

RESEARCH ARTICLE

Growing Glycans in Rosetta: Accurate *de novo* glycan modeling, density fitting, and rational sequon design

Jared Adolf-Bryfogle^{1,2,3,4,5*}, Jason W. Labonte⁶, John C. Kraft^{7,8}, Maxim Shapovalov⁹, Sebastian Raemisch^{1,2,3}, Thomas Lütke¹⁰, Frank DiMaio^{7,8}, Christopher D. Bahl^{4,5}, Jesper Pallesen^{11,12}, Neil P. King^{7,8}, Jeffrey J. Gray⁶, Daniel W. Kulp¹², William R. Schief^{1,2,3*}

1 Department of Immunology and Microbiology, The Scripps Research Institute, La Jolla, California, United States of America, **2** IAVI Neutralizing Antibody Center, The Scripps Research Institute, La Jolla, California, United States of America, **3** Consortium for HIV/AIDS Vaccine Development, The Scripps Research Institute, La Jolla, California, United States of America, **4** Institute for Protein Innovation, Boston, Massachusetts, United States of America, **5** Division of Hematology-Oncology, Boston Children's Hospital, Harvard Medical School, Boston, Massachusetts, United States of America, **6** Department of Chemistry & Biomolecular Engineering, Johns Hopkins University, Baltimore, Maryland, United States of America, **7** Department of Biochemistry, University of Washington, Seattle, Washington, United States of America, **8** Institute for Protein Design, University of Washington, Seattle, Washington, United States of America, **9** Fox Chase Cancer Center, Philadelphia, Pennsylvania, United States of America, **10** Institute of Veterinary Physiology and Biochemistry, Justus-Liebig-University Giessen, Giessen, Germany, **11** Department of Molecular and Cellular Biochemistry, Indiana University, Bloomington, Indiana, United States of America, **12** Vaccine and Immunotherapy Center, The Wistar Institute, Philadelphia, Pennsylvania, United States of America

* jadolfbr@gmail.com (JAB); schief@scripps.edu (WRS)



OPEN ACCESS

Citation: Adolf-Bryfogle J, Labonte JW, Kraft JC, Shapovalov M, Raemisch S, Lütke T, et al. (2024) Growing Glycans in Rosetta: Accurate *de novo* glycan modeling, density fitting, and rational sequon design. PLoS Comput Biol 20(6): e1011895. <https://doi.org/10.1371/journal.pcbi.1011895>

Editor: Ozlem Keskin, Koç University, TURKEY

Received: April 18, 2023

Accepted: February 6, 2024

Published: June 24, 2024

Copyright: © 2024 Adolf-Bryfogle et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All main methods developed in this manuscript are part of the Rosetta software suite, which is free for academic use with full code released (<https://www.rosettacommons.org>). All associated scripts are available as supplemental or attached as text in the supplemental section.

Funding: This work was supported by NIAID grants U19AI117905, R01AI11386, 1UM1 AI100663, and UM1 AI144462 to WRS, by PATH Malaria Vaccine Initiative under Grant OPP1141162

Abstract

Carbohydrates and glycoproteins modulate key biological functions. However, experimental structure determination of sugar polymers is notoriously difficult. Computational approaches can aid in carbohydrate structure prediction, structure determination, and design. In this work, we developed a glycan-modeling algorithm, *GlycanTreeModeler*, that computationally builds glycans layer-by-layer, using adaptive kernel density estimates (KDE) of common glycan conformations derived from data in the Protein Data Bank (PDB) and from quantum mechanics (QM) calculations. *GlycanTreeModeler* was benchmarked on a test set of glycan structures of varying lengths, or “trees”. Structures predicted by *GlycanTreeModeler* agreed with native structures at high accuracy for both *de novo* modeling and experimental density-guided building. We employed these tools to design *de novo* glycan trees into a protein nanoparticle vaccine to shield regions of the scaffold from antibody recognition, and experimentally verified shielding. This work will inform glycoprotein model prediction, glycan masking, and further aid computational methods in experimental structure determination and refinement.

from the Bill & Melinda Gates Foundation to WRS, by BMGF CAVD funding to the IAVI NAC Center to WRS, by Grant OPP1156262 from the Bill & Melinda Gates Foundation and a generous donation from the Open Philanthropy Project to NPK, NIH awards 1P01AI167966 and P50AI150464 to NPK, and by an NIAID training grant fellowship T32AI007244 to JAB. NIH award R01-GM127278 supported JWL and J.J.G. NIH/NIGMS R35 GM122517 (R. Dunbrack) supported MS. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: Dr. J.J.G. is an unpaid board member of the Rosetta Commons. Under institutional participation agreements between the University of Washington, acting on behalf of the Rosetta Commons, Johns Hopkins University may be entitled to a portion of revenue received on licensing Rosetta software including methods discussed/developed in this study. As a member of the Scientific Advisory Board, J.J.G. has a financial interest in Cyrus Biotechnology. Cyrus Biotechnology distributes the Rosetta software, which may include methods developed in this study. These arrangements have been reviewed and approved by the Johns Hopkins University in accordance with its conflict-of-interest policies. WRS is an employee of Moderna, Inc., but his contributions to this work were all conducted prior to his employment at Moderna. JAB is an employee of Johnson and Johnson Innovative Medicine, Inc., but his contributions to this work were all conducted prior to his current employment.

Author summary

Many biological proteins are chemically modified to induce specific structure and function. Carbohydrates (glycans) are one such modification that play an important role in signaling, stability, solubility, aggregation, and the immune system. In this work, we have developed and extensively benchmarked a computational protocol for predicting the structures of these glycans, while improving the Rosetta Software Suite through the development of new general analysis frameworks, such as the *SimpleMetric* system and extensive glycan tools for modeling and design. We describe the benchmarking, optimization, and use of a novel computational method for three dimensional modeling of glycans and glyco-conjugates called the *GlycanTreeModeler*. This method is unique in that it grows glycans layer-by-layer using extensive data-driven methods. We thoroughly benchmark the method and detail iterative improvement of the *GlycanTreeModeler* through scoring and kinematic improvements, and show how these methods can be useful in glycan-related computational tasks and glycan masking of a novel vaccine scaffold in-vitro and in-vivo.

Introduction

Carbohydrates and glycoproteins are ubiquitous in biological organisms [1]. Viral glycoproteins such as HIV envelope trimer, influenza hemagglutinin, and SARS-CoV-2 spike, employ N-linked glycosylation as an immune evasion strategy, taking advantage of the fact that host glycans on the surface of proteins are usually recognized as “self” by the adaptive immune system [2]. Yet, HIV broadly neutralizing antibodies often target glycans as part of their epitopes [3] [4] [5]. Small carbohydrate residues attached to serine or threonine can act in signaling pathways akin to phosphorylation [6], while glycans on the constant region of antibodies act as mediators of effector function [7] [8]. Glycans can also improve stability [9] and solubility [10], reduce aggregation [11], and even improve biological drug-targeting and vaccine design through glycan masking of off-target regions [12]. However, in the context of a series of protein nanoparticle immunogens, we recently discovered that glycan masking of the protein nanoparticle scaffold itself is unlikely to enhance antigen-specific antibody responses, especially when the displayed antigen is immunodominant over the nanoparticle scaffold [13]. But we also recently showed that high-density and high-mannose glycans on protein nanoparticle surfaces increase lymph node trafficking and antibody responses against the nanoparticle in a density- and mannose-dependent manner [14]. Thus, optimizing the density and composition of glycans displayed on protein-based vaccines—either on the antigen and/or protein nanoparticle scaffold—provides a framework for engineering glycan recognition to optimize vaccine efficacy.

The biosynthesis of glycoconjugates is complex. Carbohydrates can be attached to certain amino acid residues including serine, threonine, asparagine, and (rarely) tryptophan through covalent modification, forming glycoproteins. The attachment can be made to nitrogen, oxygen, or carbon atoms, (known as N-, O-, or C-linked glycosylation, respectively), with each process involving a multitude of enzymes, sugar moieties and resulting carbohydrate structures. These processes are stochastic in nature, producing glycoproteins that are heterogeneous in both the occupancy of a glycan at the glycosylation site (macro-heterogenicity) and the chemical makeup of the N-, C-, or O-linked glycan (micro-heterogenicity) [15].

The most common form of glycosylation observed in glycoprotein structures is N-linked glycosylation. Initiation of this process occurs during translation, by the protein

oligosaccharyltransferase (OST), which recognizes a multi-residue consensus motif, or sequon, of NX(S/T) (where X is any residue except proline), and covalently attaches a lipid-linked core-oligosaccharide to the asparagine residue through an N-glycoside linkage¹. This process is not deterministic (not every sequon results in attachment of a glycan) and certain amino acids in and around the sequon motif can affect the efficiency of this process, resulting in higher or lower glycan occupancy at the site [16] [17].

Upon successful protein folding in the endoplasmic reticulum, the initial N-linked glycan is “trimmed down” by removal of several terminal glucosyl residues, while many sugar processing enzymes in the Golgi apparatus can add or remove sugar residues from the nascent branched sugar (tree). The resulting chemical makeup of the glycan tree depends on which enzymes are available in the Golgi, which is heavily influenced by species, disease state [18], developmental stage [19]; and the local structure, sequence, and environment of the glycosylation site [20]. In addition, a particular glycosylation site can result in vastly different glycans [21], though this can be controlled to some extent through various bioengineering techniques [15] [22] [23].

Glycans are also conformationally flexible, being highly hydrophilic and typically exposed on the surface of proteins, with a large number of conformational degrees of freedom. However, as has been observed in molecular dynamics and NMR experiments, glycan conformations can be influenced by their structural environment [24]. Through the plethora of high-resolution crystallographic and cryo-EM studies, we also know that glycans can adopt stable conformations with well-defined density observed for many of the glycan residues in each tree, especially towards the root of the glycan tree, even for some unrestrained glycans [25] [26]. Presumably, these low-energy, stable conformations are occupied at higher frequency in solution. In addition, a recent QM study on glycan torsional energies showed that the QM-derived conformational preferences of glycan torsions match well with glycan structures analyzed from the protein data bank, indicating that conformational diversity is also influenced by the chemical makeup of each glycan structure [27].

Given the complex chemistry and conformational diversity involved, accurate modeling of glycans is currently a grand challenge in computational biology. Computational glycobiology tools and webapps have been developed for protein glycosylations [28], validation of carbohydrate structural chemistry [29], statistical analysis [30], and docking [31] [32]. Common methods in glycoprotein modeling typically involve molecular dynamics (MD) simulations [33] or adding glycans by manual placement and conformational tweaking into their density for structure determination [34]. Recently, a new method for automatic building of glycan structures from sequence was described [35]; this method, the CHARM-GUI Glycan Modeler, was benchmarked only up to the first and second sugar.

Here we describe a new glycan modeling algorithm built within the Rosetta software suite, a platform that incorporates state-of-the-art applications and modules for a variety of macromolecular modeling and design tasks [36]. The new algorithm provides user interfaces for the creation of tailor-made protocols [37] [38], and includes a reliable knowledge-based energy function to evaluate models and designs [39]. We build on earlier work that enabled representing and evaluating carbohydrate structures within Rosetta [40] and in loading, representing, and refining glycans from the Protein Data Bank [41]. We expand on this foundational work through the addition of new carbohydrate-specific sampling methods, an updated conformer database employing adaptive kernel density estimates, a new framework for general analysis in Rosetta (*SimpleMetrics*), and a new algorithm for accurately modeling complex carbohydrates, the *GlycanTreeModeler*.

We rigorously benchmark the new method on a set of diverse high-resolution crystal structures of glycans in symmetric crystal environments using the new analysis framework

SimpleMetrics and a new application called *rosetta_scripts_jd3*, and we show that the *GlycanTreeModeler* is capable of recapitulating native glycan structures with high accuracy both through *de novo* and density-guided modeling [42]. We then applied our glycan modeling protocol with Rosetta sequence design of glycan sequons to engineer optimal new glycans onto a protein nanoparticle vaccine scaffold and evaluated changes in immune responses. We observed reduced reactivity to the underlying protein surface in immunization experiments, thus demonstrating that glycans can be computationally engineered to tailor immunogenicity of vaccines.

Results

Benchmarking tools

In order to examine the performance of *GlycanTreeModeler*, we built a new benchmarking infrastructure in Rosetta. We developed the *SimpleMetrics* framework within the XML interface to Rosetta (*RosettaScripts* [37]), which allows for robust analysis through more than 20 associated structural and energetic metrics, with data reporting at any step in a *RosettaScripts* protocol. To facilitate large scale benchmarking, we developed a general application for parallel *RosettaScripts* computing, *rosetta_scripts_jd3*, enabling glycan calculations to be run in parallel on a high-performance computing cluster. This application can run multiple jobs within a single parallel run of Rosetta, with individually configured glycan trees to be modeled, and any associated input files for each. The *SimpleMetric* framework and *rosetta_scripts_jd3* application are reviewed in detail in [S1 Text](#).

Glycan structure set

The Rosetta *GlycanTreeModeler* algorithm was benchmarked against a set of 25 unique N-linked glycan trees in their crystal arrangement ranging from three to twelve residues, across 19 unrelated glycoprotein structures of better than 2 Å resolution, totaling 139 sugar residues. Each glycan tree was checked for chemical and structural inconsistencies (such as incorrect isoform assignments, wrong linkages, or missing atoms) using the glycosciences.de pdb-care webservice (which filtered many of our initial glycan list) [29]. It should also be noted that some of the structures are likely substructures of larger glycans. Preparation and analysis of the structures can be found in [S1 Text](#).

De novo modeling

Using the optimized protocol and scoring function found during protocol optimization (see [methods](#)), benchmarking was done on the set of 25 glycans described above. Across the benchmark dataset, the median RMSD of the glycan predictions to the native structures was 2.7 Å, while the mean was 5 Å. For the first two residues of the glycan tree, the median was 1.28 Å with a mean of 2.17 Å. Of the 25 glycan trees, 20% of the glycans were predicted at < 1 Å accuracy and 72% (18/25) of the glycans were predicted at < 5 Å accuracy (Figs 1 and 2). The largest glycan in our dataset, with twelve residues, was benchmarked at 2.5 Å. Full results for each glycan are listed in [S1 Table](#).

It is also useful to understand how well the algorithm predicts the internal structure of the glycans, as a single dihedral angle change at the root of the glycan can significantly change the overall structure of the glycan relative to the protein. For each of these structures, the same lowest-energy models were superimposed onto the input glycan. The median superimposed RMSD is 1.1 Å, with a mean of 2.7 Å. Overall, 32% (8/25) were < 1 Å RMSD, 64% < 2.5 Å RMSD and 92% of the predictions < 5 Å. Both RMSD measurements of the glycans were generally correlated to each other ([S1 Fig](#)).

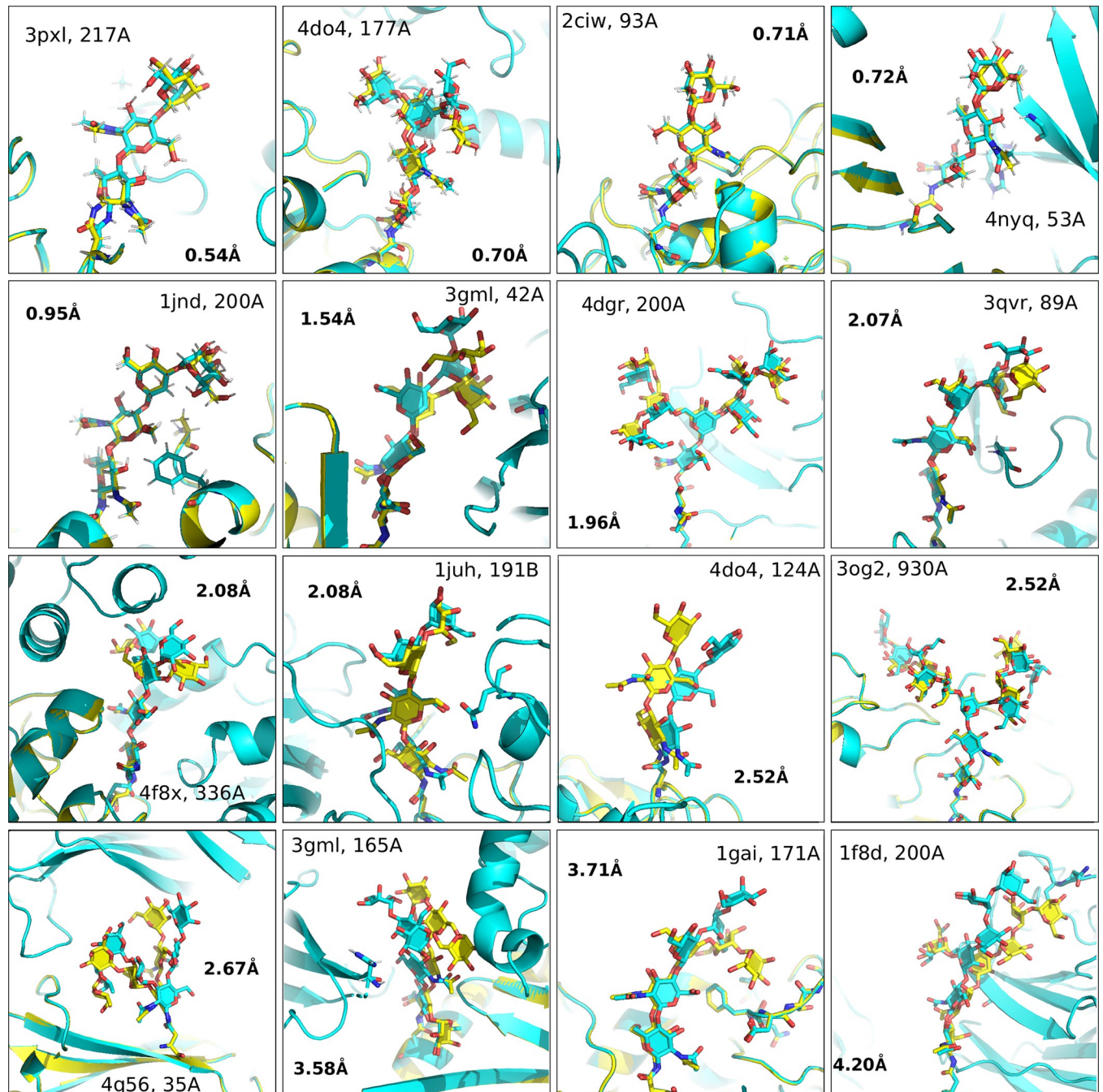


Fig 1. Near native structures from *de novo* modeling. (Top Scoring models for each glycan in the benchmark set) Yellow = Native, Cyan = Model.

<https://doi.org/10.1371/journal.pcbi.1011895.g001>

In addition, most of the glycan benchmarks in our dataset had convergent score *vs.* RMSD (funnel) plots (S2 Fig). This funnel-like quality is directly related to the ability of the scoring function to discriminate near-native models from decoys and was quantified using the PNear metric [43] that estimates the Boltzmann-weighted probability of finding a system near its native state at various near-native cutoffs (lambdas) (S1 Text). A PNear closer to 1.0 indicates the highest quality funnel possible. The worst-performing glycans in our benchmark set had poor score *vs.* RMSD funnels, indicating that the scoring function was not able to capture

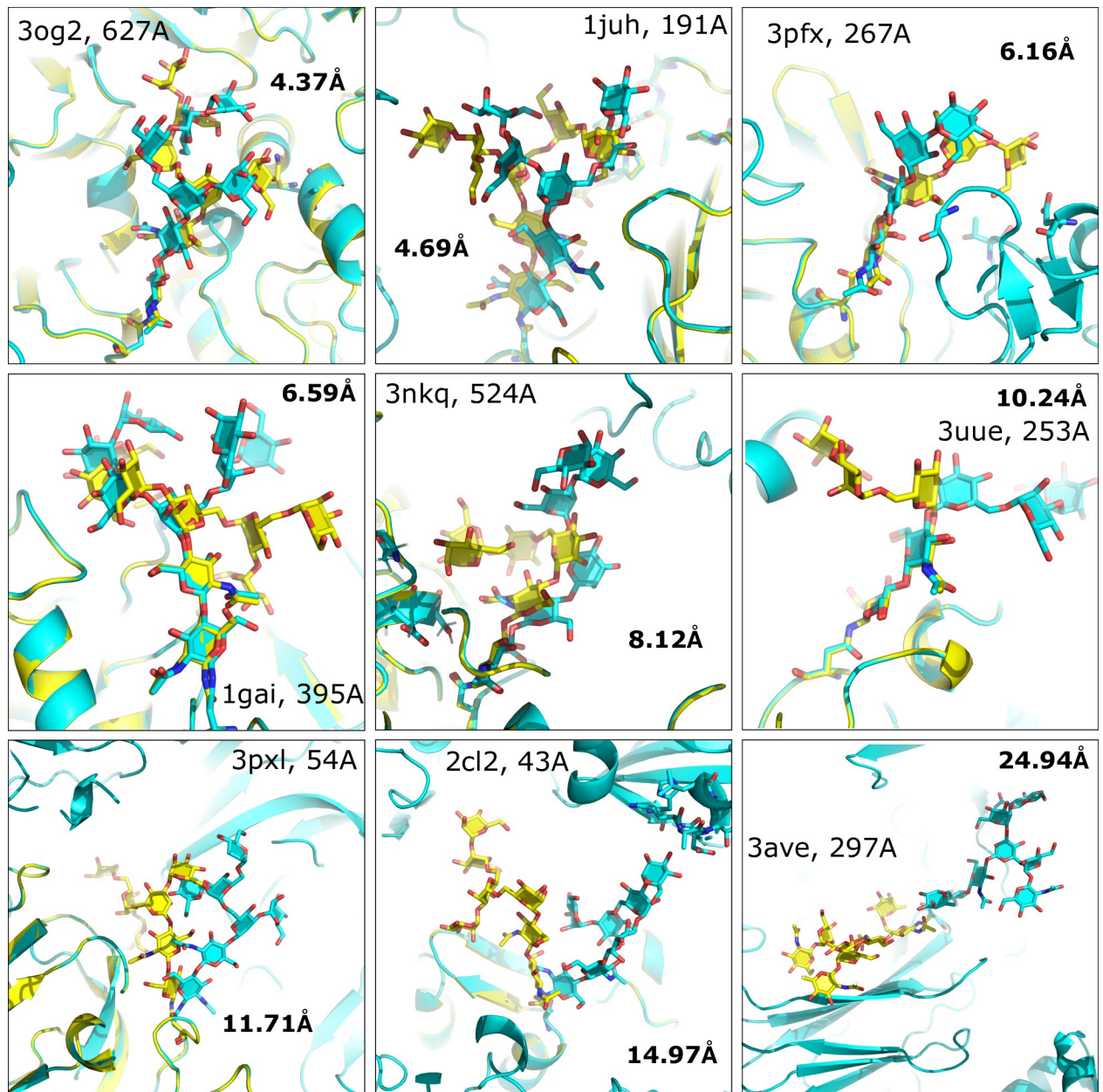


Fig 2. De novo predictions, farthest from native. (Top Scoring models for each glycan in the benchmark set) Yellow = Native, Cyan = Model

<https://doi.org/10.1371/journal.pcbi.1011895.g002>

important biophysical properties of the structure (S3 Fig). The worst-performing glycan from the Fc antibody fragment of 3ave, had an RMSD of almost 25 Å with an internal (superimposed) RMSD of 3.6 Å. In this lowest-scoring model (and others), the modeled glycan interacts with the more hydrophilic surface of a crystallographic symmetry mate rather than the more hydrophobic glycan-interacting surface of the parent protein that includes two aromatic rings (S4 Fig). This result is further detailed through the low pNear metrics of the funnel plot with all lambdas being less than .01, showing that the current energy function is unable to score these types of interactions well. However, a scoreterm that accurately represents glycan-aromatic CH- π interactions [44] may improve these results.

Solvent is implicitly represented in most Rosetta applications, but we observe that half of the benchmark glycans have significant crystallographic waters in contact with the surrounding protein. Attempting to understand the effect of waters, we modeled the worst-performing and best-performing glycans and then predicted explicit waters around the glycan for each output decoy using Rosetta-ECO [45] in order to score more native-like conformations that have these bridged waters. However, decoy discrimination as measured by pNear was significantly worse for all lambda cutoffs (even for the best-performing glycans), indicating that even with explicit waters and sufficient near-native sampling distributions, the Rosetta energy function was unable to use this information to accurately distinguish near-native decoys. (S2 Table).

In the benchmark set, the internal (superimposed) RMSDs are generally low in comparison to the overall RMSD (84% < 3 Å), showing that the energy function, guided by the QM-derived *sugar_bb* energy term, can accurately predict many glycan structures, but may need to be further improved to more accurately score glycan-protein interactions in the future.

Density building

There are an increased number of glycoprotein structures being determined. To assist structure determination, many recent glycan modeling tools have focused on their ability to aid in glycan structure building and refinement using the experimental density, especially for structures with many resolved glycans such as HIV Env. We tested the ability of the *GlycanTreeModeler* to build glycan structures using crystallographic density information to guide modeling and decoy discrimination using integrated density scoring [42]. The experiment was conducted in the same manner as *de novo* modeling, by first randomizing all backbone dihedral angles of the glycan to be modeled for each output decoy and removing all crystallographic waters. For each of the 25 glycans, the lowest-energy model was used for assessment.

Without further refinement or any additional changes to the protocol, all glycans were modeled at sub-angstrom accuracy. The best glycan in the current benchmark, with six residues, was built at 0.08 Å RMSD to native (3gml position 165A glycan), while the worst, a five-residue glycan, was modeled at 0.88 Å RMSD (1gai position 171A glycan). For both of these glycans, funnel plots were generally good, with respective PNear values of 0.99 and 0.46 at a lambda of 1.0 Å (Fig 3). For 1gai glycan 171A, the last residue in the glycan is twisted in the best model compared to the native and fits two constituent oxygens into the low residue density at a different angle than the solved structure. This twist can clearly be seen in the funnel plot where the distribution of models less than 1 Å is bimodal, indicating two primary close solutions of the electron density. (Fig 3F).

Overall, the *GlycanTreeModeler* achieved a mean heavy atom RMSD of 0.48 Å using all residues and 0.34 Å using residues that had acceptable fits into the density (133/139 total glycan residues, S1 Text). For both inclusion types, the median RMSD was 0.31 Å and 0.28 Å respectively, while the mean RMSD of the glycan root (first two sugar residues) was .23 Å (Fig 4A) (S3 Table). Values for PNear with lambda of 1.0 Å were generally quite favorable, indicating high-quality funnels, with a mean of 0.86 and median of 0.92 (Fig 4B). These results show that the *GlycanTreeModeler* can be effective for modeling known glycans into electron density, especially with existing methods refinement [41].

Sugar coating protein surfaces

The addition of glycans to exposed protein surfaces can reduce B cell receptor access to underlying surface epitopes; this approach (called “glycan masking”) has been used to decrease the amount of antibodies elicited against off-target epitopes of designed immunogens [12] [46]

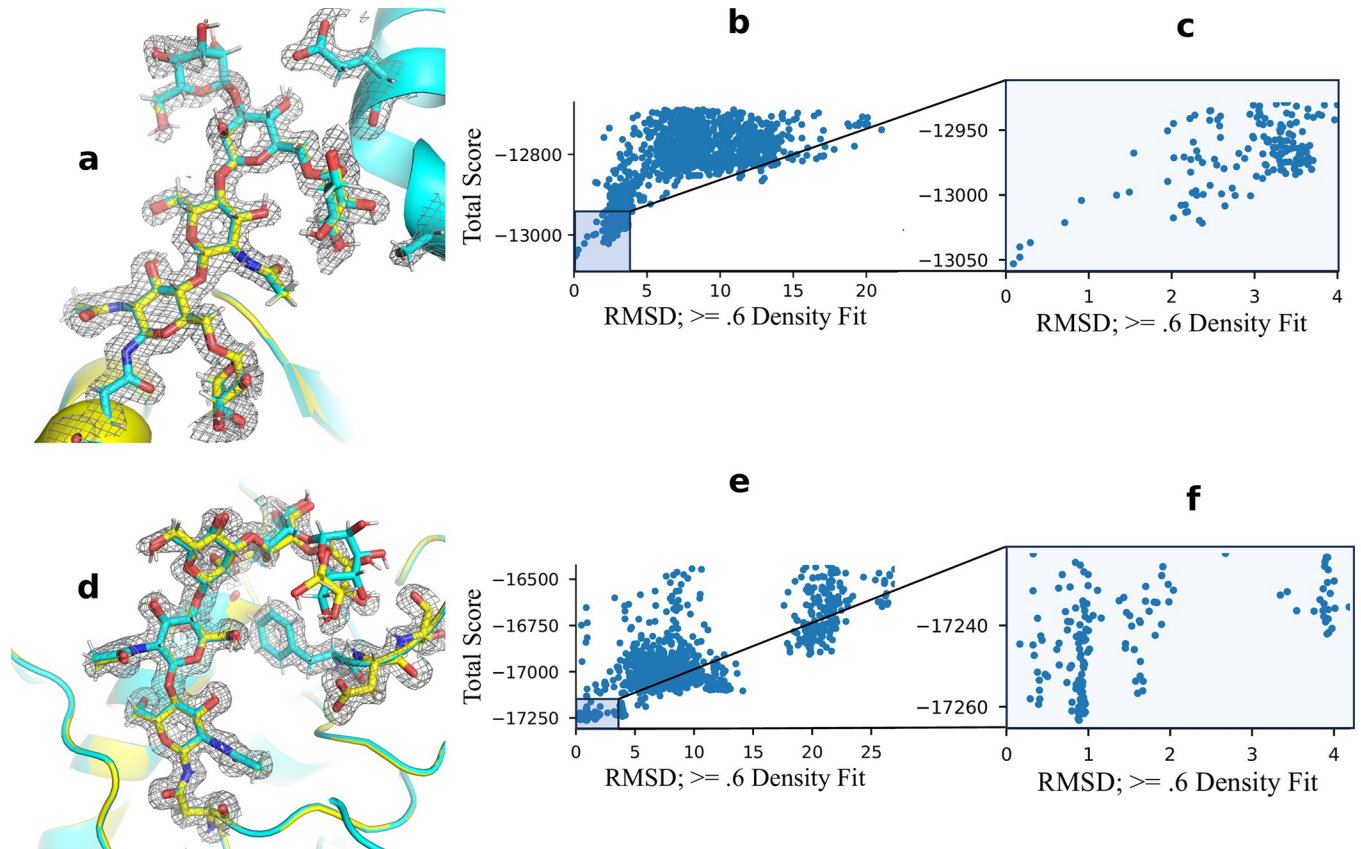


Fig 3. Best and worst results from density-guided modeling: a. Structural comparison of 3gml 165A glycan; **0.08Å RMSD**; cyan = model | yellow = native b (and e). RMSD vs. Score (funnel) plot, top 80% by energy. c (and f). Funnel plot of top 10% models with pNear metrics. d. Structural comparison of 1gai 171A glycan; **0.88Å RMSD**; cyan = model | yellow = native

<https://doi.org/10.1371/journal.pcbi.1011895.g003>

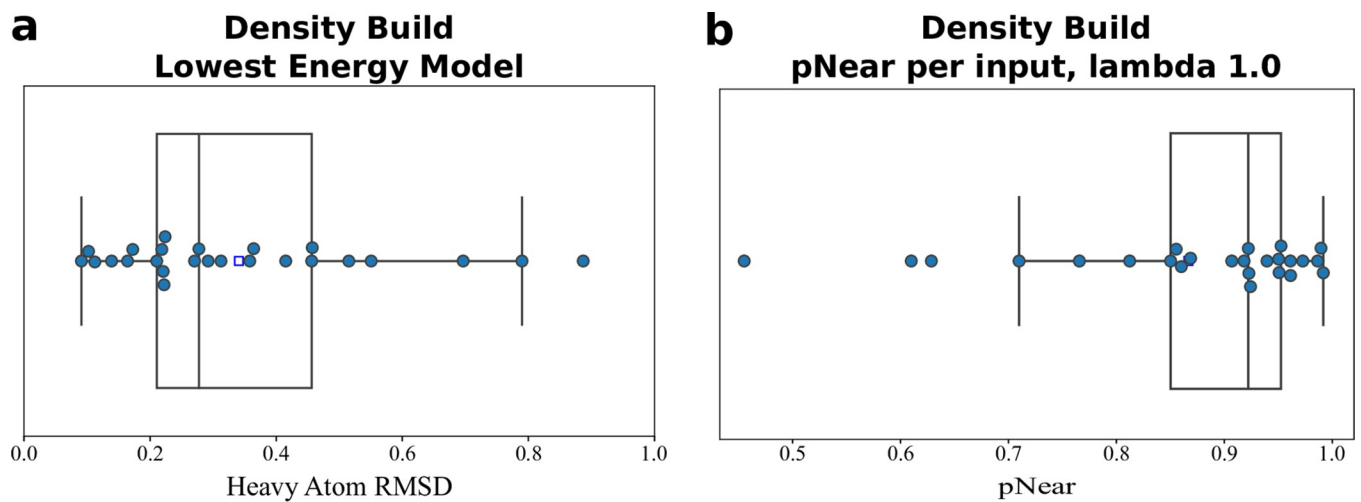


Fig 4. Density-guided modeling quality: a. Boxplot of the RMSD to native of the best-scoring decoy for each of the benchmarked input glycans. b. Boxplot of the funnel quality for each of the benchmark glycans as measured by the pNear metric. A value closer to 1.0 indicates a high-quality funnel.

<https://doi.org/10.1371/journal.pcbi.1011895.g004>

[47] [48]. Given the predictive capability of the *GlycanTreeModeler* to model the spatial arrangement of complex glycans, we used the algorithm in combination with *RosettaScript SugarCoating* methods for sequon design and computational glycosylation to iteratively design four N-linked glycans onto the outer surface of the I53-50A trimeric component of the I53-50 protein nanoparticle scaffold (Fig 5A; details of the design approach are described in *Materials and Methods* in [S1 Text](#)). Designed sequences and designed glycan positions are given in [S4 Table](#)). I53-50 was selected as a model immunogen because it is currently in clinical trials as the nanoparticle scaffold for SARS-CoV-2 [49] and RSV [50] vaccines.

When glycosylated I53-50A trimers and I53-50B pentamers were mixed *in vitro* at equimolar concentrations, the two components self-assembled into I53-50(gly) nanoparticles that display 240 glycans on the outer surface (Fig 5A and 5B). Biophysical characterization by negative stain transmission microscopy (nsTEM), dynamic light scattering (DLS), and size exclusion chromatography (SEC) confirmed the formation of monodisperse particles with the known I53-50 morphology (Fig 5B). SDS-PAGE analysis of the I53-50A(gly) trimer treated with PNGase F confirmed that the designed glycans were present in the protein (Fig 5B). Further *in vitro* characterization and antibody responses against these glycosylated I53-50A trimers has been recently described in other reports [13]. Mice were immunized three times with 5.57 μ g of I53-50 or I53-50(gly) particles. Anti-I53-50A trimer serum antibody titers were significantly lower in mice immunized with I53-50(gly) particles compared to mice immunized with I53-50 particles, whereas anti-I53-50A(gly) trimer titers were unchanged between the two groups (Figs 5C and S5). These data demonstrate that the methods presented here can be used for glycan masking through design and analysis of potential sequon motifs and the spatial arrangement of putative glycans on protein surfaces.

Discussion

The *GlycanTreeModeler* and associated tools allow modelers to accurately model glycans of interest through *de novo* and density-guided modeling. The algorithm and energy function were rigorously optimized and benchmarked with glycans of varying length and complexity at a median *de novo* RMSD of 2.7Å. In fact, even before full optimization and release, the GlycanSampler algorithm (previously the *glycan_relax app*) was used to model glycans on HIV [52], Hepatitis C [53], vaccine candidates [54] [55], and (with the final optimized version) SARS-CoV-2 [56], illustrating the general utility of the algorithm and its potential to inform chemical biology.

The modular nature of Rosetta and the tools created for this work allow them to be used in a variety of complex modeling and design tasks. The *GlycanTreeModeler* was used with previously published density tools [42] to build glycans into their crystallographic or cryoEM experimental density with sub-Angstrom accuracy. However, while the results are encouraging, a truly automated solution for glycoprotein modeling must also sample glycan chemistries, branching, and kinematics simultaneously in order to build potential glycan residues into the density of unknown glycans. Knowledge of the range of glycoforms and occupancy occurring at a glycosylation site can be obtained through mass-spectroscopy techniques [21] [57], but due to chemical and structural heterogeneity at any single glycan site, modelers will typically need to build models for multiple different glycoforms at a single site, especially for complex glycans. The tools presented here can sample and build multiple potential whole glycans at a site through the *SimpleGlycosylateMover*, but core Rosetta methods that also consider species and cell-type dependent glycan chemistries during the *GlycanTreeModeler* or end-to-end deep learning methods would be a welcome addition to the methods presented here.

By combining the tools through *RosettaScripts*, it becomes possible to computationally design glycan sequons at ideal positions on a protein, and then build and model multiple

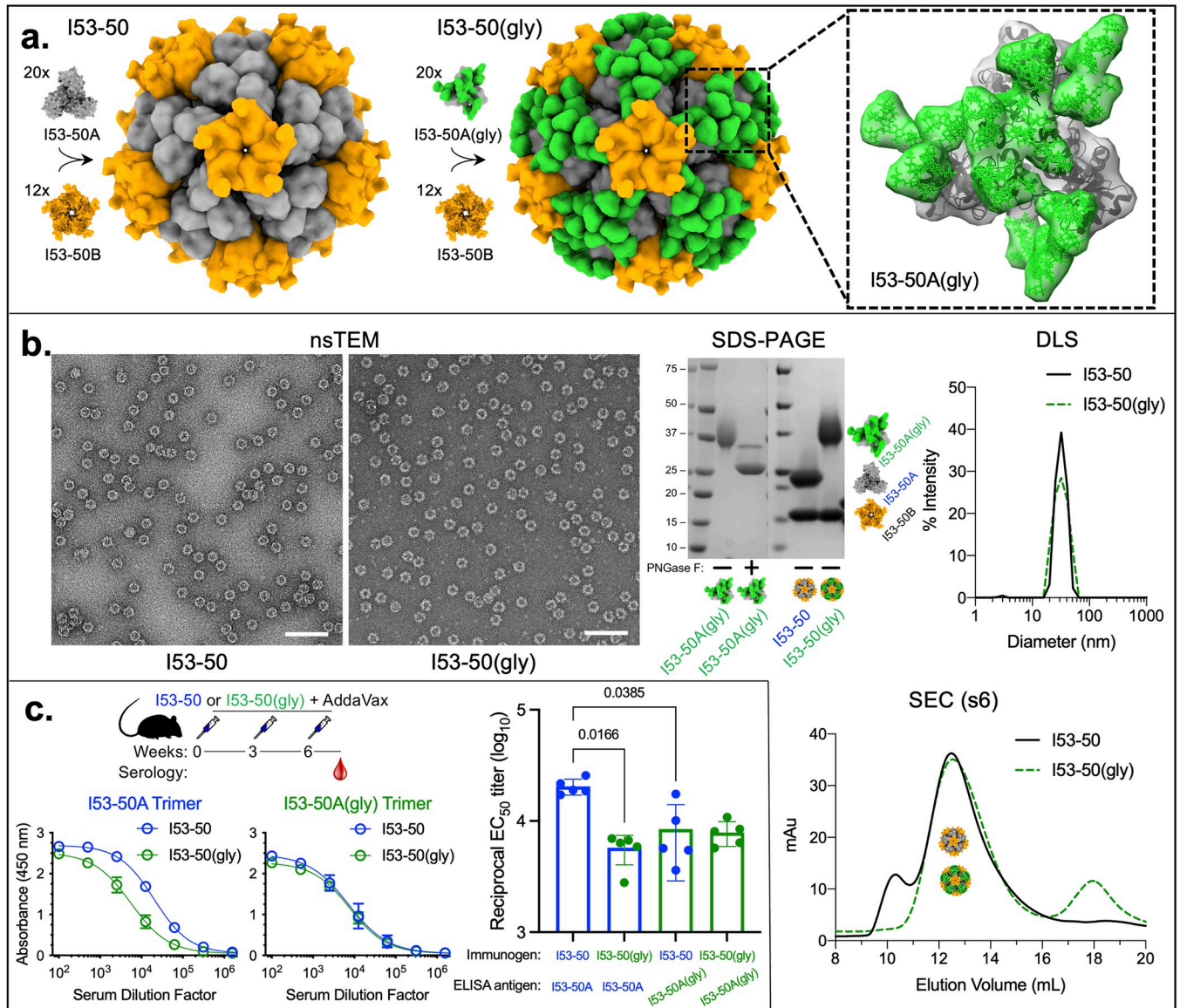


Fig 5. Characterization and reduced immunogenicity through glycosylation of the trimeric component of the I53-50 two-component protein nanoparticle scaffold that is used in clinical-stage subunit protein nanoparticle immunogens. **a.** Schematic of protein design models. On the left, twenty I53-50A trimers (gray) and twelve I53-50B pentamers (orange) self-assemble into I53-50 protein nanoparticles [51]. Rosetta sugarcoating design protocols were used to glycosylate the outer surface of I53-50A trimers with 4 N-linked glycans (green) per protomer to form I53-50 particles with 240 N-linked glycans (middle). The inset on the right is a close-up view of glycosylated I53-50A trimers with 12 total glycans on the outward-facing surface. **b.** Characterization of bare versus glycosylated I53-50 particles using negative stain transmission electron microscopy (nsTEM; scale bar, 100 nm), SDS-PAGE, dynamic light scattering (DLS), and size exclusion chromatography (SEC) on a Superose 6 Increase 10/300 GL column (GE Healthcare). In the SEC chromatogram, both I53-50 and I53-50(gly) particles reach peak elution at 12.5 mL; unassembled I53-50A and I53-50B components elute at ~18 mL. **c.** ELISA curves (left two plots) and corresponding EC₅₀ titers (right bar plot) showing reduction in anti-I53-50A antibody responses when mice were immunized with I53-50(gly) versus I53-50. BALB/c mice were immunized intramuscularly at 0, 3, and 6 weeks with 5.57 μg of I53-50 or I53-50(gly) and serum antibody binding to I53-50A trimer (left) or I53-50A(gly) trimer (right) was quantified via ELISA using 8-week sera (N = 5 mice/group). For statistical analysis, Mann-Whitney tests were used to compare among the experimental groups.

<https://doi.org/10.1371/journal.pcbi.1011895.g005>

potential glycans at a variety of sites in a symmetric manner. This general workflow was used to sugarcoat a clinically relevant nanoparticle vaccine scaffold with N-linked glycans. In vitro and in vivo testing of this glycosylated scaffold showed a decrease in the humoral immune response to the glycan-masked surface. Sugar coating therapeutics using these methods could

potentially reduce off-target effects of many preclinical biologics, especially with respect to immunogenicity.

Most glycans can sample a wide range of conformations in solution, as they are mostly polar, usually exposed to solvent, and have many conformational degrees of freedom. Thus, accurately predicting the lowest energy states (and highest occupancy conformations) for glycans is difficult. In addition, these glycans may be forced into higher-energy internal states through local and crystal contacts. While we can generalize that low energy conformations found through the *GlycanTreeModeler* should be indicative of probable solution conformations, the *GlycanTreeModeler* was not benchmarked on an experimental ensemble of glycan structures. The few glycan ensembles found through solution NMR [58] may approximate conformational ensembles in solution and could be the bases for future benchmarking and protocol/scorefunction optimization. However, even with this consideration, many of the benchmark glycans that were modeled accurately to their crystal structures are not hindered by monomer or crystal contacts, but have few interactions to protein residues in their glycan root. Additionally, predictions of the internal (superimposed) RMSDs of all glycans benchmarked were generally favorable with a median benchmarked accuracy of 1.1 Å and a mean of 2.7 Å, indicating that the glycan root, subsequent torsional preferences, and intra-glycan interactions may be determining structural factors for these isolated glycans.

Although the algorithm is capable of accurate *de novo* modeling of many glycans (especially at their base) and has been used for experimental glycan masking, there is certainly room for improvement. In nearly all of the benchmarks, the native structure is sampled adequately, but in a subset of structures, the energy function is not able to choose near-native structures. Upon further investigation of the many native glycans in the benchmark set with water-mediated hydrogen bonds, we originally hypothesized that explicit water modeling might help the energy function discriminate near-native models. However, we found that *implicit* modeling actually led to better discrimination scores through the pNear metric. In order to improve the algorithm further, the Rosetta energy function will need to be optimized to improve glycan-protein interactions, specifically in terms of hydrogen bonds, solvation, and the introduction of energy terms that better represent aromatic CH- π interactions [44]. Finally, the algorithm requires more compute time as the number of glycans to model increases, which can be prohibitive for large, multimeric glycoproteins such as HIV.

In this work, optimization of both sampling and scoring was necessary to improve overall accuracy. A key component of the algorithm is the nature-inspired kinematics used during sampling, which was shown to be an important determinant of the overall accuracy of the algorithm. The kinematics were rigorously benchmarked here, though kinematics are not always taken into account or optimized in state-of-the-art classical modeling algorithms. This benchmarking was made possible by the *SimpleMetric* framework and a new *RosettaScripts* application that were created and used continuously throughout this work. In addition, we demonstrated the usability of these methods through glycan masking the trimeric subunit of a two-component self-assembling protein nanoparticle that is used as a scaffold to multi-valency display viral glycoprotein antigens. While the glycan masking did not completely remove antibodies specific for the trimer, the experimental results did show proof-of-concept that glycan masking can significantly reduce antibody responses.

SimpleMetrics have now become a critical tool for general analysis in Rosetta and as a way to export important information for external algorithms, such as the quantum annealer [59]. As core protocols in Rosetta continue to be optimized, and as deep learning becomes a more integral aspect of modeling and design, *SimpleMetrics* should allow the robust analysis of new protocols, results, and Rosetta benchmarks, as it has for this work.

These results show that the *GlycanTreeModeler* is able to accurately predict glycan structures *de novo*, build them into known density, and be used in *SugarCoating* protein surfaces. In addition, the modular nature of the components allows them to be further developed for specific engineering tasks such as immunogenicity reduction or the optimization of developability characteristics such as half-life, solubility, and aggregation potential.

Methods

The Rosetta *GlycanTreeModeler* builds whole glycan “trees” through an algorithm that mimics the growth of natural trees. A primary difficulty in *de novo* glycan modeling is the correct prediction of the base of glycoconjugate structures. To increase the accuracy of the first few sugars of the tree, our algorithm begins modeling from the “root” (reducing end) of the glycan tree out to the branching “foliage”. Monte Carlo optimization through sampling of glycan degrees of freedom (DOFs) is carried out through the new *GlycanSampler*, which includes routines for glycosidic torsion angle (backbone) sampling, structure minimization, hydroxyl and other side-chain optimization, and neighbor protein side-chain optimization. During the protocol, the total amount of sampling scales linearly with the number of glycan residues being modeled, ensuring even sampling regardless of the size or quantity of glycans being modeled.

The *GlycanSampler* optimizes glycosidic torsion angles using statistically favorable sets of phi, psi, and omega angles (conformers) and single torsions sampled from QM-derived probabilities originally used for energetic evaluation of glycosidic linkages [27] [31]. Conformer sets are dependent on each chemically distinct pair of saccharides making up a glycosidic bond, whereas single torsions depend on the anomeric chemistry of the linkage. We derived the conformers for this work by carrying out a new bioinformatic analysis of glycans in the PDB through the use of adaptive kernel density estimates in a similar manner to what was done for the 2010 Dunbrack Backbone-dependent Rotamer Library [60] (S1 Text).

To optimize the conformations of glycan residues on different branches at the same time, the glycan tree is built layer-by-layer, with a layer defined as the residue distance to the root (Fig 6A). Once each new layer is built and optimized, all previous layers are then optimized further (Fig 6B). After all layers are built and optimized, a final optimization is conducted. The lowest energy model (decoy) found during this Monte Carlo algorithm is output at the end of the program as a PDB file. The lowest-energy structure of all the output decoys is used as the “best” model produced by the algorithm (S1 Text).

Benchmarking protocol

Benchmarking was carried out through the *SimpleMetrics* framework developed for this work. A *SimpleMetric* takes a structure and returns a metric or set of metrics, which can then be written to an output scorefile at the end of the protocol during a *RosettaScripts* execution. A number of *SimpleMetric* types were developed for textual, numeric, coupled, and per-residue data (S7 Table). These metrics enable calculation of RMSDs, Solvent-Accessible-Surface-Area (SASA), complex hydrogen bonding networks, and other biophysical properties. These metrics can also be used on-the-fly with Rosetta filters using the *SimpleMetricFilter* and simple calculations of per-residue data can be achieved using the *ResidueSummaryMetric*. Many of these metrics were used for benchmarking and analysis (S1 Text). Further, a new application, *rosetta_scripts_jd3* was created to enable large-scale benchmarking of Rosetta protocols. This application enables parallel-execution of different rosettascript protocols in parallel, with all resulting experiments tagged during score-file output. This allows for an entire experimental benchmarking pipeline to be created, run, and analyzed through a single Rosetta execution. The Python scripting language was used to load the resulting JSON scorefile for data analysis

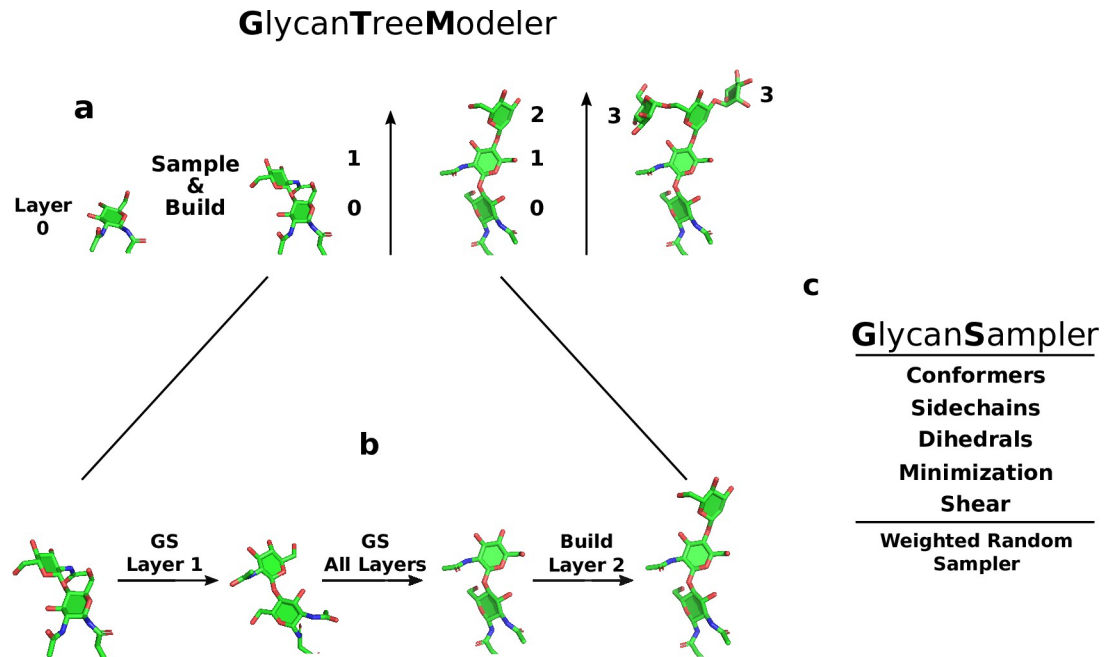


Fig 6. Glycan modeling diagram. a. Glycan trees building layer by layer. Numbers indicate distance to root of the glycan tree, which is the first residue. b. After a layer is built, Glycan Sampling is performed on the new layer, and then all layers, before building the next layer. c. Diagram showing major components of the GlycanSampler. The GS is a weighted random sampler, indicating that each DOF is sampled with a specific probability (S1 Text).

<https://doi.org/10.1371/journal.pcbi.1011895.g006>

and figure creation using the numpy [61], pandas [62], and seaborn [63] libraries. All protocol components and their availability in *RosettaScripts* is listed in S8 Table.

To assess the predictive capability of the *GlycanTreeRelax* algorithm, the dihedral angles of the glycans are randomized at the start of the algorithm, and waters are removed. Models are compared to the crystal structures using the all-heavy-atom Root Mean Square Deviation (RMSD) metric, with the lowest energy model of all output decoys used for assessment (Fig 7). The RMSD is calculated on all glycan residues that have an acceptable fit to the density in the native model, as terminal glycan residues of some glycans often cannot be observed in the density due to their higher flexibility. A description of the methods used for the RMSD calculation is provided in S1 Text.

Glycan masking

Glycan masking was carried out through the use of two new *RosettaScript* components; the *CreateGlycanSequonMover*, which designs typical and enhanced [64] [17], glycan sequons into a protein at a desired position, and the *SimpleGlycosylateMover*, which adds whole glycans of a

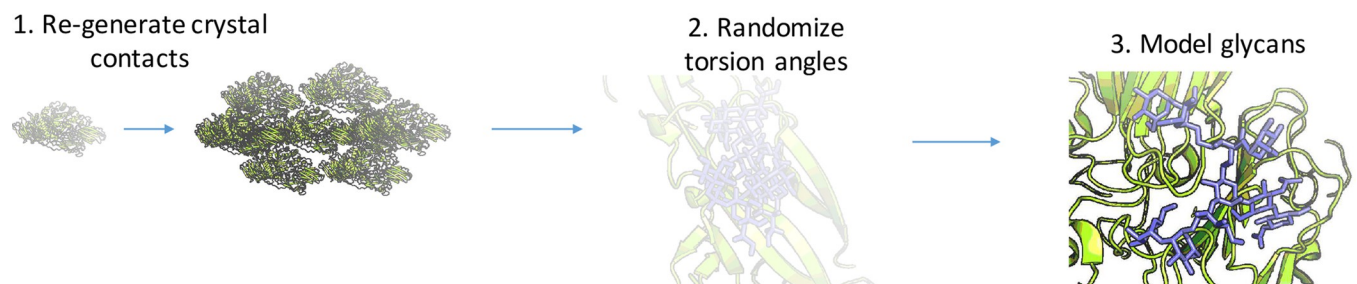


Fig 7. Schematic of benchmarking protocol.

<https://doi.org/10.1371/journal.pcbi.1011895.g007>

given IUPAC onto a protein. Glycans were then sampled using the *GlycanTreeModeler* through *RosettaScripts* at each potential glycan position individually. Low-energy and non-clashing models were used to select optimal positions for experimental validation with sequon sequences designed for each position using the *CreateGlycanSequonMover* (*S1 Text*).

Availability and Documentation

The *GlycanTreeModeler*, *GlycanSampler*, and all tools used in this work are available in the Rosetta Software Suite, which is free for non-commercial use. All tools are available as components for *RosettaScripts* and PyRosetta. In addition, the use of all core components are covered in publicly accessible tutorials [65] and detailed protocol captures [66]. Results of this study are continuously benchmarked using the Rosetta automated scientific testing framework [67].

Figures

Figures were created using matplotlib [68]. Glycans were visualized in PyMol using the Azahar plugin [69], which was expanded for this work. The cartoonize command was generally run for figures (cartoonize A) for chain A: <https://github.com/BIOS-IMASL/Azahar/pull/17>

Documentation Links:

- **RosettaScripts:**
 - https://www.rosettacommons.org/docs/latest/scripting_documentation/RosettaScripts/RosettaScripts
- **Working with Glycans:**
 - https://www.rosettacommons.org/docs/latest/application_documentation/carbohydrates/WorkingWithGlycans
- **Chapter 13 of the PyRosetta Notebooks:**
 - <https://github.com/RosettaCommons/PyRosetta.notebooks>

Supporting information

S1 Text. Detailed Supplemental Methods.
(DOCX)

S1 Table. Raw *de novo* Modeling results for each glycan tree.
(XLSX)

S2 Table. Rosetta-ICO vs Rosetta-ECO mean pNear values at various lambdas. N = 8.
(XLSX)

S3 Table. Density-guided modeling results for each glycan tree.
(XLSX)

S4 Table. Amino acid sequences of self-assembling nanoparticle components.
(XLSX)

S5 Table. GlycanSampler Components and Probabilities.
(XLSX)

S6 Table. Structures used for this work.
(XLSX)

S7 Table. Initial SimpleMetrics created and used in this work.

(XLSX)

S8 Table. *RosettaCarbohydrate* and General Extensions (RS indicates accessibility in Rosetta Scripts).

(XLSX)

S1 Fig. Superimposed RMSD comparisons of the top scoring model for each de novo modeled glycan tree.

(TIF)

S2 Fig. Score vs. RMSD funnel plots of the best predicted glycan structures with pNear at different lambda values. Shown is the top 10% of models by total energy. Blue line is the scored native structure with symmetry.

(TIF)

S3 Fig. Score vs. RMSD funnel plots of the worst predicted glycan structures with pNear at different lambda values. Shown is the top 10% of models by total energy. Blue line is the scored native structure with symmetry.

(TIF)

S4 Fig. Hydrophobic surface interactions with 3ave glycan at residue 297, chain A. F241, 243F, 262V, and 264V are shown as spheres at the glycan interface.

(TIF)

S5 Fig. Individual ELISA curves for 8-week sera from mice immunized three times with I53-50 or I53-50(gly) in the presence of AddaVax adjuvant (related to Fig 5C). (a,b) anti-I53-50A trimer or (c,d) anti-I53-50(gly) trimer antibody responses from mice immunized with (a,c) I53-50 nanoparticles (NP) or (b,d) I53-50(gly) NP.

(TIF)

S6 Fig. Density fit example.

(TIF)

S7 Fig. Probability density function of a torsion angle calculated with adaptive kernel density estimation. The von Mises kernel allows for continuous circular description of the torsion angle distribution. The experimental angles from a sample are shown with small dots at the bottom. Such 1-D density estimates were performed for each torsion comprising a glycan-glycan or amino-acid-glycan linkage type. These 64 linkage types can be found in the resulting conformer table included in [S1 Data](#).

(TIF)

S8 Fig. Kinematic sampling optimization, decoy enrichment from each individual experiment. All experiments were conducted with the same total amount of sampling. **a.** Boxplots at decoy enrichments of <1A, <2.5A, and <5.0A. First figure has mean only since most are grouped at zero. **b.** decoy enrichments of <1A, <2.5A, and <5.0A. Asterisks indicate statistically significant differences through paired t-test. Asterisk above bar indicate statistical significance with all other groups. *, $p < .05$; **, $p < .005$; ***, $p < .0005$.

(TIF)

S9 Fig. Enrichments of GlycanSampler(GS) alone compared to a successive algorithm of GlycanTreeModeler and then GlycanSampler (hybrid-GS).

(TIF)

S10 Fig. Hybrid Enrichments compared to Hybrid building two layers at a time.
(TIF)

S11 Fig. Overall pool of models at varying filters of total score. N = 150,000; 37,500 per experiment.
(TIF)

S12 Fig. Kinematic experiments—enrichment KDEs. **a.** Kernel Density Estimates of enrichment per input model for each major kinematic experiment. **b.** Box plot of enrichment of each input model per experiment less than 7.5 Å RMSD to the native crystal structure. **c.** Means of B, with paired t-test, All vs. All. * indicates $p < .05$. p-value for hybrid-build-one vs. build-by-layer $p < .005$, while vs. all-sampler $p < .0005$.
(TIF)

S13 Fig. Kinematic experiments—decoy enrichment of STEM region (Layers 0 and 1). **a.** Boxplots of each input glycan benchmark at $<1.0\text{Å}$, $<2.5\text{Å}$, and $<5.0\text{Å}$ of the glycan STEM **b.** Means of panel a. Asterisk above bar indicate statistical significance with all other groups through paired t-test. * $|p < .05$; ** $|p < .005$; *** $|p < .0005$. For **b** $<1\text{Å}$, pvalue of all-sampler vs. hybrid-build-one is **.
(TIF)

S14 Fig. Scoring experiments—decoy discrimination of various values of sugar_bb energy term. 1000 decoys were produced for each glycan and each experiment for a total of 125k decoys. Note that this is a third less than all other optimization experiments. Boxplots and bar-plots of pNear metric at each significant lambda are shown. Blue squares indicate mean. Line in box indicates median. Upper left figure shows only means as most datapoints are grouped at 0 and the box could not be seen. Paired T-test results between *sugar_bb* weight of 1.0 and .5 are indicated.
(TIF)

S15 Fig. Funnel plot quality of scoring benchmarks assessed by the pNear metric. **a.** Box-plot of pNear values for each benchmark glycan, indicating funnel plot quality for lambdas of 1.0, 2.5, and 5.0 RMSD to native. Higher pNear indicates better near-native discrimination from other decoys. Blue squares indicate mean. **b.** Means of pNear over each experiment. Significance from paired t-test; * indicates $p < .05$.
(TIF)

S16 Fig. Scoring optimization—decoy enrichments of each experiment. Asterisk above bar indicate statistical significance with all other groups through paired t-test. * $|p < .05$ ** $|p < .005$ *** $|p < .0005$ **a.** Decoy Enrichment in output models at $<1.0\text{Å}$, $<2.5\text{Å}$, and $<5.0\text{Å}$ RMSD. **b.** Decoy Enrichment in output models of the base (STEM) region indicating layers 0 and 1.
(TIF)

S17 Fig. Boxplot of the lowest energy model for all major scoring experiments across all benchmark glycans.
(TIF)

S1 Data. Associated data and scripts, including the glycan conformer table. Please see README for a full description of each component. Briefly, these include the following scripts: *create_substituted_JD.py*—Used to create substituted job description file for rosetta_scripts_jd3; *create_symmetry_mates.py*—This was used to create symmetry mates with a 12 Å radius; *density_build_example.xml*—An an example of a main job definition file before being

substituted for each glycan; *density_build_example_substituted.xml*—Substituted version; *glycan_bm_plots.py*—Tools used for plotting all benchmark experiments from Rosetta JSON files.; *glycan_conformer_table.txt*—The resultant glycan conformer table used in Rosetta and benchmarked against; *glycan_tree_relax.xml*—Main script used for de novo benchmarking.; *glycan_tree_relax_dens.xml*—Main script used for density building.; *glycan_tree_relax_with_solvation.xml*—Script used for running the GTM with explicit waters; *glycan_water_mediated_hbonds.xml*—Script for calculating water mediated hydrogen bonds; *pdb_roots_density_sym.txt*—Benchmark glycan list used for *create_substituted_JD.py*; *refine_with_map_dih2.xml*—Script used for refining input structures into the rosetta energy function using density. (ZIP)

S1 Movie. Example of glycan modeling for a single decoy modeling a single glycan tree of 9 residues where the layer size is set to one.

(MOV)

Acknowledgments

We gratefully acknowledge Rashmi Ravichandran for providing I53-50B pentamer, Deleah Pettie and Michael Murphy for assistance with expression of I53-50A(gly) design models, Alex Roederer for preparation of I53-50 and I53-50(gly) particles for mouse immunization, and Minh N Pham for performing mouse immunizations and blood draws.

Author Contributions

Conceptualization: Jared Adolf-Bryfogle, Jason W. Labonte, John C. Kraft, Sebastian Raemisch, Jesper Pallesen, Neil P. King, Daniel W. Kulp, William R. Schief.

Data curation: Jared Adolf-Bryfogle, John C. Kraft, Maxim Shapovalov, Thomas Lütteke.

Formal analysis: Jared Adolf-Bryfogle, John C. Kraft, Maxim Shapovalov.

Funding acquisition: Neil P. King, Jeffrey J. Gray, Daniel W. Kulp, William R. Schief.

Investigation: Jared Adolf-Bryfogle, Jason W. Labonte, John C. Kraft, Sebastian Raemisch, Frank DiMaio, Daniel W. Kulp, William R. Schief.

Methodology: Jared Adolf-Bryfogle, Jason W. Labonte, John C. Kraft, Sebastian Raemisch, Thomas Lütteke, Daniel W. Kulp, William R. Schief.

Project administration: Jared Adolf-Bryfogle, John C. Kraft, Frank DiMaio, Christopher D. Bahl, Neil P. King, Jeffrey J. Gray, Daniel W. Kulp, William R. Schief.

Resources: Neil P. King, Jeffrey J. Gray, Daniel W. Kulp, William R. Schief.

Software: Jared Adolf-Bryfogle, Jason W. Labonte, Maxim Shapovalov, Sebastian Raemisch, Thomas Lütteke, Jeffrey J. Gray.

Supervision: Frank DiMaio, Christopher D. Bahl, Daniel W. Kulp, William R. Schief.

Validation: Jared Adolf-Bryfogle, John C. Kraft, Jesper Pallesen, Neil P. King.

Visualization: Jared Adolf-Bryfogle, John C. Kraft.

Writing – original draft: Jared Adolf-Bryfogle, Jason W. Labonte, John C. Kraft, Maxim Shapovalov, Frank DiMaio, Christopher D. Bahl, Neil P. King, Jeffrey J. Gray, Daniel W. Kulp, William R. Schief.

Writing – review & editing: Jared Adolf-Bryfogle, Jason W. Labonte, John C. Kraft, Maxim Shapovalov, Frank DiMaio, Christopher D. Bahl, Neil P. King, Jeffrey J. Gray, Daniel W. Kulp, William R. Schief.

References

1. Ernst B, Hart GW, Sinaý P. Carbohydrates in Chemistry and Biology [Internet]. 1st ed. Wiley; 2000 [cited 2020 Dec 18]. Available from: <https://onlinelibrary.wiley.com/doi/book/10.1002/9783527618255>.
2. Ploegh HL. Viral Strategies of Immune Evasion. *Science*. 1998 Apr 10; 280(5361):248–53. <https://doi.org/10.1126/science.280.5361.248> PMID: 9535648
3. Pejchal R, Doores KJ, Walker LM, Khayat R, Huang PS, Wang SK, et al. A Potent and Broad Neutralizing Antibody Recognizes and Penetrates the HIV Glycan Shield. *Science*. 2011 Nov 25; 334(6059):1097–103. <https://doi.org/10.1126/science.1213256> PMID: 21998254
4. Julien JP, Sok D, Khayat R, Lee JH, Doores KJ, Walker LM, et al. Broadly Neutralizing Antibody PGT121 Allosterically Modulates CD4 Binding via Recognition of the HIV-1 gp120 V3 Base and Multiple Surrounding Glycans. Trkola A, editor. *PLoS Pathog*. 2013 May 2; 9(5):e1003342. <https://doi.org/10.1371/journal.ppat.1003342> PMID: 23658524
5. Falkowska E, Le KM, Ramos A, Doores KJ, Lee JH, Blattner C, et al. Broadly Neutralizing HIV Antibodies Define a Glycan-Dependent Epitope on the Prefusion Conformation of gp41 on Cleaved Envelope Trimers. *Immunity*. 2014 May; 40(5):657–68. <https://doi.org/10.1016/j.immuni.2014.04.009> PMID: 24768347
6. Wells L. Glycosylation of Nucleocytoplasmic Proteins: Signal Transduction and O-GlcNAc. *Science*. 2001 Mar 23; 291(5512):2376–8. <https://doi.org/10.1126/science.1058714> PMID: 11269319
7. Jennewein MF, Alter G. The Immunoregulatory Roles of Antibody Glycosylation. *Trends in Immunology*. 2017 May; 38(5):358–72. <https://doi.org/10.1016/j.it.2017.02.004> PMID: 28385520
8. Irvine EB, Alter G. Understanding the role of antibody glycosylation through the lens of severe viral and bacterial diseases. *Glycobiology*. 2020 Mar 20; 30(4):241–53. <https://doi.org/10.1093/glycob/cwaa018> PMID: 32103252
9. Shental-Bechor D, Levy Y. Effect of glycosylation on protein folding: A close look at thermodynamic stabilization. *Proceedings of the National Academy of Sciences*. 2008 Jun 17; 105(24):8256–61.
10. Sinclair AM, Elliott S. Glycoengineering: The effect of glycosylation on the properties of therapeutic proteins. *Journal of Pharmaceutical Sciences*. 2005 Aug; 94(8):1626–35. <https://doi.org/10.1002/jps.20319> PMID: 15959882
11. Wang W, Nema S, Teagarden D. Protein aggregation—Pathways and influencing factors. *International Journal of Pharmaceutics*. 2010 May; 390(2):89–99. <https://doi.org/10.1016/j.ijpharm.2010.02.025> PMID: 20188160
12. Duan H, Chen X, Boyington JC, Cheng C, Zhang Y, Jafari AJ, et al. Glycan Masking Focuses Immune Responses to the HIV-1 CD4-Binding Site and Enhances Elicitation of VRC01-Class Precursor Antibodies. *Immunity*. 2018 Aug; 49(2):301–311.e5.
13. Kraft JC, Pham MN, Shehata L, Brinkkemper M, Boyoglu-Barnum S, Sprouse KR, et al. Antigen- and scaffold-specific antibody responses to protein nanoparticle immunogens. *Cell Rep Med*. 2022 Oct 18; 3(10):100780. <https://doi.org/10.1016/j.xcrm.2022.100780> PMID: 36206752
14. Read BJ, Won L, Kraft JC, Sappington I, Aung A, Wu S, et al. Mannose-binding lectin and complement mediate follicular localization and enhanced immunogenicity of diverse protein nanoparticle immunogens. *Cell Rep*. 2022 Jan 11; 38(2):110217. <https://doi.org/10.1016/j.celrep.2021.110217> PMID: 35021101
15. Rini JM, Esko JD. Glycosyltransferases and Glycan-Processing Enzymes. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al., editors. *Essentials of Glycobiology* [Internet]. 3rd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2015 [cited 2020 Dec 18]. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK453021/>.
16. Rao RSP, Wollenweber B. Do N-glycoproteins have preference for specific sequons? *Bioinformatics*. 2010 Nov 1; 5(5):208–12. <https://doi.org/10.6026/97320630005208> PMID: 21364799
17. Huang YW, Yang HI, Wu YT, Hsu TL, Lin TW, Kelly JW, et al. Residues Comprising the Enhanced Aromatic Sequon Influence Protein N-Glycosylation Efficiency. *J Am Chem Soc*. 2017 Sep 20; 139(37):12947–55. <https://doi.org/10.1021/jacs.7b03868> PMID: 28820257
18. Adamczyk B, Tharmalingam T, Rudd PM. Glycans as cancer biomarkers. *Biochimica et Biophysica Acta (BBA)—General Subjects*. 2012 Sep; 1820(9):1347–53. <https://doi.org/10.1016/j.bbagen.2011.12.001> PMID: 22178561

19. Haltiwanger RS, Lowe JB. Role of Glycosylation in Development. *Annu Rev Biochem.* 2004 Jun; 73(1):491–537. <https://doi.org/10.1146/annurev.biochem.73.011303.074043> PMID: 15189151
20. Suga A, Nagae M, Yamaguchi Y. Analysis of protein landscapes around N-glycosylation sites from the PDB repository for understanding the structural basis of N-glycoprotein processing and maturation. *Glycobiology.* 2018 Oct 1; 28(10):774–85. <https://doi.org/10.1093/glycob/cwy059> PMID: 29931153
21. Riley NM, Hebert AS, Westphall MS, Coon JJ. Capturing site-specific heterogeneity with large-scale N-glycoproteome analysis. *Nat Commun.* 2019 Dec; 10(1):1311. <https://doi.org/10.1038/s41467-019-09222-w> PMID: 30899004
22. Ren WW, Jin ZC, Dong W, Kitajima T, Gao XD, Fujita M. Glycoengineering of HEK293 cells to produce high-mannose-type N-glycan structures. *The Journal of Biochemistry.* 2019 Sep 1; 166(3):245–58. <https://doi.org/10.1093/jb/mvz032> PMID: 31102532
23. Dalziel M, Crispin M, Scanlan CN, Zitzmann N, Dwek RA. Emerging Principles for the Therapeutic Exploitation of Glycosylation. *Science.* 2014 Jan 3; 343(6166):1235681. <https://doi.org/10.1126/science.1235681> PMID: 24385630
24. Woods RJ. Predicting the Structures of Glycans, Glycoproteins, and Their Complexes. *Chem Rev.* 2018 Sep 12; 118(17):8005–24.
25. Lee JH, Ozorowski G, Ward AB. Cryo-EM structure of a native, fully glycosylated, cleaved HIV-1 envelope trimer. *Science.* 2016 Mar 4; 351(6277):1043–8. <https://doi.org/10.1126/science.aad2450> PMID: 26941313
26. Pallesen J, Murin CD, de Val N, Cottrell CA, Hastie KM, Turner HL, et al. Structures of Ebola virus GP and sGP in complex with therapeutic antibodies. *Nat Microbiol.* 2016 Sep; 1(9):16128. <https://doi.org/10.1038/nmicrobiol.2016.128> PMID: 27562261
27. Nivedha AK, Makeneni S, Foley BL, Tessier MB, Woods RJ. Importance of ligand conformational energies in carbohydrate docking: Sorting the wheat from the chaff. *J Comput Chem.* 2014 Mar 15; 35(7):526–39. <https://doi.org/10.1002/jcc.23517> PMID: 24375430
28. Bohne-Lang A, von der Lieth CW. GlyProt: in silico glycosylation of proteins. *Nucleic Acids Research.* 2005 Jul 1; 33(Web Server):W214–9. <https://doi.org/10.1093/nar/gki385> PMID: 15980456
29. Lütteke T. pdb-care (PDB Carbohydrate REsidue check): a program to support annotation of complex carbohydrate structures in PDB files. *BMC Bioinformatics.* 2004;6.
30. Frank M, Lutteke T, von der Lieth CW. GlycoMapsDB: a database of the accessible conformational space of glycosidic linkages. *Nucleic Acids Research.* 2007 Jan 3; 35(Database):287–90. <https://doi.org/10.1093/nar/gkl907> PMID: 17202175
31. Nivedha AK, Thieker DF, Makeneni S, Hu H, Woods RJ. Vina-Carb: Improving Glycosidic Angles during Carbohydrate Docking. *J Chem Theory Comput.* 2016 Feb 9; 12(2):892–901. <https://doi.org/10.1021/acs.jctc.5b00834> PMID: 26744922
32. Nance ML, Labonte JW, Adolf-Bryfogle J, Gray JJ. Development and Evaluation of GlycanDock: A Protein–Glycoligand Docking Refinement Algorithm in Rosetta. *J Phys Chem B.* 2021 Jun 16; acs.jpcc.1c00910. <https://doi.org/10.1021/acs.jpcc.1c00910> PMID: 34133179
33. Kirschner KN, Yongye AB, Tschampel SM, González-Outeiriño J, Daniels CR, Foley BL, et al. GLY-CAM06: A generalizable biomolecular force field. *Carbohydrates: GLYCAM06.* *J Comput Chem.* 2008 Mar; 29(4):622–55.
34. Emsley P, Crispin M. Structural analysis of glycoproteins: building N-linked glycans with *Coot*. *Acta Crystallogr D Struct Biol.* 2018 Apr 1; 74(4):256–63.
35. Park SJ, Lee J, Qi Y, Kern NR, Lee HS, Jo S, et al. CHARMM-GUI *Glycan Modeler* for modeling and simulation of carbohydrates and glycoconjugates. *Glycobiology.* 2019 Apr 1; 29(4):320–31.
36. Leman JK, Weitzner BD, Lewis SM, Adolf-Bryfogle J, Alam N, Alford RF, et al. Macromolecular modeling and design in Rosetta: recent methods and frameworks. *Nat Methods.* 2020 Jul; 17(7):665–80. <https://doi.org/10.1038/s41592-020-0848-2> PMID: 32483333
37. Fleishman SJ, Leaver-Fay A, Corn JE, Strauch EM, Khare SD, Koga N, et al. RosettaScripts: A Scripting Language Interface to the Rosetta Macromolecular Modeling Suite. Uversky VN, editor. *PLoS ONE.* 2011 Jun 24; 6(6):e20161. <https://doi.org/10.1371/journal.pone.0020161> PMID: 21731610
38. Gray JJ, Chaudhury S, Lyskov S. The PyRosetta Interactive Platform for Protein Structure Prediction and Design. 2009 [cited 2014 Mar 26]; Available from: http://graylab.jhu.edu/~sid/pyrosetta/downloads/documentation/PyRosetta_Textbook.pdf.
39. Alford RF, Leaver-Fay A, Jeliazkov JR, O’Meara MJ, DiMaio FP, Park H, et al. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J Chem Theory Comput.* 2017 Jun 13; 13(6):3031–48. <https://doi.org/10.1021/acs.jctc.7b00125> PMID: 28430426

40. Labonte JW, Adolf-Bryfogle J, Schief WR, Gray JJ. Residue-centric modeling and design of saccharide and glycoconjugate structures. *J Comput Chem*. 2017 Feb 15; 38(5):276–87. <https://doi.org/10.1002/jcc.24679> PMID: 27900782
41. Frenz B, Rämisch S, Borst AJ, Walls AC, Adolf-Bryfogle J, Schief WR, et al. Automatically Fixing Errors in Glycoprotein Structures with Rosetta. *Structure*. 2019 Jan; 27(1):134–139.e3. <https://doi.org/10.1016/j.str.2018.09.006> PMID: 30344107
42. DiMaio F, Tyka MD, Baker ML, Chiu W, Baker D. Refinement of Protein Structures into Low-Resolution Density Maps Using Rosetta. *Journal of Molecular Biology*. 2009 Sep; 392(1):181–90. <https://doi.org/10.1016/j.jmb.2009.07.008> PMID: 19596339
43. Bhardwaj G, Mulligan VK, Bahl CD, Gilmore JM, Harvey PJ, Cheneval O, et al. Accurate de novo design of hyperstable constrained peptides. *Nature*. 2016 Oct; 538(7625):329–35. <https://doi.org/10.1038/nature19791> PMID: 27626386
44. Hudson KL, Bartlett GJ, Diehl RC, Agirre J, Gallagher T, Kiessling LL, et al. Carbohydrate–Aromatic Interactions in Proteins. *J Am Chem Soc*. 2015 Dec 9; 137(48):15152–60. <https://doi.org/10.1021/jacs.5b08424> PMID: 26561965
45. Pavlovicz RE, Park H, DiMaio F. Efficient consideration of coordinated water molecules improves computational protein-protein and protein-ligand docking discrimination. Wallner B, editor. *PLoS Comput Biol*. 2020 Sep 21; 16(9):e1008103. <https://doi.org/10.1371/journal.pcbi.1008103> PMID: 32956350
46. Ahmed FK, Clark BE, Burton DR, Pantophlet R. An engineered mutant of HIV-1 gp120 formulated with adjuvant Quil A promotes elicitation of antibody responses overlapping the CD4-binding site. *Vaccine*. 2012 Jan; 30(5):922–30. <https://doi.org/10.1016/j.vaccine.2011.11.089> PMID: 22142583
47. Lin SC, Liu WC, Jan JT, Wu SC. Glycan Masking of Hemagglutinin for Adenovirus Vector and Recombinant Protein Immunizations Elicits Broadly Neutralizing Antibodies against H5N1 Avian Influenza Viruses. Kang SM, editor. *PLoS ONE*. 2014 Mar 26; 9(3):e92822.
48. Garrity RR, Rimmelzwaan G, Minassian A, Tsai WP, Lin G, de Jong JJ, et al. Refocusing neutralizing antibody response by targeted dampening of an immunodominant epitope. *J Immunol*. 1997 Jul 1; 159(1):279–89. PMID: 9200464
49. Walls AC, Fiala B, Schäfer A, Wrenn S, Pham MN, Murphy M, et al. Elicitation of Potent Neutralizing Antibody Responses by Designed Protein Nanoparticle Vaccines for SARS-CoV-2. *Cell*. 2020 Nov 25; 183(5):1367–1382.e17. <https://doi.org/10.1016/j.cell.2020.10.043> PMID: 33160446
50. Marcandalli J, Fiala B, Ols S, Perotti M, de van der Schueren W, Snijder J, et al. Induction of Potent Neutralizing Antibody Responses by a Designed Protein Nanoparticle Vaccine for Respiratory Syncytial Virus. *Cell*. 2019 Mar 7; 176(6):1420–1431.e17. <https://doi.org/10.1016/j.cell.2019.01.046> PMID: 30849373
51. Bale JB, Gonen S, Liu Y, Sheffler W, Ellis D, Thomas C, et al. Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science*. 2016 Jul 22; 353(6297):389–94. <https://doi.org/10.1126/science.aaf8818> PMID: 27463675
52. Ringe RP, Cruz Portillo VM, Dosenovic P, Ketas TJ, Ozorowski G, Nogal B, et al. Neutralizing Antibody Induction by HIV-1 Envelope Glycoprotein SOSIP Trimers on Iron Oxide Nanoparticles May Be Impaired by Mannose Binding Lectin. Silvestri G, editor. *J Virol*. 2019 Dec 18; 94(6):e01883–19. <https://doi.org/10.1128/JVI.01883-19>
53. Urbanowicz RA, Wang R, Schiel JE, Keck Z yong, Kerzic MC, Lau P, et al. Antigenicity and Immunogenicity of Differentially Glycosylated Hepatitis C Virus E2 Envelope Proteins Expressed in Mammalian and Insect Cells. James Ou JH, editor. *J Virol*. 2019 Jan 16; 93(7):e01403–18. <https://doi.org/10.1128/JVI.01403-18>
54. Havenar-Daughton C, Sarkar A, Kulp DW, Toy L, Hu X, Deresa I, et al. The human naive B cell repertoire contains distinct subclasses for a germline-targeting HIV-1 vaccine immunogen. *Sci Transl Med*. 2018 Jul 4; 10(448):eaat0381. <https://doi.org/10.1126/scitranslmed.aat0381> PMID: 29973404
55. Xu Z, Wise MC, Chokkalingam N, Walker S, Tello-Ruiz E, Elliott STC, et al. In Vivo Assembly of Nanoparticles Achieved through Synergy of Structure-Based Protein Engineering and Synthetic DNA Generates Enhanced Adaptive Immunity. *Adv Sci*. 2020 Apr; 7(8):1902802. <https://doi.org/10.1002/advsc.201902802> PMID: 32328416
56. Gowthaman R, Guest JD, Yin R, Adolf-Bryfogle J, Schief WR, Pierce BG. CoV3D: a database of high resolution coronavirus protein structures. *Nucleic Acids Research*. 2021 Jan 8; 49(D1):D282–7. <https://doi.org/10.1093/nar/gkaa731> PMID: 32890396
57. Cao L, Diedrich JK, Ma Y, Wang N, Pauthner M, Park SKR, et al. Global site-specific analysis of glycoprotein N-glycan processing. *Nat Protoc*. 2018 Jun; 13(6):1196–212. <https://doi.org/10.1038/nprot.2018.024> PMID: 29725121
58. Freedberg DI, Kwon J. Solution NMR Structural Studies of Glycans. *Isr J Chem*. 2019 Nov; 59(11–12):1039–58.

59. Mulligan VK, Melo H, Merritt HI, Slocum S, Weitzner BD, Watkins AM, et al. Designing Peptides on a Quantum Computer [Internet]. *Bioengineering*; 2019 Sep [cited 2021 Jan 12]. Available from: <http://biorxiv.org/lookup/doi/10.1101/752485>.
60. Shapovalov MV, Dunbrack RL. A Smoothed Backbone-Dependent Rotamer Library for Proteins Derived from Adaptive Kernel Density Estimates and Regressions. *Structure*. 2011 Jun; 19(6):844–58. <https://doi.org/10.1016/j.str.2011.03.019> PMID: 21645855
61. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. *Nature*. 2020 Sep 17; 585(7825):357–62. <https://doi.org/10.1038/s41586-020-2649-2> PMID: 32939066
62. McKinney W. Data Structures for Statistical Computing in Python. In Austin, Texas; 2010 [cited 2021 Jan 7]. p. 56–61. Available from: <https://conference.scipy.org/proceedings/scipy2010/mckinney.html>.
63. Waskom M, Gelbart M, Botvinnik O, Ostblom J, Hobson P, Lukauskas S, et al. mwaskom/seaborn: v0.11.1 (December 2020) [Internet]. Zenodo; 2020 [cited 2021 Jan 7]. Available from: <https://zenodo.org/record/592845>.
64. Murray AN, Chen W, Antonopoulos A, Hanson SR, Wiseman RL, Dell A, et al. Enhanced Aromatic Sequons Increase Oligosaccharyltransferase Glycosylation Efficiency and Glycan Homogeneity. *Chemistry & Biology*. 2015 Aug; 22(8):1052–62. <https://doi.org/10.1016/j.chembiol.2015.06.017> PMID: 26190824
65. Le K, Adolf-Bryfogle J, Klima J, Lyskov S, Labonte J, Bertolani S, et al. PyRosetta Jupyter Notebooks Teach Biomolecular Structure Prediction and Design [Internet]. *ENGINEERING*; 2020 Feb [cited 2021 Jan 7]. Available from: <https://www.preprints.org/manuscript/202002.0097/v1>.
66. Schoeder CT, Schmitz S, Adolf-Bryfogle J, Sevy AM, Finn JA, Sauer MF, et al. Modeling Immunity with Rosetta: Methods for Antibody and Antigen Design. *Biochemistry*. 2021 Mar 11;acs.biochem.0c00912. <https://doi.org/10.1021/acs.biochem.0c00912> PMID: 33705117
67. Koehler Leman J, Lyskov S, Lewis S, Adolf-Bryfogle J, Alford RF, Barlow K, et al. Ensuring scientific reproducibility in bio-macromolecular modeling via extensive, automated benchmarks [Internet]. *Bioinformatics*; 2021 Apr [cited 2021 Sep 27]. Available from: <http://biorxiv.org/lookup/doi/https://doi.org/10.1101/2021.04.04.438423>
68. Hunter JD. Matplotlib: A 2D Graphics Environment. *Comput Sci Eng*. 2007; 9(3):90–5.
69. Arroyuelo A, Vila JA, Martin OA. Azahar: a PyMOL plugin for construction, visualization and analysis of glycan molecules. *J Comput Aided Mol Des*. 2016 Aug; 30(8):619–24. <https://doi.org/10.1007/s10822-016-9944-x> PMID: 27549814