to right, there are the native-like poses from the initial FRODOCK rigid body docking ranked using their FRODOCK scores compared to the other decoys. Next are the rankings of representative structures from *Upside* after a minor (<3 Å) relaxation of the FRODOCK starting poses using semirigid backbone restraints. These structures are first scored using the *Upside* energy function. Side chains are then added with SCWRL4[26] (also required for f_{nat} calculation) and rescored with the all-atom statistical potential SOAP-PP.[23] This rescoring with complete atomistic side chains examines *Upside*'s loss of accuracy due to the use of a single (albeit multi-position) bead for each side chain. Finally, there are the results for representative structures from *Upside* simulations without backbone restraints.

When starting with the bound forms of the two partners including the native side chain rotamers, the FRODOCK results have a high success rate with a majority of the targets having high quality native-like predictions (green bar). This success is due in large part to the backbones and interfacial side chains being fixed a priori in their bound conformations and rotamers, and FRODOCK can find a well-matched, interdigitated interface between the native positions of the binding partners. Once processed into *Upside*, the atomistic positions of the side chains are lost. This situation could occur with low resolution structures, for which the side chains positions are not as well determined as the backbone (e.g., in nuclear magnetic resonance spectroscopy or cryo-electron microscopy-based structure determination). Nevertheless, *Upside* still performs well on ranking native-like poses compared to the other docking algorithms of Vreven et al.[7] (Figures 2.3a) when the *Upside* backbones are kept semirigid (note however, that the predictions of the other algorithms use the unbound forms of the subunits, while *Upside* is using the bound forms. Therefore, this comparison is not completely valid; the proper test is conducted below).

The SOAP-PP results are calculated using *Upside*'s optimized structures, followed by full side chain addition by SCRWL4[26] and then scored using the SOAP-PP energy function. The results with full side chains are notably better only for predicting the native-like pose as the lowest energy structure (T1 performance: 5.4 → 12.7%). For being in the top 5+ lowest energy predictions, however, there is minimal improvement. This suggests that for most situations, there is only a very mild decrease in performance when using *Upside*'s single side chain bead at the scoring stage once the backbone is determined. However, the limitations of SCWRL4 and SOAP-PP must be considered. The χ_{1+2} accuracy of SCWRL4 was 80% on a test set of proteins when compared to the crystal positions of side chains with high electron density, and less for higher χ angles (e.g. 47% χ_3 accuracy for Arg).[26] Furthermore, in SOAP-PP's original paper, it only had 40% success in placing native-like predictions in the top 10 on a prior version of the BM5 when the subunit backbones and side chains were in their exact native positions (i.e., it had the best possible starting structures to score).[23] Thus, there is still room to benefit from a more accurate model for the side chains.

Another major finding is the drop in quality and ranking of native-like structures that occurs when the proteins are allowed to be fully flexible during the *Upside* simulations. The subunits drift away from their native bound forms, indicating that there is insufficient accuracy in the folding component of the *Upside* forcefield which cannot be compensated for by the new interprotein energy terms. We characterize this issue further in subsequent sections and investigate to what extent it may be general for all forcefield-guided dynamics methods.

Although there is movement away from the native state when starting from the bound forms, at this point we were anticipating that there would be movement towards the native state when starting with the *unbound* forms, which is the more relevant scenario that we discuss next.

The FRODOCK unbound subunit results are much worse compared to the bound form results since the tight fit at the native interface is lost due to backbone and side chain conformational differences (Figure 3c). FRODOCK still

https://doi.org/10.1021/acs.jctc.1c01255
J. Chem. Theory Comput. 2022, 18, 2016−2032

performs better than the other docking algorithms, possibly since it incorporates SOAP-PP in its scoring and may have other advancements since it is a newer algorithm.

Upside's decrease in performance due to coarse-graining of the side chains would be acceptable if it improved on the FRODOCK results of the unbound forms by compensating with its ability to sample backbone conformations and side chain rotamers to find a more native-like pose. However, *Upside* and its conformational sampling exhibit poorer performance, even under semirigid restraints (backbones held in unbound conformation), with both fewer native-like poses ranked highly and more lower quality structures. SOAP-PP is able to improve performance for some of the T1-T5 predictions for the semirigid *Upside* structures rebuilt with full side chains. Although there is still some backbone deviation, we expect the semirigid case to emphasize the role of the side chains and new interprotein energy terms. The results indicate that *Upside*'s ability to repack the CG side chain beads did not yield an overall scoring advantage over FRODOCK. However, *Upside* with semirigid restraints achieves comparable performance to many of the docking methods in ref 7, which we view as a partial success considering *Upside*'s disadvantage due to coarse-graining of the side chains.

Most importantly, our predicted structures have larger deviations from the native bound state when *Upside* is allowed full backbone flexibility, that is, even the starting poses with the subunits in their unbound conformations are overall more native-like as compared to those generated when *Upside* is allowed to optimize the backbone. In the section "Energetic Costs of Retaining Native-Like Subunits", we will return to this issue and characterize the energetic compensation required to shift the subunits from their unbound to bound backbone conformations in order to establish the magnitude of forcefield improvements needed.

The best performing methods examined by Vreven et al.[7] are SwarmDock and HADDOCK. SwarmDock performs the best in terms of the percentage of acceptable quality native-like structures or better ranked in the top 10, whereas HADDOCK has the most high-quality native-like structures. SwarmDock's success likely is partly due to its approach to backbone flexibility via normal mode deformation, allowing it to better address the more difficult targets that have a greater change between bound and unbound forms of the subunits. SwarmDock was the only successful method for the sole "easy" target, having a ΔIRMSD < 1 Å between bound and unbound forms. The authors hypothesize that this is due to SwarmDock being able to widen the narrow opening of the receptor binding site. In SwarmDock, the normal mode coefficients are updated in the search procedure in the direction of minimum energy, but the energy evaluations include only the interaction energy between the two binding partners and not their internal normal mode energy.[8] In effect, the range of allowed protein flexibility is somewhat artificial, as it is highly constrained by the selection of the normal modes with lowest frequency. With this strategy, SwarmDock is not penalized by backbone strain during search and scoring, whereas the backbone energy is an integral part of the physical forcefield that guides the dynamical motions in fully flexible models such as *Upside*. With such physical models, structural deformations are governed by the intramolecular potential energy and occurs spontaneously during the sampling, and their influence is implicitly included in scoring poses.

With HADDOCK, Vreven et al.[7] utilized bioinformatics predictions of the interfaces and knowledge of antibody CDR loops to bias the docking results to make use of HADDOCK's "ambiguous information restraints". HADDOCK's high-quality native-like structures may be a result of this extra information, in combination with HADDOCK's all-atom explicit solvent flexible refinement and an all-atom energy function. In the antibody section of this paper, we also test *Upside*'s performance with additional information to enable a more valid comparison to HADDOCK.

The latest studies for SwarmDock and HADDOCK reassessed their performance for a subset of the docking benchmark set with enhancements to their procedures (cross-docking diversification of the starting conformations for SwarmDock, use of higher levels of informational restraints for HADDOCK). These new studies show that there is still room for improvement in flexible docking for both normal mode and MD methods.[20,28]

Backbone Flexibility. To quantify the impact of backbone flexibility on docking predictions and separate it from any deficiencies in our search process, we ran a series of the best case simulations starting from the native pose (Figure 4). All complexes were run for the same time as the CAPRI evaluation runs and each point is the average of three runs calculated using the centroid structure taken from the second half of

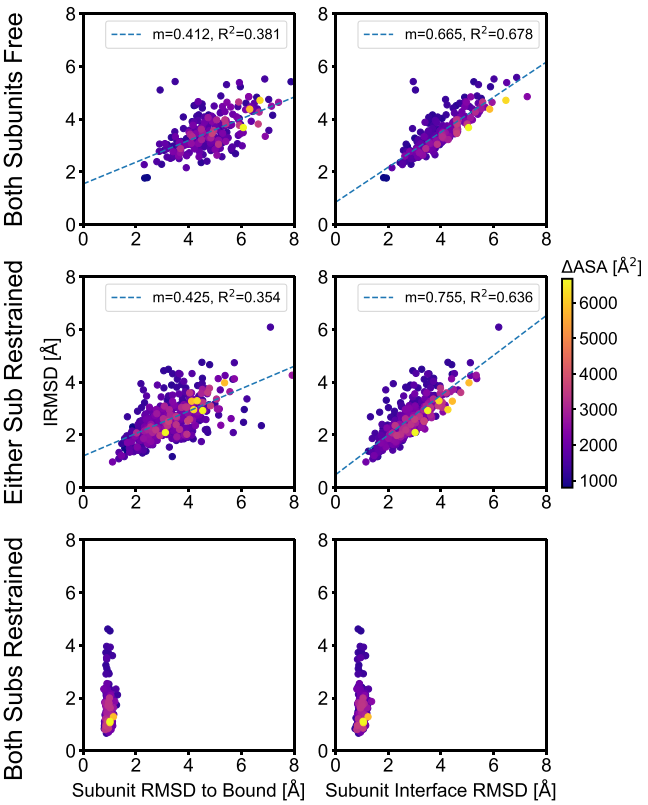


Figure 4. Impact of *Upside* backbone flexibility's on IRMSD. Left: subunit RMSD to their native bound forms. Right: RMSD of the interfaces for each individual subunit. These subunit RMSDs are the Euclidean norm of both subunits (see text). The *y*-axis is the IRMSD of the complex. Top row: no restraints applied. Middle row: harmonic restraints applied to one of the subunits to keep its backbone within ∼1 Å of the native bound state (data from both subunits are shown). Bottom row: restraints separately applied to both subunits. The points are colored according to change in accessible surface area, ΔASA.

https://doi.org/10.1021/acs.jctc.1c01255
J. Chem. Theory Comput. 2022, 18, 2016−2032

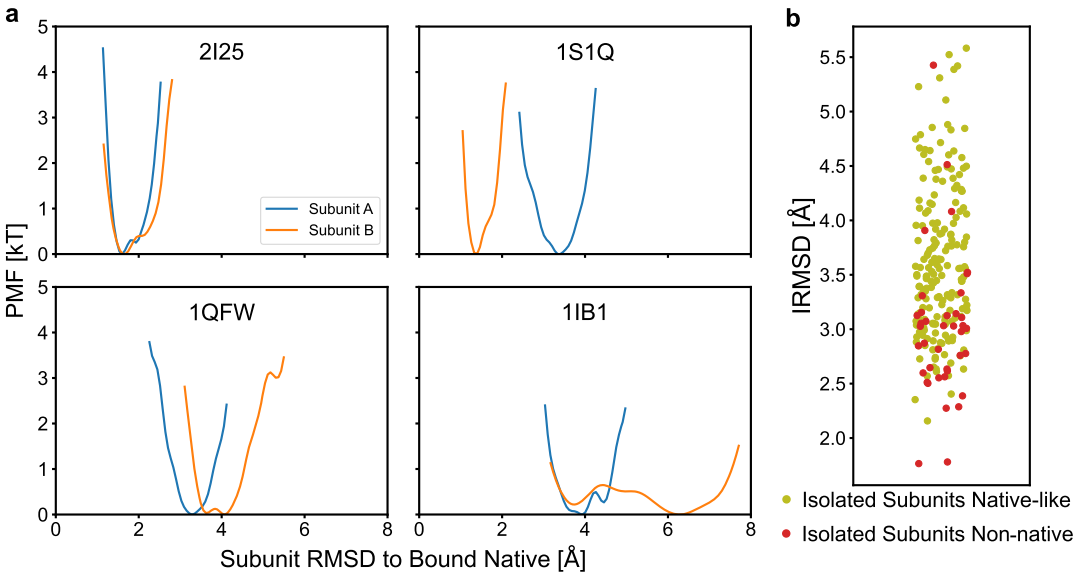


Figure 5. Energy needed to adopt the native subunit conformation. (a) PMFs from RMSD distributions referenced to the native bound state to visualize the energy required to adopt bound native-like conformations. (b) Classification plot of IRMSDs from unrestrained simulations starting from the native complex using the following threshold criterion on the RMSD distributions of the PMFs: complexes are considered to have native-like subunits when 50% of both the individual subunit RMSD distributions are below 3 Å (red dots), otherwise they are considered non-native subunits (yellow dots). The *x*-axis spread is artificial jitter to aid in the distinguishing of the points.

trajectories. The left and right columns present two different measures of backbone flexibility, either the total subunit RMSD to the native bound state (left) or the RMSD of the individual subunit interface (right). The plotted subunit RMSD is the Euclidean norm of both subunit's RMSDs ($x_2 = \sqrt{\mathrm{RMSD}_1^2 + \mathrm{RMSD}_2^2}$). There is a stronger correlation between subunit conformations and the IRMSD of the complex when considering the RMSDs at each individual subunit interface compared to whole subunit RMSDs, as expected.

When no backbone restraints are applied to the subunits (Figure 4, top row), we see that the subunit RMSDs mostly lie between 2 and 5 Å as are the IRMSD values. Conformations that are 2−5 Å RMSD from native conformation are generally considered a success for protein folding prediction, so *Upside*'s ability to maintain that RMSD for the subunit conformations could be considered laudable for a de novo coarse grained model. However, a substantial number of subunit RMSDs are above 5 Å. For those complexes, medium quality docking results are unachievable, as the requirements are IRMSD < 2 Å or LRMSD < 5 Å (given f_{nat} < 0.5). The net effect of inaccuracies for both subunits largely explains why allowing backbone flexibility often is detrimental.

The simulations are improved to 1.5−4 Å subunit RMSDs and IRMSD after applying harmonic restraints to either subunit to keep its RMSD to ∼1 Å of the original bound conformation (middle row). Furthermore, when both subunits are restrained, the IRMSD remains largely below 3 Å. This positive performance for the doubly restrained situation emphasizes *Upside*'s ability to accurately hold the native pose using the new interprotein terms (if given the native subunit conformations). For the complexes, where both subunits are restrained yet have a large IRMSD, there are significant rigid body-like translocations of the subunits from the native bound pose. For these examples, our interprotein interaction terms are inadequate. The high IRMSD points correspond to low ΔASA complexes and we further examine the impact of interfacial area in a later section.

Another depiction of the effect of subunit backbone restraints is provided by 2D heat maps (Figure S3). Here, the densities are shifted to more native-like values of IRMSD and f_{nat} for most cases with subunit backbone restraints. It is again reassuring that for most complexes the interprotein interactions are strong enough to maintain a low IRMSD.

Energetic Costs of Retaining Native-like Subunits. The previous section examined the increase in individual subunit RMSDs when the proteins are in the complexes. We now focus on the role of backbone flexibility on individual, isolated subunits to examine the magnitude of free energy required to shift the backbones into their native bound conformations, which are a necessity for obtaining good docking predictions. We again ran the subunits for the same base duration as the CAPRI evaluation runs, starting from their native bound conformations. Figure 5a shows representative plots of the potentials of mean force (PMFs) generated from Gaussian kernel density estimates of the RMSDs to the bound native state taken from the second half of the trajectories. The two lines in each plot are for each partner run separately and plotted from the lowest to the highest observed RMSDs. For subunits with RMSD ≥ 2 Å, we observe free energies of up to ∼$4k_BT$ at their lower RMSD bounds (e.g., 1QFW in Figure 5a). Conformations with a lower RMSD would have an even higher free energy cost, indicating that a substantial improvement of the folding portion of an *Upside*'s energy function would be needed to consistently obtain structures with a RMSD smaller than 2 Å.

We next examined the relationship between the individual RMSDs of the separated subunits and the IRMSDs of the unrestrained simulations of the complexes, as shown in Figure 4. The results in Figure 5b show a classification of those simulated complexes, with yellow dots corresponding to complexes with one or both subunits are non-native-like while red dots corresponding to complexes where both

https://doi.org/10.1021/acs.jctc.1c01255
J. Chem. Theory Comput. 2022, 18, 2016−2032

partners remain native-like. Complexes are considered native-like when both the partners have 50% of their RMSD distribution below 3 Å when individually simulated. Native-like subunits tend to have native-like IRMSDs, but there is large overlap between the classes. The existence of low IRMSD, but non-native-like subunits indicate that regions of the proteins away from the binding interface experience the bulk of the conformational difference. Conversely, a high IRMSD with native-like subunits is a situation where the binding partners experience rigid-body displacement, indicating a deficiency in the interprotein terms of our energy function for those complexes.

Backbone Flexibility and Other MD Methods. To determine the generality of *Upside*'s decrease in performance when backbone movement is allowed, we examined two other recent forcefield-guided methods, the CABS coarse grain model[11,12,29] and a computationally heavy all-atom explicit solvent approach.[10] Both approaches let the dynamics guide the association of the protein partners, unlike the search process used in the *Upside* docking pipeline to start with up to 1000 pre-docked poses. The docking problem can be broken into three parts, the generation of many possibly bound poses followed by their refinement and scoring. In this work, we focus on the last two steps for *Upside* as they are sufficient to address whether our model is capable of identifying the true native pose and to what degree does backbone flexibility help in this identification.

The CABS CG model is similar to *Upside* but has slightly lower level detail, for example, it includes angular dependance of SC−SC interactions but single side chain states.[11,12,29] Hence, it offers an independent insight into the potential benefits of backbone flexibility for docking. The CABS model uses replica exchange with Monte Carlo moves that capture transitions on physical timescales, which they call pseudo-dynamics. In the CABS docking study, 62 complexes are free docked, which as mentioned earlier means that the binding partners begin separated at different initial positions for each replica and associate over the course of the simulation. The partners begin in their unbound native conformations and they applied restraints individually to each partner, with strengths such that the receptor fluctuates only around 1 Å and the ligand between 2 and 12 Å.[29] This setup is overall much more flexible than the *Upside* unbound subunit semirigid case for the CAPRI evaluations, and the *Upside* case utilizes FRODOCK pre-docked starting states.

Figure 6 shows a comparison of the *Upside* unbound subunit semirigid results and the CABS results taken from Table 1 of Kurcinski et al. for the 43 complexes common to both test sets.

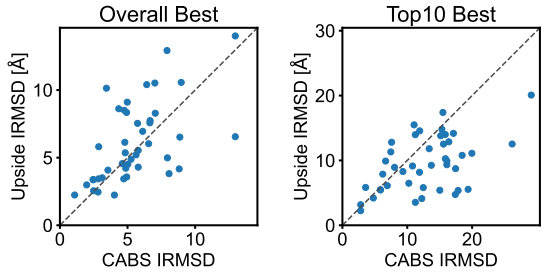


Figure 6. Comparison of *Upside* with CABS docking for complexes common to both studies. (a) Overall best IRMSD out of all poses. (b) Best IRMSD among the top 10 ranked poses.

The results are mixed when considering the lowest IRMSD observed from all poses (Figure 6a), with CABS performing better for some complexes and the *Upside* semirigid situation better for others. However, *Upside* semirigid is overall better than CABS for the lowest IRMSD in the top 10 ranked poses (Figure 6b), indicated by points below the diagonal. Thus, in the context of scoring native-like poses highly, backbone flexibility is a detriment to the CABS CG model as well when compared to the *Upside* semirigid results. "Simple" rigid-body docking, for example, using FRODOCK would have been better, considering that FRODOCK was used for the starting states of *Upside* for these targets and *Upside* tends to do worse than FRODOCK as shown previously. This test should have been ideally conducted as a "CABS semirigid" versus "CABS flexible" setup to remove influence from differences in force fields, but we think that *Upside* as a CG model is a suitable stand-in for the semirigid scenario. We also recognize that the CABS study had a focus on whether low IRMSD states could be sampled at all even if they were not among the top 10 predictions, and the authors acknowledge that scoring improvements for their model are required. However, our comparison highlights how much backbone flexibility can be a detriment when sampling and scoring are coupled.

To see whether the downside of allowing for backbone flexibility is limited to CG models and their inherent inaccuracies, we examined the MD approach of Pan et al.[10] These all-atom explicit solvent simulations use computing power of Anton 2 with enhanced sampling termed "tempered binding". They observed reversible binding to the native state for five out of six of the complexes (the sixth irreversibly bound into a native-like pose). The native state was the most populated for each, with the IRMSDs of the most stable poses for all six complexes being below 1.3 Å. *Upside* does not have the same complexes as their set but none of the top 10 poses for any complex were at that level of accuracy.

However, a few issues are worth noting. The chosen complexes in this study bind in a very rigid manner, with an IRMSD smaller than 2 Å between the unbound and bound forms of the subunits. Second, the simulations from each individual subunit are started in their bound conformation. In addition, most notably, they apply backbone torsional restraints centered at the bound native structures for both subunits for four complexes, while the remaining two complexes have such restraints applied to one subunit. As we have seen from the *Upside* results, even slight backbone deviations can be very detrimental to the accuracy of the simulation and adding some restraints can substantially improve performance. Pan et al. note that the restraints help prevent conformational degradation at the (hundreds of) microsecond timescales they need to simulate in order to observe reversible association. This observation is consistent with our results. While an in-depth assessment of the effects of torsional restraints was not provided, the association rates of barnase−barstar with and without restraints was also examined with conventional MD. Interestingly, the predicted association rate was about five times slower without restraints than with restraints compared to experiment. In the end, they also conclude that detailed forcefields will have difficulty modeling systems for which the unbound subunit conformations differ significantly from the bound states when torsional corrections cannot be relied upon.[10]

Determining Features That Contribute to Performance. We conducted an analysis of which factors contribute to

https://doi.org/10.1021/acs.jctc.1c01255
J. Chem. Theory Comput. 2022, 18, 2016−2032

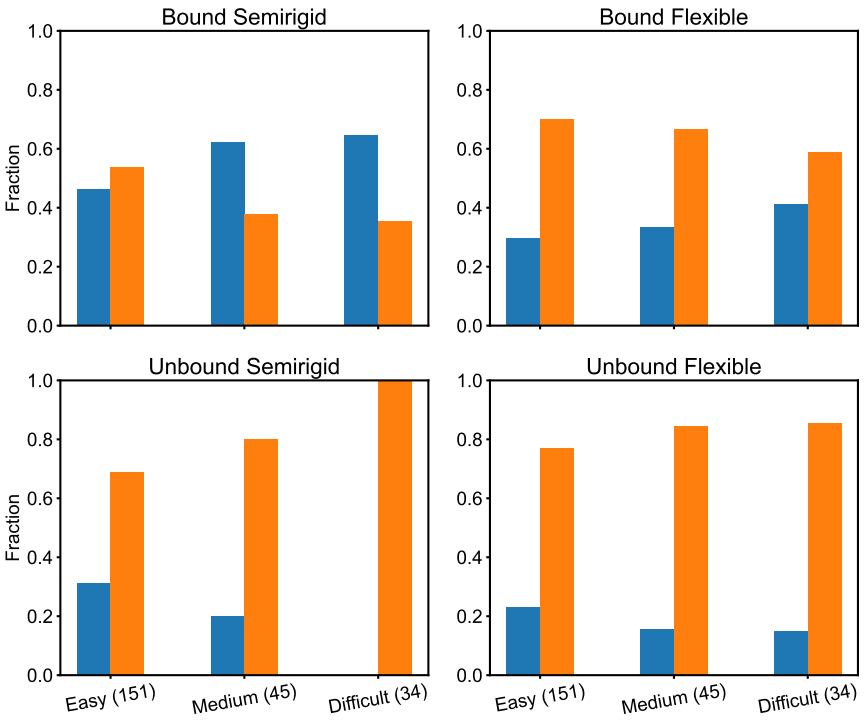


Figure 7. *Upside*'s performance with difficulty class of complexes. Each subplot corresponds to the subunit starting conformation and backbone flexibility scenarios of the CAPRI evaluation in Figure 3. Easy: IRMSD < 1.5 Å and $f_{\text{non-nat}}$ < 0.40, difficult: I-RMSD > 2.2 Å, and medium: all others.[7] The *y*-axes indicate the fraction of complexes in a difficulty class that are good (native-like poses scored in the top 10) or poor performing with *Upside*.

Upside's performance beginning with a feature reflecting backbone flexibility, the conformational change at the interface between bound and unbound forms of the subunits (Figure 7). The difficulty categories are easy: IRMSD < 1.5 Å and $f_{\text{non-nat}}$ < 0.40; difficult: I-RMSD > 2.2 Å; medium: all others.[7] The performance labels for each complex (e.g., native-like in top 10, poor performers) are taken from the previous CAPRI criteria evaluation from the respective cases of starting conformation and whether backbone flexibility was allowed during the *Upside* simulations.

The bound semirigid case shown in Figure 7 again represents a best-case scenario, since backbone conformational search is not allowed, thereby focusing on the scoring component of *Upside*'s energy function. In this situation, we find that the performance is not very different between the three difficulty classes. This finding is partly expected since we trained on the bound complexes but also it also indicates that *Upside* does not have a harder time learning the properties of the interfaces for the different difficulty classes. In the bound flexible situation, performance decreases across the board due to issues with our forcefields, as previously discussed.

In the unbound and semirigid case, we have fewer native-like (top) performers in large part because the subunit backbones are being held in their unbound conformations. A contributing factor to this lack of native-like poses comes from the rigid-body docking stage of our pipeline where FRODOCK is unable to find the general binding interface due to a loss of lock and key fit of the surfaces. Moreover, even if the general native binding interface is found, the unbound conformation subunit backbone RMSD can be a detriment to the IRMSD and CAPRI score. We also find that the performance correlates with the difficulty class of the complexes as expected. When we allow for backbone flexibility in the unbound flexible scenario, we lose some performance on the easy complexes but gain for the difficult complexes, implying that there exist some cases where our energy function can drive the backbones in the correct direction.

We next examine the effects of interface composition and size under the bound semirigid scenario in order to focus on the interactions rather than the conformation search. In Figure 8a, we compare performance based on the amount of pairwise interactions between apolar, polar, and charged residues. The values for each type of pair interaction are normalized according to the maximum of the fraction at the interface between the good and poor performers to make it easier to compare the performance (i.e., the greatest value in each subplot is 1.0).

There are significant differences between the distributions of apolar−apolar and apolar−charged interactions for good and poor performers at 95% confidence level according to the Kolmogorov−Smirnov test. In Figure 8c, we see several significant differences in the interface compositions between the training set and the test set. Notably, apolar−charged interactions tend to be more numerous for the test set and hence, *Upside* had fewer examples of high apolar−charged interfaces to learn from during training. This may explain why apolar−charged interactions tend to be greater for the poor performing complexes. This suggests that *Upside*'s docking forcefield may be slightly improved in the future by reducing the size of the test set from the current ∼24% of the entire set to ∼10%, considering that we have relatively few training examples for our number of parameters (albeit multiplied by the number of decoy poses). The training set should be divided further into *k*-fold cross-validation to find the best stopping point of training to prevent overfitting. Greater improvements may require substantially expanding the training

https://doi.org/10.1021/acs.jctc.1c01255
J. Chem. Theory Comput. 2022, 18, 2016−2032

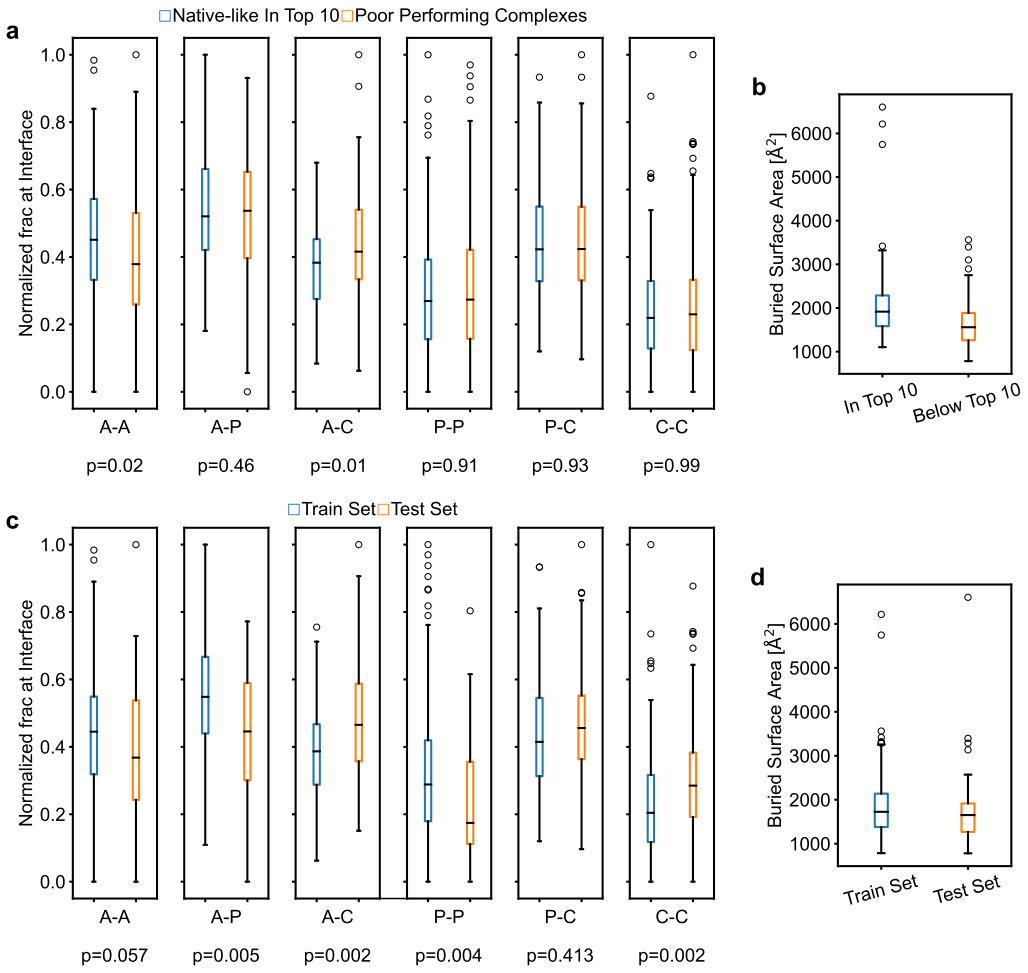


Figure 8. *Upside*'s performance and types of interactions and interface size. (a) Comparison of the fraction of the types of pairwise interactions between good and poorly performing complexes. A: apolar, P: polar, and C: charged. The values for each type of pair interaction are normalized to the difference between the good and poor performers. (b) Comparison of the interface size between the good and poor performers. (c) Comparison of the fraction of different types of pairwise interactions between the training and test sets. (d) Comparison of the interface size between the training and test sets. For panels a and c, the difference between the pairs of distributions is quantified using *p* values.

set, which we detail in the Discussion section. Irrespective of whether these suggestions would produce significant improvements, for comparison purposes, we chose the current size and membership of the test set to correspond to that of Vreven et al.[7]

Complexes with larger interfaces tend to perform better (Figure 8b). This presumably is because larger interfaces use more interactions in calculating the energy, and hence, benefit from an averaging across the interface. In addition, the interface size distribution may be broader with the native interface being more likely to be larger and more favorable.

Vreven et al.[7] were able to find a separating line of performance of the docking algorithms that they tested based on interface area combined with experimental binding energy (K_d) of the complexes, whereas each individual feature (interfacial area, K_d) was only weakly predictive of success. It suggests that binding energy (K_d), which is easier to obtain experimentally than the structure, could be used in a filter to select complexes for which we can be more confident in our predictions. The training and test sets have about the same median interface size and lower bounds (Figure 8d), so performance gains from a rearrangement in the train-test split and *k*-fold cross-validation may not be as influenced by interface size.

We next tested whether we could find a linear combination of features that separate the performance classes. We combined the pair interaction type features and the interface size feature (Figure 8) for the entire set of complexes and performed linear discriminant analysis (LDA) (Figure S4a). We used the labels of good (native-like in top 10) and poor *Upside* performance as the classes to separate in this analysis. LDA returns an output one dimension less than the number of labels/classes. However, this one-dimensional treatment does a poor job of separating the performance classes. Therefore, we increased the number of labels with information about whether the complex belonged to the training or test set for a total of four labels ([train, test] × [good, poor]), since earlier we noticed that some differences in the distribution of features correspond with differences in their distribution in the training and test sets (Figure S4b). However, even this three-dimensional LDA lacks a strong separating surface between the performance classes. Considering that *Upside*'s energy function contains non-linear terms (e.g., the environment energy) and details such as distance and orientation between side chain beads, classification may not be feasible with a linear method and with the simple features that we have chosen. Additionally, because the training procedure simultaneously optimizes thousands of parameters for all the residue types, efforts aimed at

https://doi.org/10.1021/acs.jctc.1c01255
J. Chem. Theory Comput. 2022, 18, 2016−2032

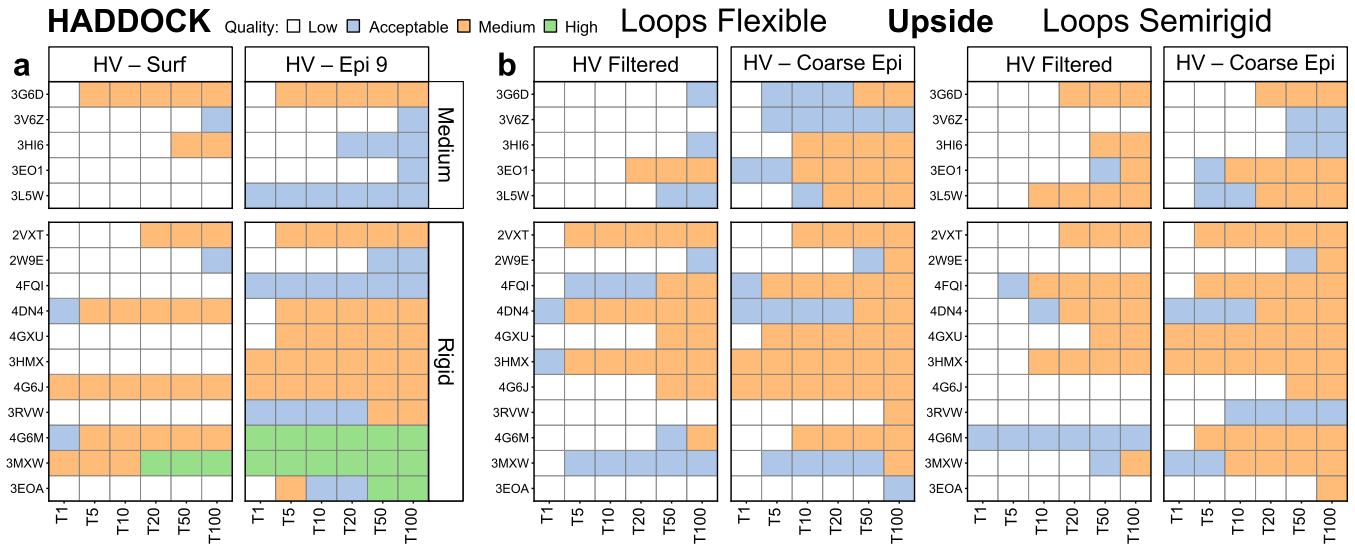


Figure 9. Ab–Ag information-driven docking predictions. (a) CAPRI criteria docking performance of HADDOCK using different levels of informational restraints (adapted with permission from ref 20). (b) *Upside* results with either flexible Ab loops or loops held semirigid. HV filtered uses loop contact information to filter poses but with no biasing potential during *Upside* runs to best correspond to HADDOCK's HV—surf protocol. *Upside*'s HV—coarse Epi protocol uses both loop and coarsely defined epitope information to filter poses, and pairwise sigmoidal contact potentials to bias interactions during *Upside* runs. *Upside*'s HV—coarse Epi roughly corresponds to HADDOCK's HV—Epi 9. Both HADDOCK and *Upside* results use the unbound forms of the subunits as inputs for docking.

pinpointing a small set of features controlling the performance would be unlikely to succeed.

Information-Driven Antibody–Antigen Docking. In some cases, additional sources of information about a complex exist beyond the structure and sequence of the unbound forms. For example, the use of experimental information for challenging complexes is recognized by CAPRI organizers, and they have provided small angle X-ray scattering and cross linking/mass spectrometry data in the round 46 CAPRI-CASP experiment for one such complex.[3] For Ab–Ag docking, the CDR loops can be identified based on the sequence of the conserved protein framework around them, and one could use hydrogen exchange or mutational scanning experiments to glean information about the location of the epitope.[30] It is illustrative to examine the extent that limitations of docking algorithms can be overcome by the incorporation of this extra information.

We follow Ambrosetti et al.'s comparative study[20] that presents the performance of individual complexes as opposed to the aggregate of all complexes, to obtain finer grained insights on the impact of both flexibility and auxiliary information on prediction accuracy. Results for the information-driven Ab–Ag docking are presented in Figure 9 according to CAPRI criteria with the same ranking of native-like structures as used in Figure 3. They found that their HADDOCK algorithm, which involves torsional and explicit solvent flexible refinement stages, does not perform well with biasing information solely from the antibody HV loops (HV-Surf), with several complexes lacking any native-like pose in the top 100 ranks (Figure 9a). However, the biasing of interactions between HV loops along with a coarse definition of epitope residues produces substantial gains (Figure 9a, column heading HV—Epi 9).

Likewise, *Upside* lacks native-like predictions for many complexes when just filtering for poses in contact with the Ab loops (Figure 9b, HV filtered). However, as with HADDOCK, *Upside* shows improvement when augmented with coarse epitope information (HV—coarse Epi) for filtering to include poses where both loops and epitope are in contact and biasing the interactions between them during the *Upside* runs.

The inclusion of loop flexibility has mixed outcomes for *Upside*. First, a comparison of the HV Filtered runs between the flexible and semi-rigid loop cases finds that some complexes either worsen their ranking or quality of their poses when backbone flexibility is allowed (e.g., 3G6D, 3HI6, 3LW5, 4FQI, and 4G6M) although others gain in ranking or quality (e.g., 2VXT, 4DN4, 3HMX, and 3MXW). When epitope information is used (e.g., HV—coarse Epi), flexibility generally produces better predictions (e.g., 3G6D, 3V6Z, 3HI6, and 4G6J). Here, the contact biasing potentials compensate for inaccuracies in the folding and protein–protein energy terms. Ambrosetti et al. similarly observed that biasing with higher levels of information was required to help with packing the antibody H3 loops during their flexible refinement stage. Improvement is required in the underlying models and forcefields of both *Upside* and HADDOCK for unaided flexible docking.

This analysis indicates that *Upside* and HADDOCK perform similar but have different strengths, appreciating that their filtering and biasing schemes are not identical. When epitope information is used and loops are flexible, *Upside* does better for some medium difficulty complexes (3V6Z, 3HI6, 3EO1). This improvement demonstrates the benefit of *Upside*'s greater backbone flexibility. However, HADDOCK produces high-quality predictions for some of the more rigid complexes (4G6M, 3MXW, 3EOA), whereas *Upside* is unable to produce any high-quality structures as noted before with the general data set in Figure 3.

As each CDR loop is known to favor certain clusters of canonical conformations,[31,32] we also compare our predicted cluster assignments of the loops to those of the native antigen bound structures. This analysis is based on the *Upside* structures obtained from the loops flexible CAPRI evaluation scenario of the preceding discussion. The PyIgClassify server[31]

https://doi.org/10.1021/acs.jctc.1c01255
J. Chem. Theory Comput. 2022, 18, 2016–2032

was used to assign CDR loop structures to known clusters according to the loop's backbone dihedral angles. Considering that the most structurally diverse CDR H3 loop poses a challenge for most contemporary computational tools,[33] we are particularly motivated to assess how *Upside*'s predictions fare.

Cluster assignments were done for all 16 antibody complexes, and with the four different scenarios for each complex using either: (1) the known bound forms of the loops; (2) the unbound forms of the loops, which were the starting structures of the *Upside* relaxation simulations; (3) best scoring native-like structure from *Upside* using only CDR loop information to assist the docking; (4) best scoring native-like structure from *Upside* using a bias between the loops and the coarse epitope to assist the docking. Figure 10 summarizes

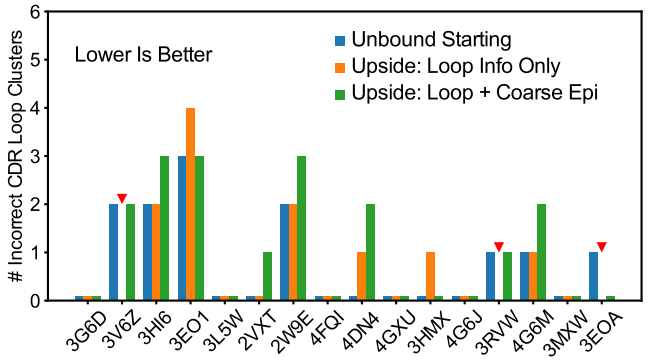


Figure 10. Number of incorrect CDR loop cluster assignments compared to the native assignment by PyIgClassify. The blue bars represent CDR loops in the unbound conformation, which are the starting points for the *Upside* relaxation simulations. The orange and green bars are for the best scoring native-like *Upside* prediction under the two different protocols. The orange bars are for the loop-information only case where the antigen was simply filtered to be in contact with the CDR loops. The green bars are for the case with bias between the loops and the coarsely defined epitope. The red triangles red ▼ designate cases where structures are missing for the no bias case because no native pose scored within the top 100 poses. The near zero bars are visual aids to indicate data exists for those cases and they have zero incorrect assignments.

CDR loop prediction accuracy, where the bars count the number of loops that have been assigned to a wrong, non-native cluster. The 3G6D complex, where no bars are visible, indicates that unbound and *Upside* structures for that complex have the same cluster assignments as the native conformations for all six CDRs. At the other end, the 3EO1 complex has three or more CDRs that are not predicted to be in the native cluster.

More importantly, the loops in the *Upside* simulations largely remain in their original clusters irrespective of the information level used to assist the docking. Buried portions likely encounter steric hindrance and interactions that trap them in their starting conformation. Literature suggests that the CDR loops (particularly H3) span a spectrum of flexibility,[34,35] and even among our test set, examples exist of complexes with loops that undergo conformational change between unbound and bound states, so there is motivation to improve the conformational sampling of the more flexible regions. Figure S5 further illustrates this point with visualizations of the loop structures for the two complexes (3EO1, 3HI6) that have the largest discrepancy between our predicted loop structure cluster assignments and the native assignments. During the *Upside* simulations, the loops generally do not change their backbone RMSD by more than ∼3 Å from the starting structures. Although the internal loop conformations are retained, the loops still undergo center-of-mass shifts and tilts when referenced to the rest of the entire antibody (Figure S6) and so the binding surface changes and explains we can obtain better CAPRI predictions in the flexible loop under the epitope bias scenario.

For comparison, Ambrosetti et al. investigated the RMSD of H3 loop of their predicted models after flexible refinement with alignment of the framework residues to the antigen bound structure (i.e., overall shifts in position and orientation were included in their RMSDs). The accuracy tended to get worse by up to ∼1.5 Å for the easy complexes that already start with low H3 RMSD in the unbound forms. They saw an improvement by up to 1.25 Å in some of their predictions for medium difficulty complexes when coarse epitope information was used, but they also saw a degradation for some others. Overall, HADDOCK has mixed performance with CDR loop prediction.

We investigated whether *Upside* is capable of sampling native-like H1 loop conformations by performing temperature replica exchange MD (TREMD) simulations of 3EO1. In the runs initiated from the unbound conformation in the absence of antigen, we found that the H1 loop can sample lower RMSD states compared to those previously found during the docking (Figure 11). The previous docking simulations were done at a

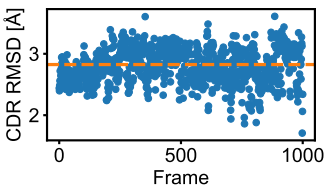


Figure 11. H1 CDR loop RMSD (in Å) from TREMD of 3EO1 antibody separated from antigen. The dashed line indicates the RMSD (Å) from the best scoring docking model.

relatively low and constant temperature. This suggests that to sample the native conformation, we likely require enhanced sampling procedures.

In conclusion, the docking algorithms, including *Upside*, do a relatively poor job of predicting antibody binding with CDR loop information only. Biasing interactions using a coarse definition of the epitope greatly improves the results, with *Upside*'s greater efficiency in flexible backbone sampling as compared HADDOCK enabling better predictions of medium difficulty complexes. While center-of-mass position and orientation of the loops shift during this flexible sampling relative to the framework residues and thus the binding surface changes, the internal structures of the loops do not change much over the course of *Upside* simulations in contact with antigen. Enhanced sampling or a conformational selection scheme (e.g., cross-docking different pre-sampled unbound conformations) should improve the prediction of the internal CDR loop structures.

DISCUSSION

Ideally, models representing proteins should capture all aspects of their behavior. However, depending on the objectives and computational resources, one often needs to make approximations. Identifying which kinds of approximations perform

https://doi.org/10.1021/acs.jctc.1c01255
J. Chem. Theory Comput. 2022, 18, 2016−2032

well for a given problem can provide new insight. We explored these issues using our fast MD algorithm, *Upside*, which makes the approximations of representing each side chain by a multiposition bead and a lack of explicit solvent yet performs reasonably well for simulating protein folding and dynamics.[21,22] By adapting *Upside* to protein−protein docking, our hope was that its backbone sampling capabilities would improve the prediction of complexes that undergo conformational change upon binding. The reality turned out to be more complicated.

We found that the addition of specific interprotein energy terms considerably improved our ability to score structures. Combinations of pairwise polar and charge side chain interactions and the burial of hydrophilic side chains were underrepresented in the original training set used for protein folding. Therefore, it is not surprising that the original forcefield designed for folding could be improved for protein−protein docking.

Even with the single bead representation of the side chains, *Upside*'s performance was comparable to traditional full side chain methods when the subunit structures were restrained in their unbound conformations. Only in a few cases where the subunits did not change conformation by more than a few Å in the *Upside* simulations did the addition of side chains with SCWRL4[26] and rescoring with SOAP-PP[23] produce a significant benefit. The CABS CG method, which also lacks explicit side chains, was also able to find some low IRMSD predictions in the top 10 scored predictions,[29] which when considered with our results indicates that moderate success in docking can be achieved without explicit side chains. However, we should note that SCWRL4 and SOAP-PP do not give a perfect reconstruction and scoring of side chains, and it is still likely that explicit side chains are required for high accuracy predictions.

Backbone Flexiblity. Surprisingly, inclusion of backbone flexibility can act both as an advantage and a disadvantage for docking. The accuracy of the approach generally decreased when full backbone flexibility was allowed even starting from the native pose and subunit conformations. This outcome is often due to the lowest energy structure of the individual subunits not being within 3 Å of their bound structures ("Both Subunits Free" panels in Figure 4). The energy is a function of both the folding and binding energy terms; the PMF-determined energy needed to shift the subunit structures to their bound conformations was frequently too large for the binding terms alone to overcome (e.g., $>4k_BT$). This steep penalty for what would seem to be a relatively minor error prevented *Upside* from being successful in flexible docking. To improve our procedure would require better training and/or more sophisticated forcefield terms that contribute to folding as well as binding.

The decrease in performance with backbone flexibility is likely endemic to most current docking methods. For example, we found that rigid docking performs better than the large conformational search of the CABS CG method.[29] Even with the extensive sampling enabled by Anton 2, the classical all-atom MD simulations of Pan et al. benefitted from imposing backbone constraints based on the bound conformation of the subunits.[10] In the case of antibody docking, biased simulations using information on antibody CDR loops and epitopes could overcome forcefield inaccuracies. With this extra information, allowing for flexible docking with *Upside* produced better results which were more competitive than HADDOCK's for medium difficulty complexes.[20]

For Ab−Ag docking, we noticed that the CDR loop conformations did not change very much after complex formation due to high energy barriers. This observation suggests that conformational selection before the initial contact plays a role, as opposed to induced fit. This is supported by the presence of more native-like loops in the antibody-only simulations. In this scenario, the natural fluctuations of the unbound antibody access the native bound conformation of the CDR loops, and this state binds the antigen. However, attempting to improve the loop conformation after the initial contact of the unbound forms of the antibody and antigen (as done in the *Upside* docking pipeline) may not succeed because of the aforementioned issue of steric hindrance experienced by the loops. The literature is divided on the relative weight of conformational selection and induced fit in antibody binding.[36,37] Better antibody loop predictions will require either fully flexible free docking that would allow more range of motion for the CDR loops prior to contact between the antibody and antigen or rigid-body cross-docking of antibodies with different pre-sampled loop conformations followed by refinement.

Recent Machine Learning Methods. RoseTTAFold and AlphaFold 2-derived protocols and models have achieved great strides in protein−protein docking.[13,15,16] In particular, AlphaFold-Multimer obtained 67% accuracy in predicting at least acceptable quality heteromeric interfaces as the top 1 prediction,[16] which is several times higher than any of the traditional methods in this paper (≤15%). Traditional methods have clearly been eclipsed for applications where only docking performance is concerned. However, deep neural network approaches with their many interconnected layers suffer from issues of interpretability,[17] and their reliance on coevolutionary information detracts from a physical understanding of the balance of interactions required accurate structure prediction.

Results from quantitative saturation scanning experiments challenged views on conservative mutations and conservation across species: there is a great deal more mutational plasticity than previously believed that preserves the structure and function.[38] Physics-based approaches might be able to predict such a diversity, whereas deep neural network methods that rely on sequence alignments may implicitly incorporate other biological constraints on a sequence space and not be generalizable to engineered systems outside of the biological context. Additionally, Alphafold-Multimer has difficulty predicting antibody complexes,[16] and physics-based methods may be more generalizable for this application as well.

Furthermore, our results that point to conformational selection for antibody loops illustrates the utility of MD tools to investigate mechanistic questions about protein binding and folding, as opposed to the recent neural network approaches. In AlphaFold 2 and RoseTTAFold, backbone sterics, especially of the peptide bond, are not explicitly represented in the primary stages of the models such that when coupled to distance information of specific residue pairs, the models are able to unnaturally search for the optimal positions of residues over long distances while the backbone "clips" or "ghosts" through itself.[13,14] Therefore, these neural network methods are unlikely to be able to differentiate between induced fit and conformational selection binding mechanisms. With MD simulations, backbone sterics and non-specific

https://doi.org/10.1021/acs.jctc.1c01255
J. Chem. Theory Comput. 2022, 18, 2016−2032

interactions hinder the search but provide an avenue to predict thermodynamics, kinetics, and pathways.

Folding versus Binding. Because *Upside* was designed for protein folding rather than docking, it is worth discussing the distinct challenges faced by these two problems. With regard to folding, proteins seemingly have a huge number of possible conformations and yet many manage to fold within seconds.[39] As Rose notes, there are strong organizational constraints imposed by backbone sterics and hydrogen bonding.[40,41] A 100-residue domain thus may have only ∼10 helices and strands, which could be arranged in about 10^3 fundamental folds. Larger proteins consist of such domains so that conformational diversity may grow manageably with length.

To find the correct structure using folding simulations, the energy terms must be balanced and considerable searching is required. In practice, imperfections in force fields can readily result in kinetic trapping. For example, *Upside* is able to fold some proteins comprising less than 100 residues, in part due to a careful consideration of backbone potential terms and training against misfolded structures. However, the method is not perfect because some misfolded states are stable and some native states are not the global energy minimum. This issue becomes much more problematic with longer sequences.

In contrast, the search problem is substantially simpler for rigid-body protein docking. According to Janin's model of barnase−barstar binding, about 70,000 poses are needed to find the native well if states are discretized every 14° with respect to each of the angles about their center-to-center vector.[42] For larger complexes, the number of required poses likely is proportional to the square of the surface area, and the area grows as (molecular weight) $^{0.7}$.[43] Hence, the number of decoys grows manageably and it is possible to generate and score the 10^5−10^6 poses needed to sample the docking space.

Nevertheless, when docking the unbound subunit structures and presented with a docked set of 10^5−10^6 poses, the traditional approaches only have a success rate of finding a native-like pose in the top 10 predictions (i.e., 10 predictions are needed for one to be native-like) for only ∼30% of the complexes. Improvement likely necessitates generating docking poses with subunits that have structures closer to their true bound conformations. Relaxation of the complexes with *Upside* results in 3−4 Å Cα RMSD for the subunits at best; this accuracy can be considered good for folding prediction, but it still produces only medium quality docking poses and the overall docking accuracy is not much improved. Generating high accuracy binding poses requires having subunit structures that are very close to their bound structures and effectively, the challenge of high accuracy binding prediction becomes a challenge in high accuracy structure refinement.

In the case of *Upside*, increasing accuracy may entail joint training of folding and interprotein energy terms. Another major hindrance for binding compared to folding is that there are many more structures to train on for folding. For example, RoseTTAFold was trained on ∼23,000 non-redundant clusters of protein chains.[14] In the present study, we were constrained to 175 complexes for training from the BM5,[7] and some overfitting to the training set is likely. A major contributor to the limitation of this dataset is that both bound and unbound structures of the complex's constituents are required. To overcome this issue and presumably improve our docking performance, we could create a larger training data set of a few thousand binding pairs taken from the AlphaFold Multimer work[16] using only the bound structures, while reserving the complexes with both bound and unbound structures available for testing. We can even leverage AlphaFold 2's accuracy and reporting of uncertainty in its predictions[13] to generate unbound structures to further expand the size of our data set.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.1c01255.

Training, native state simulations, and antibody CDR loop conformations (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Benoît Roux − *Department of Biochemistry and Molecular Biology and Department of Chemistry, University of Chicago, Chicago, Illinois 60637, United States;* orcid.org/0000-0002-5254-2712; Email: roux@uchicago.edu

Tobin R. Sosnick − *Department of Biochemistry and Molecular Biology and Pritzker School of Molecular Engineering, University of Chicago, Chicago, Illinois 60637, United States;* orcid.org/0000-0002-2871-7244; Email: trsosnic@uchicago.edu

Authors

Nabil F. Faruk − *Graduate Program in Biophysical Sciences, University of Chicago, Chicago, Illinois 60637, United States;* orcid.org/0000-0003-3687-7114

Xiangda Peng − *Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, Illinois 60637, United States*

Karl F. Freed − *Department of Chemistry, University of Chicago, Chicago, Illinois 60637, United States;* orcid.org/0000-0001-6240-2599

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.1c01255

Author Contributions

N.F.F.: conceptualization, formal analysis, investigation, methodology, software, validation, writing. X.P.: methodology, software. K.F.F.: conceptualization, supervision. B.R.: conceptualization, funding acquisition, investigation, methodology, project administration, supervision, writing (review and editing). T.R.S.: conceptualization, funding acquisition, investigation, methodology, project administration, supervision, writing (review and editing). All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

The source code and examples for the docking version of *Upside* can be found at https://github.com/nffaruk/upside-docking. The release tagged v2.0.0 corresponds to this paper.

■ ACKNOWLEDGMENTS

We thank John M. Jumper for his early assistance with the modification of the *Upside* code. We also thank Anthony A. Kossiakoff for helpful discussions, especially with antibody docking. This work was supported by an NSERC PGS D PGSD3-475231-2015 Committee no. 187 (N.F.F.), NSF grants MCB-1517221 (B.R.) and MCB-2023077 (T.R.S.), and an NIH grant R01 GM055694 (T.R.S.).

REFERENCES

(1) Porter, K. A.; Desta, I.; Kozakov, D.; Vajda, S. What Method to Use for Protein-Protein Docking? *Curr. Opin. Struct. Biol.* **2019**, *55*, 1−7.

(2) Ramírez-Aportela, E.; López-Blanco, J. R.; Chacón, P. FRODOCK 2.0: fast protein−protein docking server. *Bioinformatics* **2016**, *32*, 2386−2388.

(3) Lensink, M. F.; Brysbaert, G.; Nadzirin, N.; Velankar, S.; Chaleil, R. A. G.; Gerguri, T.; et al. Blind prediction of homo- and hetero-protein complexes: The CASP13-CAPRI experiment. *Proteins: Struct., Funct., Bioinf.* **2019**, *87*, 1200−1221.

(4) Dapkūnas, J.; Olechnovič, K.; Venclovas, Č. Structural modeling of protein complexes: Current capabilities and challenges. *Proteins: Struct., Funct., Bioinf.* **2019**, *87*, 1222−1232.

(5) Lensink, M. F.; Velankar, S.; Wodak, S. J. Modeling protein-protein and protein-peptide complexes: CAPRI 6th edition. *Proteins* **2017**, *85*, 359−377.

(6) Quignot, C.; Rey, J.; Yu, J.; Tufféry, P.; Guerois, R.; Andreani, J. InterEvDock2: an expanded server for protein docking using evolutionary and biological information from homology models and multimeric inputs. *Nucleic Acids Res.* **2018**, *46*, W408−W416.

(7) Vreven, T.; Moal, I. H.; Vangone, A.; Pierce, B. G.; Kastritis, P. L.; Torchala, M.; et al. Updates to the integrated protein-protein interaction benchmarks: Docking benchmark version 5 and affinity benchmark version 2. *J. Mol. Biol.* **2015**, *427*, 3031−3041.

(8) Moal, I. H.; Bates, P. A. SwarmDock and the Use of Normal Modes in Protein-Protein Docking. *Int. J. Mol. Sci.* **2010**, *11*, 3623−3648.

(9) van Zundert, G. C. P.; Rodrigues, J. P. G. L. M.; Trellet, M.; Schmitz, C.; Kastritis, P. L.; Karaca, E.; et al. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* **2016**, *428*, 720−725.

(10) Pan, A. C.; Jacobson, D.; Yatsenko, K.; Sritharan, D.; Weinreich, T. M.; Shaw, D. E. Atomic-level characterization of protein-protein association. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 4244−4249.

(11) Kolinski, A. Protein modeling and structure prediction with a reduced representation. *Acta Biochim. Pol.* **2004**, *51*, 349−371.

(12) Kmiecik, S.; Gront, D.; Kolinski, M.; Wieteska, L.; Dawid, A. E.; Kolinski, A. Coarse-Grained Protein Models and Their Applications. *Chem. Rev.* **2016**, *116*, 7898−7936.

(13) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583−589.

(14) Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G. R.; et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **2021**, *373*, 871−876.

(15) Bryant, P.; Pozzati, G.; Elofsson, A. Improved prediction of protein-protein interactions using AlphaFold2 and extended multiple-sequence alignments. 2021, bioRxiv2021.09.15.460468. Available from: https://www.biorxiv.org/content/10.1101/2021.09.15.460468v1.

(16) Evans, R.; O'Neill, M.; Pritzel, A.; Antropova, N.; Senior, A.; Green, T.; et al. Protein complex prediction with AlphaFold-Multimer. 2021, bioRxiv2021.10.04.463034. Available from: https://www.biorxiv.org/content/10.1101/2021.10.04.463034v1.

(17) Jiménez-Luna, J.; Grisoni, F.; Schneider, G. Drug discovery with explainable artificial intelligence. *Nat. Mach. Intell.* **2020**, *2*, 573−584.

(18) Janin, J. l.; Henrick, K.; Moult, J.; Eyck, L. T.; Sternberg, M. J. E.; Vajda, S.; et al. CAPRI: A Critical Assessment of PRedicted Interactions. *Proteins* **2003**, *52*, 2−9.

(19) Lensink, M. F.; Nadzirin, N.; Velankar, S.; Wodak, S. J. Modeling protein-protein, protein-peptide, and protein-oligosaccharide complexes: CAPRI 7th edition. *Proteins* **2020**, *88*, 916−938.

(20) Ambrosetti, F.; Jiménez-García, B.; Roel-Touris, J.; Bonvin, A. M. J. J. Modeling Antibody-Antigen Complexes by Information-Driven Docking. *Structure* **2020**, *28*, 119−129.

(21) Jumper, J. M.; Faruk, N. F.; Freed, K. F.; Sosnick, T. R. Accurate calculation of side chain packing and free energy with applications to protein molecular dynamics. *PLoS Comput. Biol.* **2018**, *14*, No. e1006342.

(22) Jumper, J. M.; Faruk, N. F.; Freed, K. F.; Sosnick, T. R. Trajectory-based training enables protein simulations with accurate folding and Boltzmann ensembles in cpu-hours. *PLoS Comput. Biol.* **2018**, *14*, No. e1006578.

(23) Dong, G. Q.; Fan, H.; Schneidman-Duhovny, D.; Webb, B.; Sali, A. Optimized atomic statistical potentials: assessment of protein interfaces and loops. *Bioinformatics* **2013**, *29*, 3158−3166.

(24) Skinner, J. J.; Wang, S.; Lee, J.; Ong, C.; Sommese, R.; Sivaramakrishnan, S.; et al. Conserved salt-bridge competition triggered by phosphorylation regulates the protein interactome. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, 13453−13458.

(25) Gaffney, K. A.; Guo, R.; Bridges, M. D.; Muhammednazaar, S.; Chen, D.; Kim, M.; et al. Lipid bilayer induces contraction of the denatured state ensemble of a helical-bundle membrane protein. *Proc. Natl. Acad. Sci. U.S.A.* **2022**, *119*, No. e2109169119.

(26) Krivov, G. G.; Shapovalov, M. V.; Dunbrack, R. L. Improved prediction of protein side-chain conformations with SCWRL4. *Proteins* **2009**, *77*, 778−795.

(27) Lensink, M. F.; Méndez, R.; Wodak, S. J. Docking and scoring protein complexes: CAPRI 3rd Edition. *Proteins* **2007**, *69*, 704−718.

(28) Torchala, M.; Gerguri, T.; Chaleil, R. A. G.; Gordon, P.; Russell, F.; Keshani, M.; et al. Enhanced sampling of protein conformational states for dynamic cross-docking within the protein-protein docking server SwarmDock. *Proteins: Struct., Funct., Bioinf.* **2020**, *88*, 962−972.

(29) Kurcinski, M.; Kmiecik, S.; Zalewski, M.; Kolinski, A. Protein-Protein Docking with Large-Scale Backbone Flexibility Using Coarse-Grained Monte-Carlo Simulations. *Int. J. Mol. Sci.* **2021**, *22*, 7341.

(30) Paterson, Y.; Englander, S. W.; Roder, H. An Antibody Binding Site on Cytochrome c Defined by Hydrogen Exchange and Two-Dimensional NMR. *Science* **1990**, *249*, 755−759.

(31) Adolf-Bryfogle, J.; Xu, Q.; North, B.; Lehmann, A.; Dunbrack, R. L., Jr PyIgClassify: a database of antibody CDR structural classifications. *Nucleic Acids Res.* **2015**, *43*, D432−D438.

(32) Wong, W. K.; Leem, J.; Deane, C. M. Comparative Analysis of the CDR Loops of Antigen Receptors. *Front. Immunol.* **2019**, *10*, 2454.

(33) Marks, C.; Deane, C. M. Antibody H3 Structure Prediction. *Comput. Struct. Biotechnol. J.* **2017**, *15*, 222−231.

(34) Krishnan, L.; Sahni, G.; Kaur, K. J.; Salunke, D. M. Role of Antibody Paratope Conformational Flexibility in the Manifestation of Molecular Mimicry. *Biophys. J.* **2008**, *94*, 1367−1376.

(35) Jeliazkov, J. R.; Sljoka, A.; Kuroda, D.; Tsuchimura, N.; Katoh, N.; Tsumoto, K.; et al. Repertoire Analysis of Antibody CDR-H3 Loops Suggests Affinity Maturation Does Not Typically Result in Rigidification. *Front. Immunol.* **2018**, *9*, 413.

(36) Keskin, O. Binding induced conformational changes of proteins correlate with their intrinsic fluctuations: a case study of antibodies. *BMC Struct. Biol.* **2007**, *7*, 31.

(37) Wang, W.; Ye, W.; Yu, Q.; Jiang, C.; Zhang, J.; Luo, R.; et al. Conformational selection and induced fit in specific antibody and antigen recognition: SPE7 as a case study. *J. Phys. Chem. B* **2013**, *117*, 4912−4923.

(38) Pál, G.; Kouadio, J.-L. K.; 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.* **2006**, *281*, 22378−22385.

(39) Dill, K. A.; Ozkan, S. B.; Shell, M. S.; Weikl, T. R. The Protein Folding Problem. *Annu. Rev. Biophys.* **2008**, *37*, 289−316.

(40) Rose, G. D.; Fleming, P. J.; Banavar, J. R.; Maritan, A. A backbone-based theory of protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 16623−16633.

(41) Rose, G. D. Protein folding - seeing is deceiving. *Protein Sci.* **2021**, *30*, 1606−1616.

https://doi.org/10.1021/acs.jctc.1c01255
J. Chem. Theory Comput. 2022, 18, 2016−2032

(42) Janin, J. The kinetics of protein-protein recognition. *Proteins: Struct., Funct., Bioinf.* **1997**, *28*, 153−161.
(43) Glyakina, A. V.; Bogatyreva, N. S.; Galzitskaya, O. V. Accessible Surfaces of Beta Proteins Increase with Increasing Protein Molecular Mass More Rapidly than Those of Other Proteins. *PLoS One* **2011**, *6*, No. e28464.

Recommended by ACS

mkgridXf: Consistent Identification of Plausible Binding Sites Despite the Elusive Nature of Cavities and Grooves in Protein Dynamics

Damien Monet, Arnaud Blondel, *et al.*

JULY 09, 2019
JOURNAL OF CHEMICAL INFORMATION AND MODELING READ

Discovering Protein Conformational Flexibility through Artificial-Intelligence-Aided Molecular Dynamics

Zachary Smith, Pratyush Tiwary, *et al.*

AUGUST 25, 2020
THE JOURNAL OF PHYSICAL CHEMISTRY B READ

Fast Calculation of Protein–Protein Binding Free Energies Using Umbrella Sampling with a Coarse-Grained Model

Jagdish Suresh Patel and F. Marty Ytreberg

DECEMBER 29, 2017
JOURNAL OF CHEMICAL THEORY AND COMPUTATION READ

Simulation of Reversible Protein–Protein Binding and Calculation of Binding Free Energies Using Perturbed Distance Restraints

Jan Walther Perthold and Chris Oostenbrink

SEPTEMBER 12, 2017
JOURNAL OF CHEMICAL THEORY AND COMPUTATION READ

Get More Suggestions >

https://doi.org/10.1021/acs.jctc.1c01255
J. Chem. Theory Comput. 2022, 18, 2016−2032