THE JOURNAL OF PHYSICAL CHEMISTRY B
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCB

Article

Enhanced Protein Complex Prediction via Rosetta, AlphaFold, and Nondifferential Covalent Labeling Mass Spectrometry

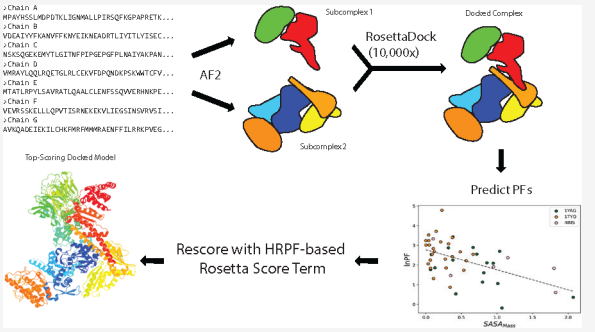
Zachary C. Drake, Alexa G. Fowler, Ashley A. Blum, and Steffen Lindert*

Cite This: J. Phys. Chem. B 2025, 129, 6489−6497

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Covalent labeling (CL) mass spectrometry is a versatile structural technique for elucidating structural information about proteins, including protein−protein complexes. When integrated with computational protein structure prediction, CL can generate accurate protein complex models. Structural insights from CL are more readily interpreted when differential data (e.g., monomer versus complex) are available, as changes in labeling can directly pinpoint interface residues. However, isolating and labeling monomeric subunits can be experimentally challenging, leading to the common scenario in which labeling data are available only in the complex form. However, importantly, nondifferential CL still encodes useful structural information that has yet to be utilized for automated protein complex prediction. In this work, we present a framework for using nondifferential CL in protein complex prediction. We introduced a new hydroxyl radical protein footprinting (HRPF)-derived scoring term that penalizes docked models based on their agreement with experimentally measured CL data. In a benchmark set, the best-scoring models showed an average root-mean-square deviation (RMSD) improvement of 4.6 Å when HRPF data was used. With the inclusion of CL data, four of the top-scoring complexes exhibited RMSDs below 5Å, whereas none did without it. This study demonstrates that CL can enhance protein complex prediction, even when only bound-state CL measurements are available.



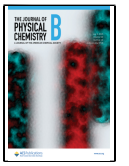
Downloaded via UNIV OF CALIFORNIA LOS ANGELES on August 14, 2025 at 16:16:16 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

INTRODUCTION

Protein−protein complexes are fundamental to virtually all biological processes, acting as the molecular machinery that drives cellular functions. Elucidating the structures of these complexes is pivotal to understanding their functions. Structural insights facilitate the identification of active sites, characterizing interface interactions, and detecting conformational changes, which are key factors in understanding how the complexes function. This understanding, in turn, has the potential to aid in the development of new and more effective therapeutics. The Protein Data Bank (PDB) contains over 230,000 entries, with over 96,000 entries corresponding to protein−protein complexes.[1] Techniques such as X-ray crystallography,[2] nuclear magnetic resonance (NMR) spectroscopy,[3,4] and cryo-electron microscopy[5] frequently provide atomic-resolution structures of proteins. However, these have inherent limitations, including sample preparation, size restrictions, accessibility, and cost, which may slow or hinder structural determination. By providing sparse structural data, structural mass spectrometry (MS)-based methods offer a valuable alternative in cases where traditional high-resolution structural elucidation methods may not be feasible. While structural MS does not yield atomic-resolution structures of proteins, it has several advantages over traditional methods. Structural MS is generally applicable to a wide range of protein sizes, can handle small sample quantities, and requires less time for sample preparation.

Several MS-based approaches provide different types of structural information. Hydrogen−deuterium exchange (HDX),[6] chemical cross-linking,[7] ion mobility (IM),[8] and surface-induced dissociation (SID)[9] reveal solvent accessibility, residue distance restraints, shape and size, complex connectivity, and stoichiometry, respectively. Covalent labeling-MS (CL-MS)[10−12] is another powerful technique that assesses solvent exposure by quantifying the extent of irreversible labeling of residues by specific or nonspecific reagents. CL measures solvent accessibility of residues, with those residues interacting with the solvent being more likely to be labeled than buried, less accessible residues. However, these techniques provide only low-resolution structural information, which is insufficient for generating a complete structure. A high-confidence structural model can be generated only when

Received: April 26, 2025
Revised: June 13, 2025
Accepted: June 17, 2025
Published: June 24, 2025



ACS Publications
© 2025 American Chemical Society

https://doi.org/10.1021/acs.jpcb.5c02872
J. Phys. Chem. B 2025, 129, 6489−6497

combining the CL-MS data with computational protein structure prediction methods.

The field of protein structure prediction has been revolutionized by AlphaFold,[13,14] showcasing the power of artificial intelligence in protein folding and design. However, AlphaFold faces challenges, particularly with dynamic proteins that adopt multiple conformations,[15] orphan proteins with limited homologous sequences,[16] and protein complexes. Its performance for complexes is significantly lower compared to that of monomeric structures. AlphaFold2-Multimer has been shown to accurately predict 60% of dimeric complexes, substantially worse than its predictive capabilities for monomeric structures.[17] Hybrid methods that integrate sparse experimental data with computational modeling, such as those in the Rosetta macromolecular modeling suite, have proven effective in improving structural predictions.[18,19] Incorporation of sparse experimental data as additional scoring terms to the Rosetta REF15 (and previously Talaris14) score function[20,21] has been shown to drastically improve the accuracy of predicted monomers[22−39] and complexes.[40−52]

Hydroxyl radical protein footprinting (HRPF) is a commonly used CL technique that generates radicals via the synchrotron radiolysis of water,[53] Fenton chemistry,[54] or laser-induced photolysis of hydrogen peroxide.[55] Hydroxyl radicals react nonspecifically with nearly all amino acids, offering an advantage over other labeling reagents.[11] HRPF data, analyzed via MS, provide solvent accessibility information about specific residues. In previous work, we demonstrated that differences in labeling between unbound and bound states of a protein complex could infer locations of and residues in protein−protein interfaces.[56] We introduced a differential CL scoring term, showing that integrating CL data into RosettaDock improved complex prediction, even surpassing AlphaFold2-Multimer in several cases. However, isolating the unbound state for the labeling experiments is not always feasible. Thus, we aimed to develop an alternative modeling approach using only bound-state HRPF labeling data.

Since covalent labeling reactivity depends on the side-chain accessibility,[10] residues on a protein's surface that are not involved in inter- or intramolecular interactions are more likely to be labeled. Solvent accessibility metrics, such as solvent accessible surface area (SASA), have been shown to correlate with HRPF oxidation rates.[22,57−60] To account for the intrinsic residue reactivities to hydroxyl radicals, Huang et al. developed a protection factor (PF) analysis, inspired by a similar approach used in HDX. This method converts the kinetic properties of measured oxidative rates into structure-oriented protection factors (PFs),[61] allowing direct comparisons across different peptides. HRPF-derived protection factors have been shown to correlate with metrics of solvent accessibility such as SASA and neighbor count, which have previously been used to computationally score monomeric protein structures.[23,24,43]

In this study, we present a framework for integrating nondifferential HRPF-based covalent labeling into protein complex prediction. We introduce a score term, NonDiffCL$_{Score_Term}$, which penalizes docked complexes based on the agreement of predicted and experimentally measured protection factors. We first observed a correlation between residue protection factors derived from HRPF and SASA, indicating that residues that were more protected from hydroxyl radicals were less solvent accessible. We then used AlphaFold2 to generate protein subunit structures for docking, incorporating ensembles of backbone perturbations generated by using Rosetta methods to sample additional conformations. The methods used include relax[62] for backbone and side-chain optimization, backrub,[63] and shear[64] for local backbone flexibility while preserving structure and geometry, and normal-mode analysis (nma)[65] for exploring low-frequency, large-scale backbone motions. Applying our score term improved the average root-mean-square deviation (RMSD) of top-scoring models across 12 different dockings of three complexes by 4.6 Å, with one-third of the predicted complexes or subcomplexes yielding top-scoring models with RMSDs less than 5 Å.

METHODS

Protein Benchmark Set. The three protein complexes in the benchmark dataset were actin bound to gelosin fragment 1, or actin/gs1 (heterodimer, PDB ID: 1YAG),[66] insulin (hexamer of heterodimers, PDB ID: 4INS),[67] and the Arp2/3 complex (heteroheptamer, PDB ID: 1TYQ).[68] Modification rates for each complex dataset were measured using HRPF experiments where hydroxyl radicals were derived from synchrotron radiolysis. Native crystal structures were used for each complex to benchmark all predicted models. Hydroxyl radical footprinting data were available for the bound states of all three protein complexes.[69−71]

Comparison of PFs and Native Structure SASA. Footprinting rate constants (k_{fp}) were taken from the original data sources. Table S1 lists labeled residues for each protein complex, their labeling rates, and R_i values. k_{fp} were measured for peptides, and across the dataset, there were 29 definite residue assignments, corresponding to only one modified residue in the peptide, and 19 indefinite residue assignments, indicating multiple modified residues. Rate constants were converted to protection factors (PFs) using eq 1, where the sum is over all modified residues of a given peptide, R_i represents the intrinsic reactivity of a residue i to hydroxyl radicals, and k_{fp} represents the footprinting rate constant. R_i values were taken from Huang et al., which were derived under similar synchrotron-based hydroxyl radical protein footprinting conditions[61]

$$\mathrm{PF} = \sum_i \frac{R_i}{k_{fp}} \tag{1}$$

For definite assignments of modified residues, no sum was needed.

Residue SASA was calculated using the "per_residue_sc_sasa" application within the Rosetta software suite using a probe radius of 1.4 Å.[20] For peptides with indefinite rate assignments, the SASA values were averaged across all modified residues. SASA values were normalized by protein mass (SASA$_{Mass}$) to facilitate comparisons across proteins of different sizes, similar to how other methods (such as relative solvent accessibility (RSA)[72] and molecular accessible surface area (MASA)[73]) normalize SASA to account for structural or residue-specific variations.

A linear correlation analysis was performed between SASA$_{Mass}$ calculated using the native structures of our benchmark set and the natural logarithm of HRPF-derived PFs. The slope and intercept of the trendline from this analysis were −1.00 and 2.75, respectively.

Scoring Strategy. In this study, we proposed that SASA could be used to predict the degree of protection a given residue has from hydroxyl radical modification. This can

https://doi.org/10.1021/acs.jpcb.5c02872
J. Phys. Chem. B 2025, 129, 6489−6497

subsequently be used to assess the quality of a computationally docked complex based on the agreement with experimental data. By utilizing the linear parameters obtained from the correlation between ln PFs and $SASA_{Mass}$ values of proteins in our benchmark set, ln PFs were predicted for computationally docked complexes, as shown in eq 2, where the slope is −1.00 and the intercept is 2.75.

$$\ln PF_{predicted} = -1.00 \times SASA_{mass} + 2.75 \quad (2)$$

This method has been integrated within Rosetta through the implementation of a new nondifferential covalent labeling score term ($NonDiffCL_{Score_Term}$). This score term penalized docked models based on the agreement between experimental and predicted ln PFs. Predicted ln PFs of a docked complex were determined by inputting calculated $SASA_{Mass}$ values of modified residues into our linear analysis model (eq 2). The difference between experimental and predicted ln PF (Δ_{lnPF}) was then used as input into a sigmoidal penalty function, eq 3, which assigned greater penalties (P_i) to larger Δ_{lnPF} values. Parameters A and B control the steepness and midpoint of the sigmoid function and were 7.2 and 1.3, respectively.

$$NonDiffCL_{Score_Term} = \sum_i P_i = \sum_i \left(1 - \frac{1}{1 + e^{A(\Delta_{lnPF} - B)}}\right) \quad (3)$$

The penalties calculated with $NonDiffCL_{Score_Term}$ were then weighted by 92.0 and added to the Rosetta REF15 score, as shown in eq 4.

$$NonDiffCL_{Score} = REF15 + 92.0 \times NonDiffCL_{Score_Term} \quad (4)$$

Ensemble Docking. Two of the protein complexes in our benchmark set had oligomeric states greater than those of dimers: 4INS and 1TYQ. Docking with RosettaDock requires only two docking partners. As a result, protein complexes with more than two chains must either be split into multiple docking stages or be treated as multichain subcomplexes for docking. To avoid compounded error, we chose the latter method. Multiple dockings of different subcomplexes were performed to demonstrate different approaches to obtaining full complexes for insulin and Arp2/3. Not all possible unique combinations were used due to an unfeasibly large number of possible docking configurations. Supporting Information Table S1 lists the chains for subunits and subcomplexes used as docking partners.

AlphaFold2[13,74] (AF2) version v.2.3.2 was used to generate protein subunits and subcomplexes, specifying either monomeric or multimer weights as necessary. The recommended settings were applied, including the use of all genetic databases (-db_preset = full_dbs) and templates from the PDB before 2024−01−01. Top-ranked AF2 predictions of each docking partner listed in Table 1 were used as subunits or subcomplexes for docking.

RosettaDock[75] was used to dock the AF2 subunits/subcomplexes using ensemble docking, allowing for additional conformational sampling of each docking partner. Ensembles were compiled from perturbations of AF2 subunits/subcomplexes. Four different methods were used to generate perturbed backbones: backrub,[63] normal-mode analysis (nma),[65] shear,[64] and relax.[62] Ensembles of 10 perturbed structures from each method were created, yielding a total of 40 structures for each docking partner. An input starting structure of each full complex was constructed by using the unperturbed AF2 structures. The ensembles and input structure were then prepacked using the Rosetta 'docking_prepack_protocol' application[76] to allow for side chains outside of the interface to adopt their lowest-energy conformations. For each set of docked partners in Table S2, 10,000 docked models were generated. The position and orientation of the second docking partner were randomized using the '-randomize2' flag in the RosettaDock protocol to allow for additional sampling of initial starting conformations. The top-scoring models were then identified by the lowest Rosetta REF15 score or the lowest $NonDiffCL_{Score}$, respectively.

Table 1. Average RMSD, TM-Score, and DockQ Score with and without the Inclusion of Nondifferential HRPF Data of the Six Models with the Largest Improvement

	average RMSD (Å)	average TM-score	average DockQ score
without nondifferential HRPF data	9.8	0.75	0.47
with nondifferential HRPF data	3.2	0.85	0.63

Analysis Metrics. Accuracy of computationally docked models and AF2 input structures was measured using α-carbon root-mean-square deviation (RMSD), template modeling score (TM-Score),[77] and DockQ score.[78] The native crystal structure of each complex was used as a reference for each of those metrics. α-Carbon RMSDs were calculated using PyMol[79] with the 'align' tool, where zero cycles of refinement with outlier rejections were specified. TM-Score was used to analyze the topological similarity of a docked complex to the native reference. TM-Scores range between 0.0 and 1.0, where a perfect match corresponds to a TM-Score of 1.0. TM-Scores can be categorized as either having random structural similarity (0.0 < TM-Score < 0.17) or high fold similarity (0.5 < TM-Score < 1.00) to the native structure. DockQ is a protein−protein docking quality measurement. Similar to TM-Score, DockQ scores range from 0.0 to 1.0, with a perfect match being equal to 1.0. DockQ score can be assigned to four categories: incorrect (0 < DockQ score < 0.23), acceptable quality (0.23 ≤ DockQ score < 0.49), medium quality (0.49 ≤ DockQ score < 0.80), and high quality (DockQ score > 0.80).

RESULTS AND DISCUSSION

Correlation of Protection Factor and Solvent Accessible Surface Area. The goal of this work was to develop a framework for using nondifferential covalent labeling data to enhance protein complex prediction using Rosetta and AlphaFold2. Our previous approach to integrating differential covalent labeling data into computational protein−protein complex structure prediction relied on the relationship between labeling changes and interface distance. Specifically, we analyzed the percentage change in labeling between the unbound and bound states of the complexes. However, for cases where differential labeling data are not available, an alternative approach was needed. We opted to extract the solvent accessibility of residues solely from bound-state labeling data.

To determine whether the relationship between solvent accessibility and PFs extended to bound protein complexes, we investigated a benchmark set of three proteins with bound-state hydroxyl radical protein footprinting data: actin bound to

https://doi.org/10.1021/acs.jpcb.5c02872
J. Phys. Chem. B 2025, 129, 6489−6497

gelosin fragment 1 (actin/gs1), insulin, and the actin-related protein 2/3 complex (Arp2/3). Crystal structures with resolutions better than 2.55 Å and bound-state modification rates of labeled peptides were available for all complexes. Labeled peptides of these proteins had either definite or indefinite assignments, meaning that they contained either a single or multiple modifiable residues. PFs were calculated for these assignments, as described in the Methods section. A linear least-squares regression was performed between ln PFs and mass-normalized SASAs, or $SASA_{Mass}$, calculated from the crystal structures. Mass normalization of the SASA was performed to account for variances in the solvent exposure with protein size. The resulting trendline, shown in Figure 1,

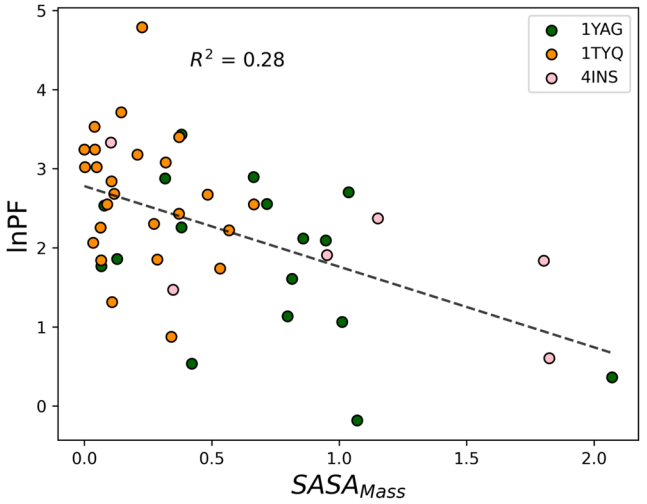


Figure 1. Linear regression of the natural logarithm of the protection factor and mass-normalized solvent accessible surface area. Residues that are exposed to more solvent tend to be less protected from hydroxyl radicals than more buried residues. PDB-IDs for each complex are labeled in the legend.

yielded a coefficient of determination (R^2) of 0.28. Although this is a modest trend from very sparse data, it is similar to previously reported ranges for correlations (0.3−0.4) to SASA and other solvent accessibility metrics.[22,24] suggesting that more-solvent-exposed residues were less protected from hydroxyl radicals and that this information could likely be used to guide protein structure prediction.

Protein Complex Prediction with Nondifferential Covalent Labeling Data. Protein−protein docking within Rosetta was performed by using the RosettaDock application. Here, we present the development of a new nondifferential HRPF-based score term for use in RosettaDock, NonDiffCL$_{Score_Term}$. This score term penalizes large deviations between experimentally measured PFs and predicted PFs of the modified residues. These penalties were summed, weighted, and added to the REF15 score function, forming a new NonDiffCL$_{Score}$. Top-scoring structures using this new score were expected to better reflect nondifferential covalent labeling data.

The effectiveness of the NonDiffCL$_{Score_Term}$ was evaluated using computationally docked complexes generated using RosettaDock. The chains comprising the respective docking partners for each complex are given in Table S2. Subunit/subcomplex structures were initially modeled using AlphaFold2, and structural ensembles with different perturbations were generated to sample alternative backbone conformations. Table S3 contains RMSDs of each subunit and subcomplex generated with AlphaFold2. While these complexes are likely present in AlphaFold2's training set, AlphaFold2-Multimer still performs notably worse on multimers than monomers and is likely to struggle with realistic full protein complex targets.[17,80] Our approach also assumes knowledge of complex stoichiometry when generating subcomplexes with AlphaFold-Multimer or docking the subcomplexes with RosettaDock, since all current prediction tools require explicit specification of the oligomeric state and cannot infer it from sequence alone. Using these AlphaFold2-generated subunit ensembles as input, 10,000 ensemble-docked complexes were generated with RosettaDock. In total, 12 independent docking simulations were performed (see Table S2), resulting in 120,000 docked models.

For each docked complex, $SASA_{Mass}$ values were calculated for each labeled residue. These SASA values were used to predict protection factors using eq 2, from which penalties were then assigned to each model according to eq 3. These penalties were then combined with the Rosetta total score (Figure S1). Figure 2 compares the α-carbon root-mean-square

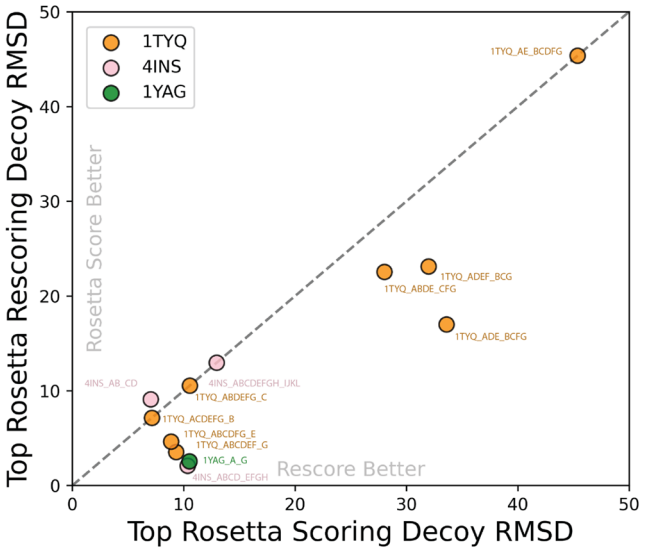


Figure 2. Performance comparison of the top-scoring docked model RMSDs before and after using the NonDiffCL$_{Score_Term}$ score term. Points below the dashed line indicate superior performance after rescoring with the nondifferential CL score term. Identities of protein complexes are labeled and indicated by their color.

deviations (RMSDs) of the top-scoring models before and after incorporating our score term. Prior to applying NonDiffCL$_{Score_Term}$, zero out of the 12 top-scoring models had RMSDs below 5 Å, with an average RMSD of 18.0 Å. After our score term was applied, four out of the 12 top-scoring models had RMSDs below 5 Å, with an improved average RMSD of 13.4 Å and an average improvement of 4.6 Å. Table S4 provides the individual RMSDs, TM-Scores, and DockQ scores of each top-scoring model before and after the inclusion of NonDiffCL$_{Score_Term}$.

A total of four top-scoring models (at least one per complex) exhibited RMSDs below 5 Å after being rescored with HRPF-MS data. Table 1 shows the average analysis metrics for these four top-scoring models before and after rescoring. On average, their RMSD improved by 6.5 Å, while TM-Score and DockQ

https://doi.org/10.1021/acs.jpcb.5c02872
J. Phys. Chem. B 2025, 129, 6489−6497

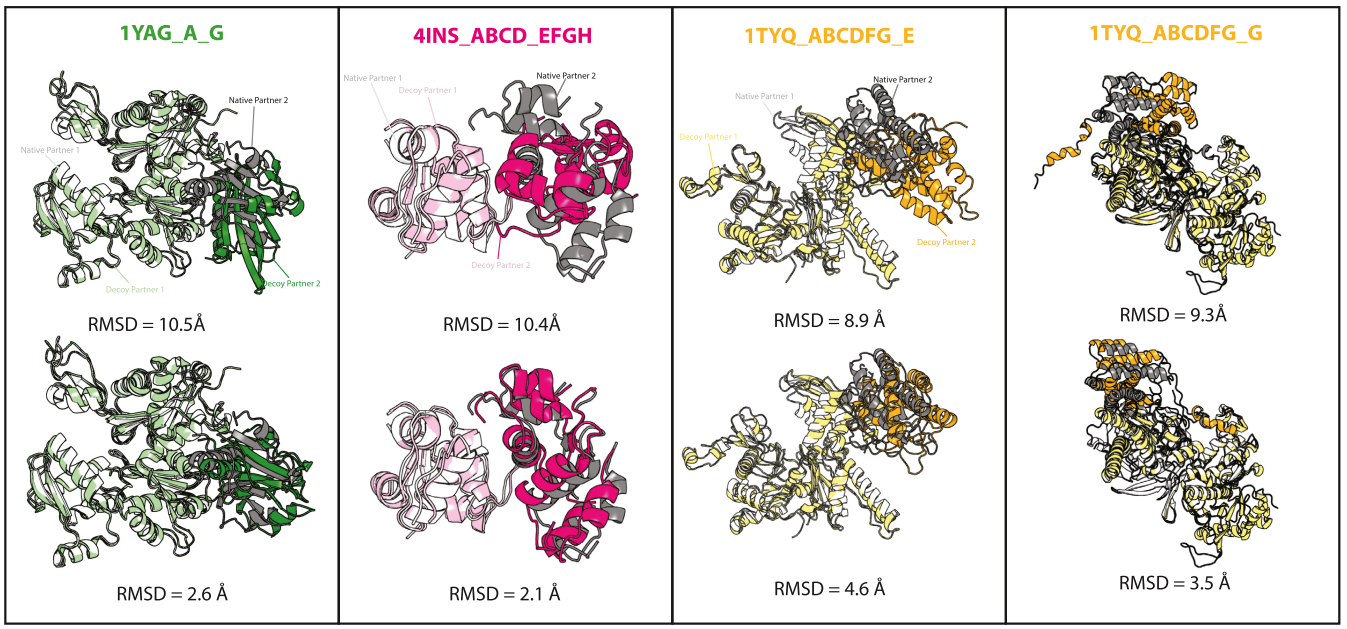


Figure 3. Alignments of top-scoring docked models before (top) and after (bottom) rescoring. The RMSD of each model is calculated against the native crystal structure. Native (gray) and model (different colors) partners are labeled once for each unique protein.

score increased by 0.1 and 0.16, respectively. Figure 3 illustrates the top-scoring models for each of these four complexes, both before and after rescoring, aligned with the native structure.

HRPF Labeling Guides Model Refinement. We visualized the native and top-scoring models with and without the inclusion of our score term for the ABCDFG_E docking scenario for 1TYQ (Figure 4). We wanted to investigate residues with crucial contributions to rescoring and evaluate their environments to better gauge how the score term performed. For this docking scenario, the top-scoring model had an RMSD of 8.9 Å prior to rescoring, and 4.6 Å after rescoring, yielding an improvement of 4.3 Å. Residues M9 and Y47 significantly contributed to rescoring, and both are located on the 21 kDa subunit of the Arp2/3 complex, or "E" chain. Residues M9 and Y47 had experimentally derived natural logarithms of protection factors of 3.71 and 3.4, respectively (Figure 4a). Thus, both residues were indicated to be highly protected, which was observed in the native structure, with residues M9 and Y47 having relative SASA values of 0.14 and 0.37, respectively (Figure 4b). Prior to rescoring, residues M9 and Y47 in the top-scoring model were overly solvent-exposed (SASA: 0.68 vs native 0.37 for Y47, and SASA: 0.19 vs native 0.14 for M9), leading to a larger penalty since the data indicated that these residues should be more protected. After rescoring, SASA values for M9 and Y47 were adjusted to 0.14 and 0.27, respectively, aligning more closely with native values. This example nicely highlights how our score term penalizes residue placements based on indicated buriedness or exposure by HRPF measurements.

CONCLUSIONS

In this study, we developed a novel framework for integrating nondifferential hydroxyl radical protein footprinting (HRPF)-based covalent labeling into protein complex structure prediction. Our results established a correlation between HRPF-derived protection factors and $SASA_{Mass}$, reinforcing the hypothesis that residues with higher protection factors are generally less solvent accessible. Using this linear regression, PFs were predicted for residues in docked complexes from SASA values and penalized based on agreement with experimental PFs. By leveraging solvent accessibility metrics and incorporating their agreement to experiments into Rosetta's docking pipeline through our newly developed $NonDiffCL_{Score_Term}$, we demonstrated significant improvements in docking accuracy. We validated the effectiveness of our score term by evaluating the performance of 12 different docking scenarios across three protein complexes, showing that the inclusion of HRPF data led to an average RMSD improvement of 4.6 Å for top-scoring models. Notably, four complexes achieved top-scoring models with RMSDs below 5 Å, further highlighting the potential of incorporating HRPF-MS data to guide docking refinements. Additionally, we explored how our score term used HRPF data to correctly place labeled residues, in agreement with their observed protection factors.

This study expands the applicability of HRPF-MS data for protein complex modeling, providing a viable alternative when differential covalent labeling data is unavailable. The integration of experimental data with computational modeling continues to be a strong approach for structural biology, bridging the gap between low-resolution experimental techniques and high-accuracy structural predictions. Future work will focus on expanding to other CL techniques and reagents, as well as combining additional complementary experimental MS data types to enhance docking. Although we observed a modest correlation between HRPF-derived PFs and $SASA_{Mass}$, machine learning models have been shown to be a powerful approach for more complex feature detection for a better regression model.[81] Future work might focus on developing a similar model using nondifferential HRPF data for protein complexes. By combining Rosetta and AlphaFold2 with nondifferential covalent labeling mass spectrometry, this work demonstrates a promising hybrid approach to improve protein complex structure prediction, paving the way for more

https://doi.org/10.1021/acs.jpcb.5c02872
J. Phys. Chem. B 2025, 129, 6489−6497

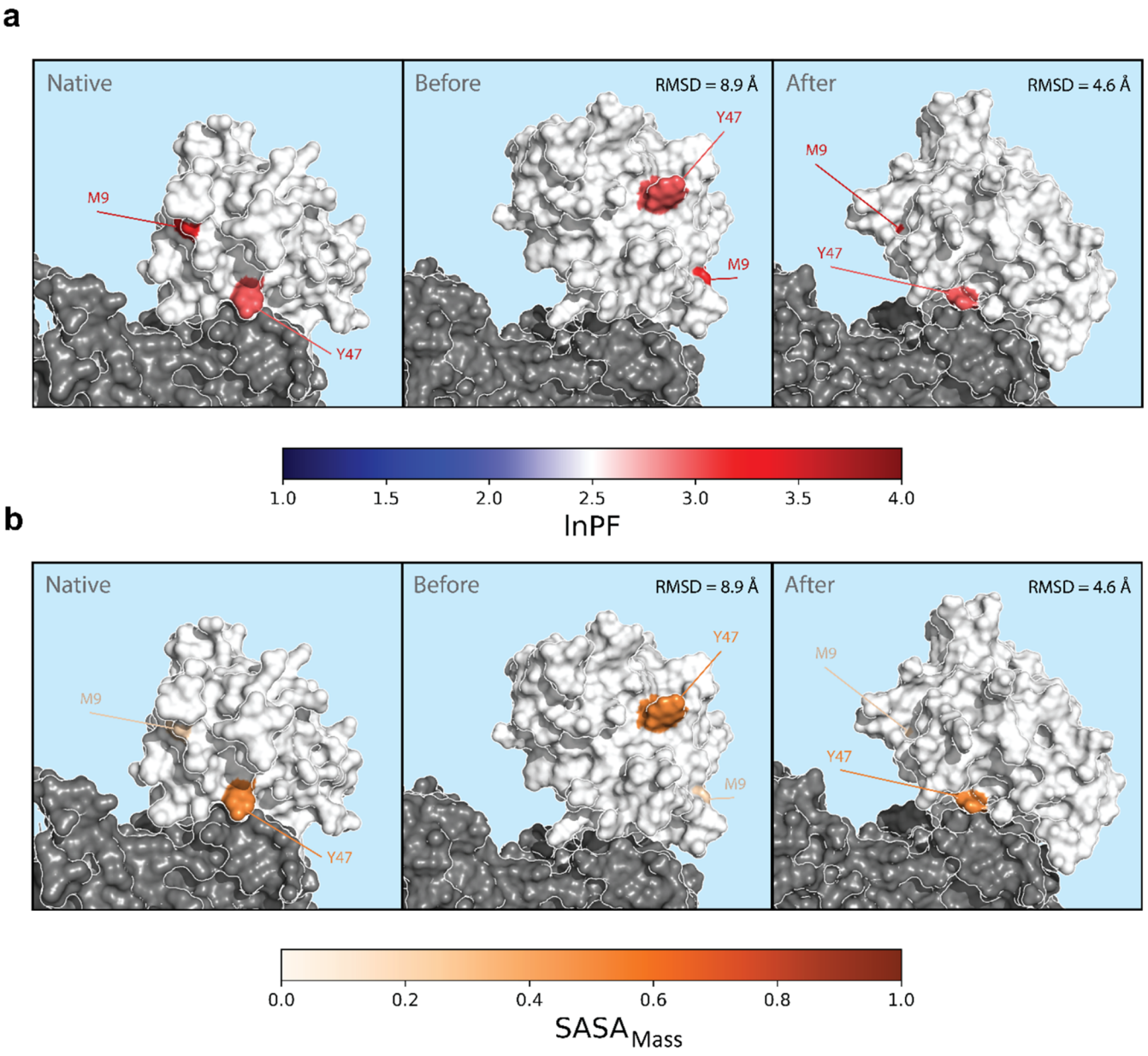


Figure 4. Comparison of the native structure and the top-scoring models before and after the incorporation of HRPF data from the docking of the ABCDFG_E Arp2/3 complex. Docking partners 1 and 2 are shown in dark gray and light gray, respectively. (a) ln PF values of labeled residues are represented by color (see top color bar). (b) $SASA_{Mass}$ values of labeled residues are represented by color (see bottom color bar). The left panel depicts native structures, the middle panel shows the top-scoring model before rescoring, and the right panel presents the top-scoring model after rescoring. Less-protected residues are labeled in blue, and more-protected residues are labeled in red. Less-solvent-exposed residues are labeled in light orange, and more-solvent-exposed residues are labeled in dark orange.

accurate and experimentally informed computational docking methods.

■ ASSOCIATED CONTENT

Data Availability Statement

The cl_complex_rescore application is available for free through the Rosetta software suite at https://github.com/rosettacommons. A tutorial with instructions on how to follow our docking protocol and how to use nondifferential HRPF data with the application is available in the Supporting Information. The crystal structure data used in this study were obtained from the Protein Data Bank (https://www.rcsb.org) with accession codes 1YAG, 4INS, and 1TYQ. A file containing a subset of 100 docked models for each of the four protein complexes with the largest improvement, as well as the labeling data used in this work, is available in the Supporting Information.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.5c02872.

$NonDiffCL_{Score}$ vs RMSD to the crystal structure for 10,000 docked models generated for each complex using AlphaFold2 models as docking input; list of labeled residues and labeling rates; list of docking partners and RMSDs of the top-ranked AF2-generated structures; list of analysis metrics of the top-scoring docked decoys

https://doi.org/10.1021/acs.jpcb.5c02872
J. Phys. Chem. B 2025, 129, 6489−6497

before and after rescoring with nondifferential HRPF; and tutorial for Rosetta application (PDF)

AUTHOR INFORMATION

Corresponding Author

Steffen Lindert − *Department of Chemistry and Biochemistry, Ohio State University, Columbus, Ohio 43210, United States;* orcid.org/0000-0002-3976-3473; Phone: 614-292-8284; Email: lindert.1@osu.edu; Fax: 614-292-1685

Authors

Zachary C. Drake − *Department of Chemistry and Biochemistry, Ohio State University, Columbus, Ohio 43210, United States*

Alexa G. Fowler − *Department of Chemistry and Biochemistry, Ohio State University, Columbus, Ohio 43210, United States*

Ashley A. Blum − *Department of Chemistry and Biochemistry, Ohio State University, Columbus, Ohio 43210, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.5c02872

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank the members of the Lindert laboratory for many useful discussions and the Ohio Supercomputer Center[82] for valuable computational resources. This work was supported by the National Institutes of Health (R01GM127267) and a Sloan Research Fellowship to S.L.

REFERENCES

(1) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235−242.

(2) Smyth, M. S.; Martin, J. H. x ray crystallography. *Mol. Pathol.* **2000**, *53* (1), 8−14.

(3) Kay, L. E. NMR studies of protein structure and dynamics. *J. Magn. Reson.* **2011**, *213* (2), 477−491.

(4) Hu, Y.; Cheng, K.; He, L.; Zhang, X.; Jiang, B.; Jiang, L.; Li, C.; Wang, G.; Yang, Y.; Liu, M. NMR-Based Methods for Protein Analysis. *Anal. Chem.* **2021**, *93* (4), 1866−1879.

(5) Yip, K. M.; Fischer, N.; Paknia, E.; Chari, A.; Stark, H. Atomic-resolution protein structure determination by cryo-EM. *Nature* **2020**, *587* (7832), 157−161.

(6) James, E. I.; Murphree, T. A.; Vorauer, C.; Engen, J. R.; Guttman, M. Advances in Hydrogen/Deuterium Exchange Mass Spectrometry and the Pursuit of Challenging Biological Systems. *Chem. Rev.* **2022**, *122* (8), 7562−7623.

(7) Piersimoni, L.; Kastritis, P. L.; Arlt, C.; Sinz, A. Cross-Linking Mass Spectrometry for Investigating Protein Conformations and Protein−Protein Interactions−A Method for All Seasons. *Chem. Rev.* **2022**, *122* (8), 7500−7531.

(8) Christofi, E.; Barran, P. Ion Mobility Mass Spectrometry (IM-MS) for Structural Biology: Insights Gained by Measuring Mass, Charge, and Collision Cross Section. *Chem. Rev.* **2023**, *123* (6), 2902−2949.

(9) Stiving, A. Q.; VanAernum, Z. L.; Busch, F.; Harvey, S. R.; Sarni, S. H.; Wysocki, V. H. Surface-Induced Dissociation: An Effective Method for Characterization of Protein Quaternary Structure. *Anal. Chem.* **2019**, *91* (1), 190−209.

(10) Mendoza, V. L.; Vachet, R. W. Probing protein structure by amino acid-specific covalent labeling and mass spectrometry. *Mass Spectrom. Rev.* **2009**, *28* (5), 785−815.

(11) Limpikirati, P.; Liu, T.; Vachet, R. W. Covalent labeling-mass spectrometry with non-specific reagents for studying protein structure and interactions. *Methods* **2018**, *144*, 79−93.

(12) Borotto, N. B. The path forward for protein footprinting, covalent labeling, and mass spectrometry-based protein conformational analyses. *J. Mass Spectrom.* **2024**, *59* (7), No. e5064.

(13) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583−589.

(14) Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A. J.; Bambrick, J.; et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **2024**, *630* (8016), 493−500.

(15) Chakravarty, D.; Schafer, J. W.; Chen, E. A.; Thole, J. F.; Ronish, L. A.; Lee, M.; Porter, L. L. AlphaFold predictions of fold-switched conformations are driven by structure memorization. *Nat. Commun.* **2024**, *15* (1), No. 7296.

(16) Wang, L.; Wen, Z.; Liu, S.-W.; Zhang, L.; Finley, C.; Lee, H.-J.; Fan, H.-J. S. Overview of AlphaFold2 and breakthroughs in overcoming its limitations. *Comput. Biol. Med.* **2024**, *176*, No. 108620.

(17) Bryant, P.; Noé, F. Improved protein complex prediction with AlphaFold-multimer by denoising the MSA profile. *PLoS Comput. Biol.* **2024**, *20* (7), No. e1012253.

(18) Seffernick, J. T.; Lindert, S. Hybrid methods for combined experimental and computational determination of protein structure. *J. Chem. Phys.* **2020**, *153* (24), No. 240901.

(19) Biehn, S. E.; Lindert, S. Protein Structure Prediction with Mass Spectrometry Data. *Annu. Rev. Phys. Chem.* **2022**, *73*, 1−19.

(20) Alford, R. F.; Leaver-Fay, A.; Jeliazkov, J. R.; O'Meara, M. J.; DiMaio, F. P.; Park, H.; Shapovalov, M. V.; Renfrew, P. D.; Mulligan, V. K.; Kappel, K.; et al. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **2017**, *13* (6), 3031−3048.

(21) O'Meara, M. J.; Leaver-Fay, A.; Tyka, M. D.; Stein, A.; Houlihan, K.; DiMaio, F.; Bradley, P.; Kortemme, T.; Baker, D.; Snoeyink, J.; Kuhlman, B. Combined Covalent-Electrostatic Model of Hydrogen Bonding Improves Structure Prediction with Rosetta. *J. Chem. Theory Comput.* **2015**, *11* (2), 609−622.

(22) Aprahamian, M. L.; Chea, E. E.; Jones, L. M.; Lindert, S. Rosetta Protein Structure Prediction from Hydroxyl Radical Protein Footprinting Mass Spectrometry Data. *Anal. Chem.* **2018**, *90* (12), 7721−7729.

(23) Aprahamian, M. L.; Lindert, S. Utility of Covalent Labeling Mass Spectrometry Data in Protein Structure Prediction with Rosetta. *J. Chem. Theory Comput.* **2019**, *15* (5), 3410−3424.

(24) Biehn, S. E.; Lindert, S. Accurate protein structure prediction with hydroxyl radical protein footprinting data. *Nat. Commun.* **2021**, *12* (1), No. 341.

(25) Marzolf, D. R.; Seffernick, J. T.; Lindert, S. Protein Structure Prediction from NMR Hydrogen-Deuterium Exchange Data. *J. Chem. Theory Comput.* **2021**, *17* (4), 2619−2629.

(26) Biehn, S. E.; Limpikirati, P.; Vachet, R. W.; Lindert, S. Utilization of Hydrophobic Microenvironment Sensitivity in Diethylpyrocarbonate Labeling for Protein Structure Prediction. *Anal. Chem.* **2021**, *93* (23), 8188−8195.

(27) Nguyen, T. T.; Marzolf, D. R.; Seffernick, J. T.; Heinze, S.; Lindert, S. Protein structure prediction using residue-resolved protection factors from hydrogen-deuterium exchange NMR. *Structure* **2022**, *30* (2), 313−320.E3.

(28) Turzo, S. B. A.; Seffernick, J. T.; Rolland, A. D.; Donor, M. T.; Heinze, S.; Prell, J. S.; Wysocki, V. H.; Lindert, S. Protein shape sampled by ion mobility mass spectrometry consistently improves protein structure prediction. *Nat. Commun.* **2022**, *13* (1), No. 4377.

(29) Turzo, S. M. B. A.; Seffernick, J. T.; Lyskov, S.; Lindert, S. Predicting ion mobility collision cross sections using projection approximation with ROSIE-PARCS webserver. *Briefings Bioinf.* **2023**, *24* (5), No. bbad308, DOI: 10.1093/bib/bbad308.

(30) Khaje, N. A.; Eletsky, A.; Biehn, S. E.; Mobley, C. K.; Rogals, M. J.; Kim, Y.; Mishra, S. K.; Doerksen, R. J.; Lindert, S.; Prestegard, J. H.; Sharp, J. S. Validated determination of NRG1 Ig-like domain structure by mass spectrometry coupled with computational modeling. *Commun. Biol.* **2022**, *5* (1), No. 452.
(31) Cavalli, A.; Salvatella, X.; Dobson, C. M.; Vendruscolo, M. Protein structure determination from NMR chemical shifts. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (23), 9615−9620.
(32) Vernon, R.; Shen, Y.; Baker, D.; Lange, O. F. Improved chemical shift based fragment selection for CS-Rosetta using Rosetta3 fragment picker. *J. Biomol. NMR* **2013**, *57* (2), 117−127.
(33) Alexander, N.; Bortolus, M.; Al-Mestarihi, A.; McHaourab, H.; Meiler, J. De novo high-resolution protein structure determination from sparse spin-labeling EPR data. *Structure* **2008**, *16* (2), 181−195.
(34) Leelananda, S. P.; Lindert, S. Using NMR Chemical Shifts and Cryo-EM Density Restraints in Iterative Rosetta-MD Protein Structure Refinement. *J. Chem. Inf. Model.* **2020**, *60* (5), 2522−2532.
(35) Biehn, S. E.; Picarello, D. M.; Pan, X.; Vachet, R. W.; Lindert, S. Accounting for Neighboring Residue Hydrophobicity in Diethylpyrocarbonate Labeling Mass Spectrometry Improves Rosetta Protein Structure Prediction. *J. Am. Soc. Mass Spectrom.* **2022**, *33* (3), 584−591.
(36) Orbán-Németh, Z.; Beveridge, R.; Hollenstein, D. M.; Rampler, E.; Stranzl, T.; Hudecz, O.; Doblmann, J.; Schlögelhofer, P.; Mechtler, K. Structural prediction of protein models using distance restraints derived from cross-linking mass spectrometry data. *Nat. Protoc.* **2018**, *13* (3), 478−494.
(37) Allison, T. M.; Degiacomi, M. T.; Marklund, E. G.; Jovine, L.; Elofsson, A.; Benesch, J. L. P.; Landreh, M. Complementing machine learning-based structure predictions with native mass spectrometry. *Protein Sci.* **2022**, *31* (6), No. e4333.
(38) Stahl, K.; Graziadei, A.; Dau, T.; Brock, O.; Rappsilber, J. Protein structure prediction with in-cell photo-crosslinking mass spectrometry and deep learning. *Nat. Biotechnol.* **2023**, *41* (12), 1810−1819.
(39) Politis, A.; Park, A. Y.; Hall, Z.; Ruotolo, B. T.; Robinson, C. V. Integrative Modelling Coupled with Ion Mobility Mass Spectrometry Reveals Structural Features of the Clamp Loader in Complex with Single-Stranded DNA Binding Protein. *J. Mol. Biol.* **2013**, *425* (23), 4790−4801.
(40) Seffernick, J. T.; Harvey, S. R.; Wysocki, V. H.; Lindert, S. Predicting Protein Complex Structure from Surface-Induced Dissociation Mass Spectrometry Data. *ACS Cent. Sci.* **2019**, *5* (8), 1330−1341.
(41) Seffernick, J. T.; Canfield, S. M.; Harvey, S. R.; Wysocki, V. H.; Lindert, S. Prediction of Protein Complex Structure Using Surface-Induced Dissociation and Cryo-Electron Microscopy. *Anal. Chem.* **2021**, *93* (21), 7596−7605.
(42) Nance, M. L.; Labonte, J. W.; Adolf-Bryfogle, J.; Gray, J. J. Development and Evaluation of GlycanDock: A Protein−Glycoligand Docking Refinement Algorithm in Rosetta. *J. Phys. Chem. B* **2021**, *125* (25), 6807−6820.
(43) Huang, W.; Ravikumar, K. M.; Parisien, M.; Yang, S. Theoretical modeling of multiprotein complexes by iSPOT: Integration of small-angle X-ray scattering, hydroxyl radical footprinting, and computational docking. *J. Struct. Biol.* **2016**, *196* (3), 340−349.
(44) Borysik, A. J. Simulated Isotope Exchange Patterns Enable Protein Structure Determination. *Angew. Chem., Int. Ed.* **2017**, *56* (32), 9396−9399.
(45) Roberts, V. A.; Pique, M. E.; Hsu, S.; Li, S. Combining H/D Exchange Mass Spectrometry and Computational Docking To Derive the Structure of Protein−Protein Complexes. *Biochemistry* **2017**, *56* (48), 6329−6342.
(46) Tran, M. H.; Martina, C. E.; Moretti, R.; Nagel, M.; Schey, K. L.; Meiler, J. RosettaHDX: Predicting antibody-antigen interaction from hydrogen-deuterium exchange mass spectrometry data. *J. Struct. Biol.* **2025**, *217* (1), No. 108166.
(47) Bolz, R. M.; Seffernick, J. T.; Drake, Z. C.; Harvey, S. R.; Wysocki, V. H.; Lindert, S. Energy Resolved Mass Spectrometry Data from Surfaced Induced Dissociation Improves Prediction of Protein Complex Structure. *Anal. Chem.* **2025**, *97* (4), 2375−2383.
(48) Hauri, S.; Khakzad, H.; Happonen, L.; Teleman, J.; Malmström, J.; Malmström, L. Rapid determination of quaternary protein structures in complex biological samples. *Nat. Commun.* **2019**, *10* (1), No. 192.
(49) Kahraman, A.; Herzog, F.; Leitner, A.; Rosenberger, G.; Aebersold, R.; Malmström, L. Cross-Link Guided Molecular Modeling with ROSETTA. *PLoS One* **2013**, *8* (9), No. e73411.
(50) Webb, B.; Viswanath, S.; Bonomi, M.; Pellarin, R.; Greenberg, C. H.; Saltzberg, D.; Sali, A. Integrative structure modeling with the Integrative Modeling Platform. *Protein Sci.* **2018**, *27* (1), 245−258.
(51) Rantos, V.; Karius, K.; Kosinski, J. Integrative structural modeling of macromolecular complexes using Assembline. *Nat. Protoc.* **2022**, *17* (1), 152−176.
(52) Adzhubei, A. A.; Laughton, C. A.; Neidle, S. An approach to protein homology modelling based on an ensemble of NMR structures: application to the Sox-5 HMG-box protein. *Protein Eng., Des. Sel.* **1995**, *8* (7), 615−625.
(53) Maleknia, S. D.; Brenowitz, M.; Chance, M. R. Millisecond radiolytic modification of peptides by synchrotron X-rays identified by mass spectrometry. *Anal. Chem.* **1999**, *71* (18), 3965−3973.
(54) Nehls, T.; Heymann, T.; Meyners, C.; Hausch, F.; Lermyte, F. Fenton-Chemistry-Based Oxidative Modification of Proteins Reflects Their Conformation. *Int. J. Mol. Sci.* **2021**, *22* (18), No. 9927.
(55) Johnson, D. T.; Di Stefano, L. H.; Jones, L. M. Fast photochemical oxidation of proteins (FPOP): A powerful mass spectrometry-based structural proteomics tool. *J. Biol. Chem.* **2019**, *294* (32), 11969−11979.
(56) Drake, Z. C.; Seffernick, J. T.; Lindert, S. Protein complex prediction using Rosetta, AlphaFold, and mass spectrometry covalent labeling. *Nat. Commun.* **2022**, *13* (1), No. 7846.
(57) Sharp, J. S.; Becker, J. M.; Hettich, R. L. Analysis of protein solvent accessible surfaces by photochemical oxidation and mass spectrometry. *Anal. Chem.* **2004**, *76* (3), 672−683.
(58) Hambly, D. M.; Gross, M. L. Laser flash photolysis of hydrogen peroxide to oxidize protein solvent-accessible residues on the microsecond timescale. *J. Am. Soc. Mass Spectrom.* **2005**, *16* (12), 2057−2063.
(59) Kiselar, J. G.; Maleknia, S. D.; Sullivan, M.; Downard, K. M.; Chance, M. R. Hydroxyl radical probe of protein surfaces using synchrotron X-ray radiolysis and mass spectrometry. *Int. J. Radiat. Biol.* **2002**, *78* (2), 101−114.
(60) Xie, B.; Sood, A.; Woods, R. J.; Sharp, J. S. Quantitative Protein Topography Measurements by High Resolution Hydroxyl Radical Protein Footprinting Enable Accurate Molecular Model Selection. *Sci. Rep.* **2017**, *7* (1), No. 4552.
(61) Huang, W.; Ravikumar, K. M.; Chance, M. R.; Yang, S. Quantitative mapping of protein structure by hydroxyl radical footprinting-mediated structural mass spectrometry: a protection factor analysis. *Biophys. J.* **2015**, *108* (1), 107−115.
(62) Nivón, L. G.; Moretti, R.; Baker, D. A Pareto-Optimal Refinement Method for Protein Design Scaffolds. *PLoS One* **2013**, *8* (4), No. e59004.
(63) Davis, I. W.; Arendall, W. B.; Richardson, D. C.; Richardson, J. S. The Backrub Motion: How Protein Backbone Shrugs When a Sidechain Dances. *Structure* **2006**, *14* (2), 265−274.
(64) Rohl, C. A.; Strauss, C. E. M.; Misura, K. M. S.; Baker, D. Protein Structure Prediction Using Rosetta. In *In Methods in Enzymology*; Academic Press, 2004; Vol. *383*, pp 66−93.
(65) Fleishman, S. J.; Leaver-Fay, A.; Corn, J. E.; Strauch, E.-M.; Khare, S. D.; Koga, N.; Ashworth, J.; Murphy, P.; Richter, F.; Lemmon, G.; et al. RosettaScripts: A Scripting Language Interface to the Rosetta Macromolecular Modeling Suite. *PLoS One* **2011**, *6* (6), No. e20161.
(66) Vorobiev, S.; Strokopytov, B.; Drubin, D. G.; Frieden, C.; Ono, S.; Condeelis, J.; Rubenstein, P. A.; Almo, S. C. The structure of

https://doi.org/10.1021/acs.jpcb.5c02872
J. Phys. Chem. B 2025, 129, 6489−6497

nonvertebrate actin: Implications for the ATP hydrolytic mechanism. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100* (10), 5760−5765.
(67) Baker, E. N.; Blundell, T. L.; Cutfield, J. F.; Dodson, E. J.; Dodson, G. G.; Hodgkin, D. M. C.; Hubbard, R. E.; Isaacs, N. W.; Reynolds, C. D.; Sakabe, K.; et al. The structure of 2Zn pig insulin crystals at 1.5 Å resolution. *Philos. Trans. R. Soc. London B, Biol. Sci.* **1988**, *319* (1195), 369−456.
(68) Nolen, B. J.; Littlefield, R. S.; Pollard, T. D. Crystal structures of actin-related protein 2/3 complex with bound ATP or ADP. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (44), 15627−15632.
(69) Guan, J.-Q.; Almo, S. C.; Reisler, E.; Chance, M. R. Structural Reorganization of Proteins Revealed by Radiolysis and Mass Spectrometry: G-Actin Solution Structure Is Divalent Cation Dependent. *Biochemistry* **2003**, *42* (41), 11992−12000.
(70) Kiselar, J. G.; Datt, M.; Chance, M. R.; Weiss, M. A. Structural Analysis of Proinsulin Hexamer Assembly by Hydroxyl Radical Footprinting and Computational Modeling*. *J. Biol. Chem.* **2011**, *286* (51), 43710−43716.
(71) Kiselar, J. G.; Mahaffy, R.; Pollard, T. D.; Almo, S. C.; Chance, M. R. Visualizing Arp2/3 complex activation mediated by binding of ATP and WASp using structural mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (5), 1552−1557.
(72) Tien, M. Z.; Meyer, A. G.; Sydykova, D. K.; Spielman, S. J.; Wilke, C. O. Maximum Allowed Solvent Accessibilites of Residues in Proteins. *PLoS One* **2013**, *8* (11), No. e80635.
(73) Konstantinidis, K.; Karakasiliotis, I.; Anagnostopoulos, K.; Boulougouris, G. C. On the estimation of the molecular inaccessible volume and the molecular accessible surface of a ligand in protein−ligand systems. *Mol. Syst. Des. Eng.* **2021**, *6* (11), 946−963.
(74) Evans, R.; O'Neill, M.; Pritzel, A.; Antropova, N.; Senior, A.; Green, T.; Žídek, A.; Bates, R.; Blackwell, S.; Yim, J.et al.Protein complex prediction with AlphaFold-Multimer *bioRxiv* 2021 DOI: 10.1101/2021.10.04.463034.
(75) Gray, J. J.; Moughon, S.; Wang, C.; Schueler-Furman, O.; Kuhlman, B.; Rohl, C. A.; Baker, D. Protein−Protein Docking with Simultaneous Optimization of Rigid-body Displacement and Side-chain Conformations. *J. Mol. Biol.* **2003**, *331* (1), 281−299.
(76) Wang, C.; Schueler-Furman, O.; Baker, D. Improved side-chain modeling for protein-protein docking. *Protein Sci.* **2005**, *14* (5), 1328−1339.
(77) Zhang, C.; Shine, M.; Pyle, A. M.; Zhang, Y. US-align: universal structure alignments of proteins, nucleic acids, and macromolecular complexes. *Nat. Methods* **2022**, *19* (9), 1109−1115.
(78) Mirabello, C.; Wallner, B.DockQ_v2: Improved automatic quality measure for protein multimers, nucleic acids, and small molecules *bioRxiv* 2024 DOI: 10.1101/2024.05.28.596225.
(79) Schrodinger, L. *The PyMol. Molecular Graphics System, Version ~1.8.*, 2015.
(80) Yang, Z.; Zeng, X.; Zhao, Y.; Chen, R. AlphaFold2 and its applications in the fields of biology and medicine. *Signal Transduction Targeted Ther.* **2023**, *8* (1), No. 115.
(81) Day, E. H.; Lindert, S. Extracting Residue Solvent Exposure from Covalent Labeling Data with Machine Learning: A Hybrid Approach for Protein Structure Prediction. *J. Am. Soc. Mass Spectrom.* **2025**, *36* (6), 1336−1347.
(82) The Ohio Supercomputer Center. 1987.



https://doi.org/10.1021/acs.jpcb.5c02872
J. Phys. Chem. B 2025, 129, 6489−6497