Automatically fixing errors in glycoprotein structures with Rosetta

Brandon Frenz[1,2], **Sebastian Rämisch**[3], **Andrew J. Borst**[1], **Alexandra C. Walls**[1], **Jared Adolf-Bryfogle**[3], **William R. Schief**[3], **David Veesler**[1], and **Frank DiMaio**[1,2,*]

[1]Department of Biochemistry, University of Washington, Seattle, WA 98195, USA

[2]Institute For Protein Design, University of Washington, Seattle, WA 98195, USA

[3]Department of Immunology and Microbiology, The Scripps Research Institute, La Jolla, CA, USA.

Summary

Recent advances in single-particle cryo-electron microscopy (cryoEM) have resulted in determination of an increasing number of protein structures with resolved glycans. However, existing protocols for the refinement of glycoproteins at low resolution have failed to keep up with these advances. As a result, numerous deposited structures contain glycan stereochemical errors. Here, we describe a Rosetta-based approach for both cryoEM and X-ray crystallography refinement of glycoproteins which is capable of correcting conformational and configurational errors in carbohydrates. Building upon a previous Rosetta framework, we introduced additional features and score terms enabling automatic detection, setup and refinement of glycan-containing structures. We benchmarked this approach using twelve crystal structures and showed that glycan geometries can be automatically improved while maintaining good fit to the crystallographic data. Finally, we used this method to refine carbohydrates of the human coronavirus NL63 spike glycoprotein and of an HIV envelope glycoprotein, demonstrating its usefulness for cryoEM refinement.
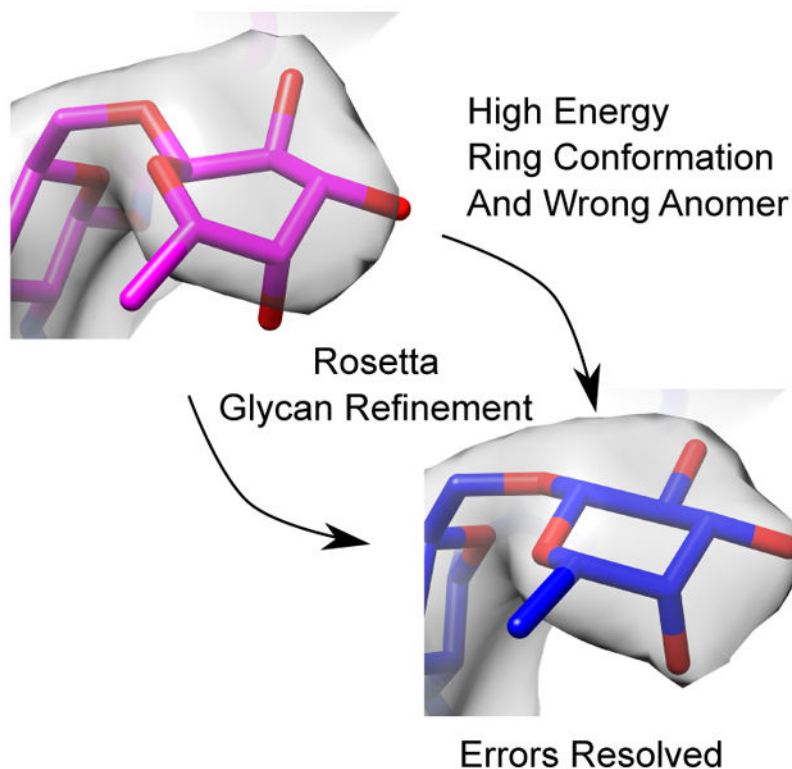
Graphical Abstract

High Energy
Ring Conformation
And Wrong Anomer

Rosetta
Glycan Refinement

Errors Resolved

Etoc Blurb:

Frenz et al. have developed a new method for refinement of glycoprotein structures against low resolution cryoEM and x-ray crystallography data. This new method is able to make significantly larger changes to the glycan geometry compared to previous methods, including the ability to changing the glycan's anomer.

Introduction

Carbohydrates are some of the most stereochemically complex biological molecules found in nature. In addition to their energetic and structural roles in living systems, their role when covalently linked to proteins is critically important in a myriad of molecular recognition processes. Viruses frequently use a glycan shield as a tool to evade the immune system, making glycoproteins the focus of intense interest for vaccinology initiatives targeting these viruses (Astronomo and Burton, 2010). Glycoproteins have historically been difficult to study structurally due to difficulties in overexpressing properly glycosylated proteins as well as their innate flexibility. Indeed, carbohydrates are frequently removed for crystallization studies (Derewenda, 2004). Despite these challenges, several thousand glycoprotein crystal structures have been determined. Recent advances in cryoEM have allowed structural studies on previously intractable glycoproteins (Walls et al., 2016a; Walls et al., 2016b; Xiong, X. et al., 2017; Walls et al., 2017; Lyumkis et al., 2013). However, the vast majority of glycoprotein structures suffer from the fact that the resolution of the electron density or electron potential maps for the carbohydrate chains is too limited to allow for accurate

atomic modeling of individual glycan moieties. As a result, the implementation of prior knowledge is necessary to obtain reliable and stereochemically realistic structural models.

In 2004, a study reported that 30% of all PDB entries with covalently linked carbohydrates contain errors in nomenclature and/or chemistry (Derewenda, 2004). The potential for widespread erroneous carbohydrate structural analysis is increasing with the rise of cryoEM as a near-atomic resolution structural technique, as exemplified by the numerous recent structures solved with unrealistic high-energy ring conformations (Agirre et al., 2015a). These observations emphasize the inadequacy and underutilization of available tools and highlight the need for the development of dedicated algorithms for glycoprotein structural refinement. To address this issue, we aimed to create a tool for the automatic detection and refinement of glycan coordinates while resolving incorrect starting glycan conformations.

We developed an approach to identify, correct, and refine glycoproteins guided by low-resolution crystallographic or cryoEM data to expedite structure determination and interpretation. The approach builds upon previous Rosetta-based structure determination tools guided by low-resolution experimental data (Frenz et al., 2017; Wang et al., 2015), and makes use of a previously developed framework for modeling of carbohydrates (Labonte et al., 2017). Compared to previous glycan refinement methods (Gristick et al. 2017), our approach: a) uses a physically realistic forcefield to ensure glycan geometry remains correct even under large conformational changes, and b) adds the ability to change the anomer of the glycan. These protocols are publicly available with the latest Rosetta release and enable refinement of glycoproteins structures against cryoEM (Wang et al., 2016) and X-ray crystallography data, the latter taking advantage of the combined Phenix-Rosetta reciprocal-space refinement pipeline (Terwilliger et al., 2012).

Results

We developed a refinement protocol to detect and correct poorly modelled glycan configurations, and to refine glycan coordinates guided by either low-resolution crystallographic data, via Phenix-Rosetta integration, (DiMaio et al., 2013a) or cryoEM density data. Of particular interest were the sugar rings of glycans, which may adopt a variety of different conformations with a range of energies (Figure 1A); therefore, we designed our protocol to specifically address these ring conformations. For more on glycan ring conformations see Agirre et al. 2017a. Several glycan-refinement specific methods have been developed, including: tools for automatic detection and setup of glycan-containing structures for subsequent refinement; a score term that enables Cartesian refinement of carbohydrate chains; a dedicated routine for handling sugar ring stereochemistry; and a protocol for correcting the geometry of incorrectly built glycans while optimizing the fit to density. The protocol has a large radius of convergence and is able to make substantial modifications to the input models. For example, the protocol can refine high-energy sugar ring conformers into low-energy conformers and correct erroneous anomeric configurations (see Figure 1). Runtime is about 6 times slower than Phenix refinement: a 300 residue protein was refined in ~25 minutes compared to ~4 mins for Phenix refinement alone. The approach is fully described in Methods.

To test our refinement protocol, we considered a benchmark set of 12 N-linked glycan-containing protein structures determined by X-ray crystallography at resolutions ranging between 1.9 Å and 3.5 Å and comprising a total of 133 glycan units. We identified four incorrect anomeric configurations and 23 high-energy ring conformations using the Privateer software (Agirre et al., 2015b) and detected by visual inspection that one structure was missing a glycosidic bond (Table 1). We compared refinement of our Rosetta based method to that of Phenix refinement alone and, for models with high energy ring conformations in the input, Phenix refinement with constraints generated by Privateer (Gristick et al. 2017). Following refinement of these 12 structures with our Phenix-Rosetta protein structure refinement pipeline, which alternates real and reciprocal-space refinement, we were able to markedly improve the carbohydrate geometry, as assessed by Privateer. All the errors detected in the input coordinates were corrected in the output models. In particular, four incorrect anomeric carbon configurations were resolved by adjusting to the correct anomeric state and 23 high-energy ring conformations were refined into a corresponding low-energy conformation with only a slight decline in agreement to the experimental data, consistent with the idea that these glycans are being forced into poor geometry in order to over fit the density (Supplemental Figure 1). The best alternative method tested, Phenix refinement with constraints from Privateer, was able to resolve only two of the four incorrect anomers and 12 of the 23 sugars in high energy conformations (Table 1).

A comparison of real-space correlations of the refined and initial models, using both 2mFo-dFc density maps as well as Polder omit maps (Liebschner et al., 2017) is also shown in Supplemental Figure 1. While the geometry consistently improves following Rosetta refinement, the real-space correlations show mixed results: while in some sugars, we see a better fit to the data, in other cases we see a slight worsening. This might be due to the fact our relatively more-restrained model does a worse job at explaining the density resulting from heterogeneous conformations.

To illustrate the improvements resulting from our protocol, we examined the structures of human IL-17AF (PDB 5n92) and IgG1-Fc (5k65). In the IL-17AF structure, fucose 507 was modeled with an incorrect beta configuration which has been resolved to the correct alpha connection in the refined model (Figure 1 C). In IgG1-Fc the fucose 507 is also problematic as it is in a high-energy boat conformation which was automatically detected and corrected in the Rosetta-refined model, which has the expected low-energy $^{1}C_4$ (Figure 1 D). In the IgG1-Fc structure, the most proximal N-acetyl glucosamine of the N-linked glycosylation is not bonded to residue Asn 297 and is not properly fit into density. After refinement and rephasing, the carbohydrate moiety is fit with better agreement to the electron density map and its covalent linkage to Asn 297 is properly formed (Figure 1 E). These residues are also shown in the polder omit map (Supplemental Figure 2) (Liebschner et al., 2017).

Refinement of Glycoproteins using CryoEM Data

In addition to the crystal structures selected for our benchmark set, this glycan refinement pipeline was applied to several viral spike glycoprotein structures recently determined using cryoEM. These include the structures of a human coronavirus (HCoV-NL63) spike glycoprotein (Walls et al., 2016b) and an HIV envelope glycoprotein (Borst et al. manuscript

in preparation) determined at 3.4 Å and 3.8 Å resolution, respectively. These two proteins mediate receptor binding and fusion of the viral and host membranes at the onset of infection. Since coronavirus spike and HIV envelope glycoproteins are the targets of vaccine design initiatives, it is key to define the structure and function of their carbohydrate components.

HCoV-NL63 is a human-infecting respiratory virus decorated by a spike glycoprotein trimer covered by an extensive glycan shield that has been suggested to assist in immune evasion. We used cryoEM to image this glycoprotein and obtained a resolution at 3.4 Å resolution. In this reconstruction, 93 N-linked glycosylation sites were visible in the map with at least two N-acetylglucosamine moieties present at most of the sites. Modeling of the protein was done using a homology model of the mouse hepatitis virus (Walls et al., 2016c) in combination with Rosetta *de novo* (Wang et al., 2015), RosettaES (Frenz et al., 2017), Rosetta density-guided iterative refinement (DiMaio et al., 2015), and hand-tracing using Coot (Emsley and Cowtan, 2004). The glycans were initially placed in the map at their approximate positions and the Rosetta glycan-refinement protocol was subsequently used to improve the fit to density along with ensuring proper stereochemistry of the sugar rings and of the glycan bond lengths and angles. The efficiency of the protocol is illustrated by its ability to correct for inaccuracies typically observed in hand-traced models, such as: the incorrect geometry of the glycosidic bond between Asn 291 and N-acetyl glucosamine 1404 (Figure 2 A); the unrealistic length of the glycosidic bond connecting residue Asn 1174 and the N-linked glycan chain that was not optimally fit into density (Figure 2 B); and high-energy sugar conformations (Figure 2 C).

The HIV envelope glycoprotein trimer is also decorated with an extensive glycan shield that limits access to neutralizing antibodies and thwarts the humoral immune response. Conversely, the arms race between viral evolution mechanisms and the immune system of infected individuals has also led to the elicitation of antibodies that bind glycan-containing epitopes. To assess the efficacy of our glycan refinement protocol at lower resolutions, we analyzed a 3.8 Å resolution reconstruction of an HIV trimer in complex with an antibody antigen-binding fragment (Fab) determined by cryoEM (Borst et al., manuscript in preparation). In this map, a total of 60 N-linked glycans were resolved and manually docked in density using Coot (Emsley and Cowtan, 2004) before refinement using Rosetta. The robustness of the refinement protocol was demonstrated in its ability to correct for the unrealistic length of the glycosidic bond connecting residue Asn 301 and the N-linked glycan chain which was not optimally fit into density (Figure 2 D), the incorrect geometry of the glycosidic bond between Asn 386 and N-acetyl glucosamine 1386 (Figure 2 E), and the inadequate stereochemistry of the glycosidic bonds between Man 1200 and Man 1203 (Figure 2 F).

Discussion

Here we describe a method for refining glycan atomic coordinates against cryoEM and X-ray crystallography data using Rosetta. Since Rosetta uses a physically realistic all-atom force field, it is well suited for modeling into near-atomic resolution density maps, which is the resolution regime achieved for most cryoEM structures. This Rosetta glycan refinement

protocol expands upon previous iterations (Labonte et al., 2017) by avoiding fitting stereochemically unfavorable glycan structures into sparse experimental data and instead yielding physically realistic geometries based on prior knowledge of saccharide chemical properties. Using a benchmark set of twelve deposited crystal structures, we demonstrated our algorithm is capable of correcting models containing significant errors in glycan geometry (Table 1) above and beyond previous methods (Gristick et al. 2017). This is likely due to the increased radius of convergence of our refinement, as well as the ability to flip anomeric state. This functionality that should prove beneficial for large-scale model validation efforts such as 'PDB redo' (Joosten et al., 2011; Terwilliger et al., 2012). We further demonstrated the strength and versatility of this algorithm for improvement of glycan stereochemistry in cryoEM models determined at near-atomic resolution with two examples of glycoproteins structures recently obtained and refined using Rosetta.

The expanded Rosetta glycan framework presented here supports automated formatting of input coordinates and detection of glycan connections, as well as cartesian scoring of the most commonly found glycoprotein saccharides, including several glucose, fucose, and mannose derivatives. Future work will further expand the variety of carbohydrates handled, in order to encompass the full known spectrum of glycans found in nature. Considering the increasing number of glycoprotein structures determined at near-atomic resolution in the past few years, we anticipate this algorithm will be a valuable tool for the structural biology community. Rosetta glycan refinement is easy to use and should help novice modelers assign realistic glycan conformations while simultaneously avoiding those not supported by the experimental density.

Star Methods

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Frank DiMaio (dimaio@uw.edu)

METHOD DETAILS

The methods developed for this manuscript were built within the Rosetta protein structure modelling package, using a previously developed glycan-modelling framework (Labonte 2017). This previous glycan-modeling framework is well suited for *de novo* glycan modeling where ideal glycan geometries are built and refined, but was poorly suited at refinement problems where one wishes to refine externally generated (and possibly high-energy) glycan conformers.

To develop the general-purpose refinement tools outlined in this manuscript, several problems in this framework needed to be addressed. The inability to read and write glycans in a standard format (disallowing interoperability with other software packages), limited the applicability of the approach. Furthermore, the glycan score function was unsuited to refining glycans with nonideal bond-lengths or bond angles. In particular, the inability to energetically assess ring conformations with non-ideal bond lengths and bond angles and the inability to resolve discrepancies between the geometry and the anomeric name assignments

limited the applicability of the approach. Finally, specific refinement protocols for cryoEM and crystallographic refinement needed to be developed. The remainder of this section details each of these changes.

Writing Glycans in Standard PDB format—One of the major limitations of Rosetta's carbohydrate framework was the inability to write the carbohydrates in standard PDB format. While the standard pdb format for glycan nomenclature has a number of limitations, including the requirement that glycans be assigned unintuitive three-letter codes in order to cover as many conformations as possible, the typical protocol for model building with glycans usually involves multiple software packages utilizing this nomenclature. To resolve this we implemented functionality to map the Rosetta glycan names back to the pdb 3 letter codes in a specialized database. This database is easily expandable to support additional glycan types following the syntax shown in Supplemental Figure 3.

These changes have been implemented Rosetta release 3.9, and are enabled by the following flags:

- -write_glycan_pdb_codes : output glycans using standard PDB codes rather than a Rosetta-specific naming scheme

- -output_alternate_atomids : write atom names using standard PDB nomenclature

- -write_pdb_link_records : outputs polysaccharide connections using PDB-formatted "LINK" records.

Automatic Configuration of Glycoproteins—While the standard Rosetta carbohydrate framework is well suited for *de novo* glycan modeling and other problems dealing with models produced and manipulated entirely by Rosetta, it has major limitations when it comes to software compatibility. This presents significant challenges for employing Rosetta for refinement problems in concert with other software packages. These limitations include a strict syntax requirement, defined by the necessary inclusion of explicit link records for every glycan moiety, and a strict limitation on glycan ordering.

To resolve these issues, we implemented a number of improvements to Rosetta including a refactoring of the way Rosetta handles connectivity. Specifically, any standard PDB file can be read without requiring specific organization or generation of specific LINK records. We also implemented a method to automatically determine glycan connectivity using the geometry of the structure. Explicit link records will now override any automatically detected bonds, meaning a user can rely on the auto-detection code to identify all reasonable geometries while explicitly declaring connections for which the geometry is too poor for auto-detection to work properly.

The rosetta flag **-auto_detect_glycan_connections** is used to trigger the use of the auto-detection code, and the flag **-maintain links** is used to overwrite auto-detected connections with explicit links when present.

Energy-terms for Scoring of Glycans—The previous implementation of carbohydrate scoring only considered torsion-space refinement of glycans, in which ideal bond length and

bond angle were assumed. Low-energy sugar ring conformations were sampled but were not assessed energetically and instead fixed to the sampled ideal conformation in the course of structure refinement. When attempting to develop a general refinement protocol for glycan-containing structures against cryoEM or X-ray crystallographic data, this proved to be a significant limitation. First, previous work (DiMaio et al., 2013; Conway et al., 2014) demonstrated that Cartesian-space refinement allowing for deviations in bond angle and geometry is important in both structure refinement and native structure discrimination tasks. Moreover, the restriction of fixed sugar ring conformations throughout refinement made it difficult to correct a large number of previously identified errors in glycan-containing structures.

To address this issue, several components were necessary: a) extending Rosetta's bond geometry term to correctly model the energetics of bond geometry deviations in carbohydrates; b) a scoring term capable of assessing the energetics of different ring conformations (and deviations from those ring conformations); and c) a scoring term that ensures each sugar adopts the correct anomeric state. Combined with the protocol described in the next section, this energy function alone is sufficient to correct significant errors in manually built glycans.

To extend Rosetta's bond geometry term for glycans ideal bond length and bond angles values were taken from phenix using elbow with am1 optimization when the default values in Rosetta were insufficient (Moriarty et al., 2009) and added to the Rosetta database for each glycan in our benchmark set. These glycans currently include alpha and beta glucose, N-acetyl glucosamine, alpha and beta mannose, and alpha and beta fucose. Other glycans are still supported (but will need to be added to the naming database to use pdb formatted outputs), however their bond lengths and angles use standard values based on atom type and should be checked carefully and ideal values updated when outliers occur. Additional constraints were added to ensure the planarity of the amide bond of N-acetyl glucosamine.

To assess the quality of carbohydrate ring conformations, we have made use of a ring conformer database to create a "rotameric" ring model in which the torsions and angles of *38* low-energy ring conformations are tabulated (French and Dowd, 1994). For a given conformation, the nearest ring "rotamer" is identified (using torsional RMSd), and harmonic constraints are generated toward that particular conformation. By adding the -ideal_sugars flag, Rosetta only considers the lowest energy ring conformation, typically the only conformation observed in glycans on a protein surface that have only minimal interaction with the protein or other chemical groups (Agirre et al., 2015a). All experiments reported in this paper make use of this flag.

Glycans adopt two anomeric states, referred to as alpha or beta, based on the position of the anomeric reference atom to the anomeric center carbon. To ensure our score function is capable of resolving incorrect anomers, chirality constraints are added around the anomeric (C1) carbon that enforce the correct anomer. For the common glycans these constraints are implemented as pseudo-torsional constraints on the O of the glycosidic bond, and the C1, C5, and C6 carbons, with an ideal torsion of $-120°$ for the alpha form and $0°$ for the beta form, for glycans with non-standard naming schemes the corresponding atoms, based on the

geometry, are used instead. These constraints, combined with the anomeric hydrogen flipper described in the following section, are sufficient to resolve the incorrect chirality around the anomeric position.

Protocols enabling the Automatic Repair of Errors in Glycoproteins—To build glycoprotein conjugates into sparse cryo-electron microscopy data, we first assigned each protein residue position utilizing a combination of Rosetta *de novo* (Wang et al., 2015) RosettaES (Frenz et al., 2017), homology modeling with RosettaCM (Song et al. 2013), and manual model building in COOT (Emsley et al. 2004). These initial models were then run through the the Rosetta cryoEM refinement protocol (Wang et al., 2016). Next, each glycan was manually docked into the experimental density near the amino acid to which they are covalently bound. This glycosylated structure was subsequently minimized in cartesian space using a reduced weight on the repulsive score term, fa_rep = 0.05. This was followed by idealizing the anomeric hydrogen, in order to resolve any incorrect anomers, and then relaxing the system using the Rosetta "fastrelax" protocol (Tyka et al., 2011). Anomeric hydrogens were then idealized by generating a new residue free of hydrogens and then generating its ideal hydrogen positions based on the location of the non-hydrogen atoms. The anomeric hydrogen of the residue to be modified is then moved to the cartesian coordinates of its idealized counterpart. This step is capable of changing the anomeric form of the structure to match the name provided in the input. Therefore, users are encouraged to use Privateer to detect errors in their structure before and after refinement to ensure the glycans are being refined to the correct conformation.

In order to use Rosetta glycan refinement with crystallography data, we took advantage of Rosetta's integration with Phenix (DiMaio et al., 2013) to refine a number of previously published crystal structures containing glycans. Modifications were made to the high resolution crystal refinement protocols to account for the presence of sugars. Specifically, several steps of minimization, with ramping repulsive weights, interspersed with calls to idealize the anomeric hydrogens, as described above, were placed at the beginning of the protocol. The schedule for ramping repulsive weights in our minimization cycles was set to mimic that of the Rosetta fast relax protocol that has been shown to work well for these types of problems. The successively increasing weight of the fa_rep score term was set to 0.011, 0.1375, 0.3025, and 0.55. Calls to idealize the anomeric hydrogen position are done after minimization cycles 1 and 2 to ensure all incorrect anomers have been resolved. The two additional cycles of minimization following the second call to idealize the anomeric hydrogens ensures that all clashes with the anomeric hydrogen are resolved.

QUANTIFICATION AND STATISTICAL ANALYSIS

Anomalies in the glycoprotein structures were assessed using privateer to report mismatches between the name and conformation of the structure as well as high energy conformational states of the sugars (Table 1). Phenix was used to create polder omit maps and calculate the real space correlation to the glycans in both these maps and the 2mFo-dFc density map (Supplemental Figure 1). Statistical methods of analysis were not used.

DATA AND SOFTWARE AVAILABILITY

This code is available through the Rosetta software package on any release after January 1st 2018. The package is free for academic users and information on licensing can be found at www.rosettacommons.org.

For each protocol, implementation uses "RosettaScripts" (Fleishman et al., 2011) which allows a flexible, XML syntax for describing protocols implemented in Rosetta. The XML scripts used for both protocols are included in the Rosetta distribution and will be kept up to date. The cryoEM refinement script can be found in /main/source/scripts/rosetta_scripts/cryoem/ in the Rosetta source. The script for refinement with crystallography data can be found with the other crystal refinement scripts in the public apps section of the Rosetta software package.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References:

Agirre J, Davies G, Wilson K, and Cowtan K (2015a). Carbohydrate anomalies in the PDB. Nat. Chem. Biol 11, 303–303. [PubMed: 25885951]

Agirre J, Iglesias-Fernández J, Rovira C, Davies GJ, Wilson KS, and Cowtan KD (2015b). Privateer: software for the conformational validation of carbohydrate structures. Nat. Struct. Mol. Biol 22, 833–834. [PubMed: 26581513]

Agirre J, Davies GJ, Wilson KS, and Cowtan KD (2017a). Carbohydrate structure: the rocky road to automation. Curr. Opin. Struct. Biol 44, 39–47. [PubMed: 27940408]

Agirre J (2017b). Strategies for carbohydrate model building, refinement and validation. Acta Crystallogr D Struct Biol 73, 171–186. [PubMed: 28177313]

Astronomo RD, and Burton DR (2010). Carbohydrate vaccines: developing sweet solutions to sticky situations? Nat. Rev. Drug Discov 9, 308–324. [PubMed: 20357803]

Borst AJ, James ZN, Zagotta WN, Ginsberg M, Rey FA, DiMaio F, Backovic M, and Veesler D (2017). The Therapeutic Antibody LM609 Selectively Inhibits Ligand Binding to Human alpha-V beta-3 Integrin via Steric Hindrance.

Cingolani G, Panella A, Perrone MG, Vitale P, Di Mauro G, Fortuna CG, Armen RS, Ferorelli S, Smith WL, and Scilimati A (2017). Structural basis for selective inhibition of Cyclooxygenase-1 (COX-1) by diarylisoxazoles mofezolac and 3-(5-chlorofuran-2-yl)-5-methyl-4-phenylisoxazole (P6). Eur. J. Med. Chem 138, 661–668. [PubMed: 28710965]

Conway P, Tyka MD, DiMaio F, Konerding DE, and Baker D (2014). Relaxation of backbone bond geometry improves protein energy landscape modeling. Protein Sci. 23, 47–55. [PubMed: 24265211]

De Nardis C, Hendriks LJA, Poirier E, Arvinte T, Gros P, Bakker ABH, and de Kruif J (2017). A new approach for generating bispecific antibodies based on a common light chain format and the stable architecture of human immunoglobulin G. J. Biol. Chem 292, 14706–14717. [PubMed: 28655766]

Derewenda Z (2004). The use of recombinant methods and molecular engineering in protein crystallization. Methods 34, 354–363. [PubMed: 15325653]

DiMaio F, Echols N, Headd JJ, Terwilliger TC, Adams PD, and Baker D (2013). Improved low-resolution crystallographic refinement with Phenix and Rosetta. Nat. Methods 10, 1102–1104. [PubMed: 24076763]

DiMaio F, Song Y, Li X, Brunner MJ, Xu C, Conticello V, Egelman E, Marlovits T, Cheng Y, and Baker D (2015). Atomic-accuracy models from 4.5-Å cryo-electron microscopy data with density-guided iterative local refinement. Nat. Methods 12, 361–365. [PubMed: 25707030]

Emsley P, and Cowtan K (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr 60, 2126–2132. [PubMed: 15572765]

Favuzza P, Guffart E, Tamborrini M, Scherer B, Dreyer AM, Rufer AC, Erny J, Hoernschemeyer J, Thoma R, Schmid G, et al. (2017). Structure of the malaria vaccine candidate antigen CyRPA and its complex with a parasite invasion inhibitory antibody. Elife 6.

Fleishman SJ, Leaver-Fay A, Corn JE, Strauch E-M, Khare SD, Koga N, Ashworth J, Murphy P, Richter F, Lemmon G, et al. (2011). RosettaScripts: a scripting language interface to the Rosetta macromolecular modeling suite. PLoS One 6, e20161. [PubMed: 21731610]

Moriarty NW, Grosse-Kunstleve RW, and Adams PD (2009). electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. Acta Crystallogr. D Biol. Crystallogr 65, 1074–1080. [PubMed: 19770504]

French AD, and Dowd MK (1994). Analysis of the ring-form tautomers of psicose with MM3 (92). J. Comput. Chem 15, 561–570.

Frenz B, Walls AC, Egelman EH, Veesler D, and DiMaio F (2017). RosettaES: a sampling strategy enabling automated interpretation of difficult cryo-EM maps. Nat. Methods 14, 797–800. [PubMed: 28628127]

Goepfert A, Lehmann S, Wirth E, and Rondeau J-M (2017). The human IL-17A/F heterodimer: a two-faced cytokine with unique receptor recognition properties. Sci. Rep 7, 8906. [PubMed: 28827714]

Gorelik A, Randriamihaja A, Illes K, and Nagar B (2017). A key tyrosine substitution restricts nucleotide hydrolysis by the ectoenzyme NPP5. FEBS J. 284, 3718–3726. [PubMed: 28898552]

Gristick HB, Wang H, and Bjorkman PJ (2017). X-ray and EM structures of a natively glycosylated HIV-1 envelope trimer. Acta Crystallogr D Struct Biol 73, 822–828. [PubMed: 28994411]

Kanai R, Kar K, Anthony K, Gould LH, Ledizet M, Fikrig E, Marasco WA, Koski RA, and Modis Y (2006). Crystal structure of west nile virus envelope glycoprotein reveals viral surface epitopes. J. Virol 80, 11000–11008. [PubMed: 16943291]

Joosten RP, te Beek TAH, Krieger E, Hekkelman ML, Hooft RWW, Schneider R, Sander C, and Vriend G (2011). A series of PDB related databases for everyday needs. Nucleic Acids Res. 39, D411–D419. [PubMed: 21071423]

Labonte JW, Adolf-Bryfogle J, Schief WR, and Gray JJ (2017a). Residue-centric modeling and design of saccharide and glycoconjugate structures. J. Comput. Chem 38, 276–287. [PubMed: 27900782]

Liebschner D, Afonine PV, Moriarty NW, Poon BK, Sobolev OV, Terwilliger TC, and Adams PD (2017). Polder maps: improving OMIT maps by excluding bulk solvent. Acta Crystallogr D Struct Biol 73, 148–157. [PubMed: 28177311]

Lin CC, Wu BS, and Wu WG (2017). Crystal structure of snake venom phosphodiesterase (PDE) from Taiwan cobra (Naja atra atra) in complex with AMP.

Lobner E, Humm A-S, Mlynek G, Kubinger K, Kitzmüller M, Traxlmayr MW, Djinovi-Carugo K, and Obinger C (2017). Two-faced Fcab prevents polymerization with VEGF and reveals thermodynamics and the 2.15 Å crystal structure of the complex. MAbs 9, 1088–1104. [PubMed: 28816592]

Lyumkis D, Julien J-P, de Val N, Cupo A, Potter CS, Klasse P-J, Burton DR, Sanders RW, Moore JP, Carragher B, et al. (2013). Cryo-EM structure of a fully glycosylated soluble cleaved HIV-1 envelope trimer. Science 342, 1484–1490. [PubMed: 24179160]

Modis Y, Ogata S, Clements D, and Harrison SC (2005). Variable surface epitopes in the crystal structure of dengue virus type 3 envelope glycoprotein. J. Virol 79, 1223–1231. [PubMed: 15613349]

Qin Z, Yang D, You X, Liu Y, Hu S, Yan Q, Yang S, and Jiang Z (2017). The recognition mechanism of triple-helical β-1,3-glucan by a β-1,3-glucanase. Chem. Commun 53, 9368–9371.

Schwarzenbacher R, Zeth K, Diederichs K, Gries A, Kostner GM, Laggner P, and Prassl R (1999). Crystal structure of human beta2-glycoprotein I: implications for phospholipid binding and the antiphospholipid syndrome. EMBO J. 18, 6228–6239. [PubMed: 10562535]

Song Y, DiMaio F, Wang RY-R, Kim D, Miles C, Brunette T, Thompson J, and Baker D (2013). High-resolution comparative modeling with RosettaCM. Structure 21, 1735–1742. [PubMed: 24035711]

Terwilliger TC, Dimaio F, Read RJ, Baker D, Bunkóczi G, Adams PD, Grosse-Kunstleve RW, Afonine PV, and Echols N (2012). phenix.mr_rosetta: molecular replacement and model rebuilding with Phenix and Rosetta. J. Struct. Funct. Genomics 13, 81–90. [PubMed: 22418934]

Tyka MD, Keedy DA, André I, Dimaio F, Song Y, Richardson DC, Richardson JS, and Baker D (2011). Alternate states of proteins revealed by detailed energy landscape mapping. J. Mol. Biol 405, 607–618. [PubMed: 21073878]

Walls AC, Tortorici MA, Bosch B-J, Frenz B, Rottier PJM, DiMaio F, Rey FA, and Veesler D (2016a). Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer. Nature 531, 114–117. [PubMed: 26855426]

Walls AC, Tortorici MA, Frenz B, Snijder J, Li W, Rey FA, DiMaio F, Bosch B-J, and Veesler D (2016b). Glycan shield and epitope masking of a coronavirus spike protein observed by cryo-electron microscopy. Nat. Struct. Mol. Biol 23, 899–905. [PubMed: 27617430]

Walls AC, Tortorici MA, Bosch B-J, Frenz B, Rottier PJM, DiMaio F, Rey FA, and Veesler D (2016c). Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer. Nature 531, 114–117. [PubMed: 26855426]

Wang RY-R, Kudryashev M, Li X, Egelman EH, Basler M, Cheng Y, Baker D, and DiMaio F (2015). De novo protein structure determination from near-atomic-resolution cryo-EM maps. Nat. Methods 12, 335–338. [PubMed: 25707029]

Wang RY-R, Song Y, Barad BA, Cheng Y, Fraser JS, and DiMaio F (2016). Automated structure refinement of macromolecular assemblies from cryo-EM maps using Rosetta. Elife 5.

Wu L, Jiang J, Jin Y, Kallemeijn WW, Kuo C-L, Artola M, Dai W, van Elk C, van Eijk M, van der Marel GA, et al. (2017). Activity-based probes for functional interrogation of retaining β-glucuronidases. Nat. Chem. Biol 13, 867–873. [PubMed: 28581485]

- New method for refinement of carbohydrates with low resolution electron density.

- Improved physical geometry of glycans in protein structures.

- Compatible with cryoEM and x-ray crystallography data
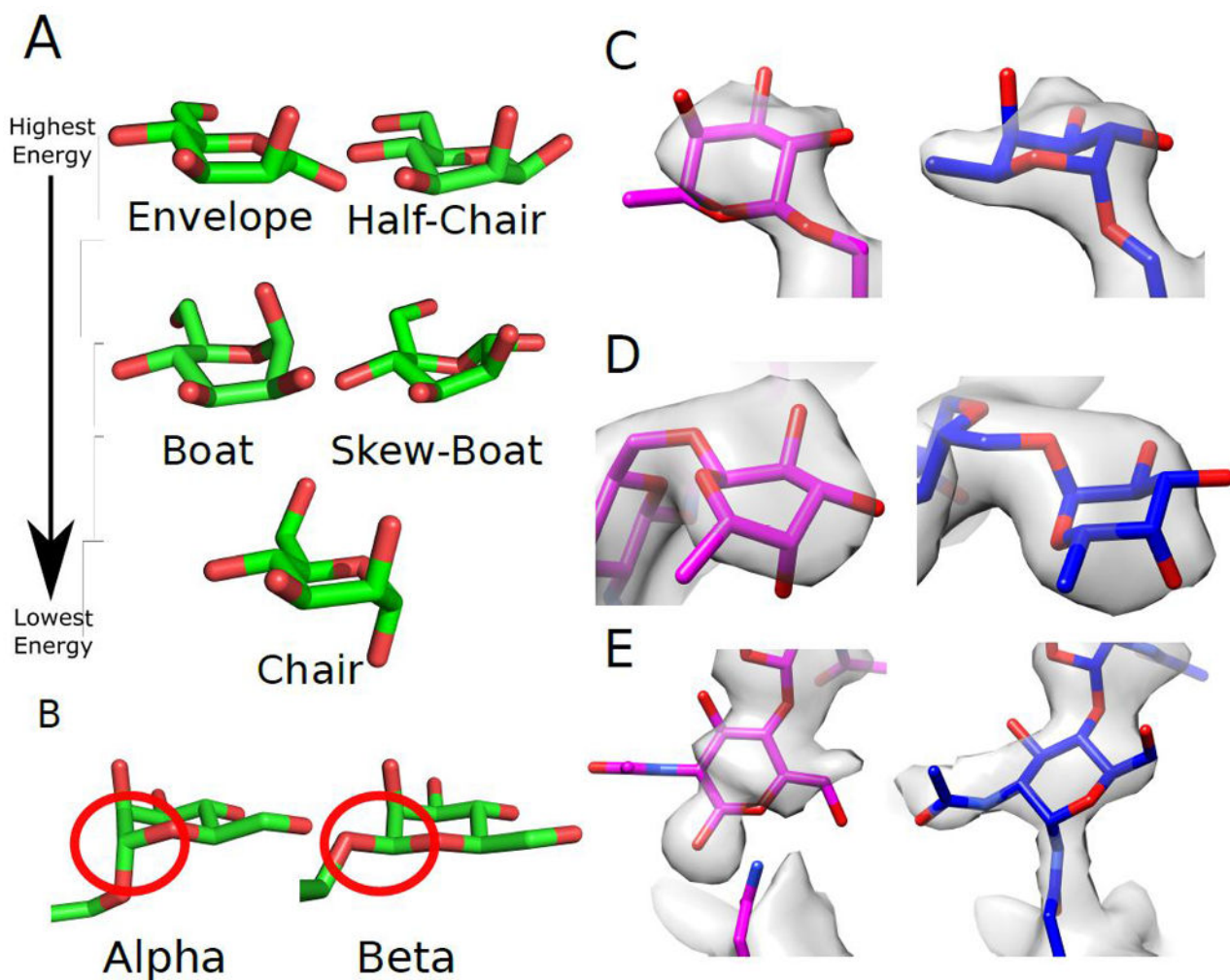
Figure 1. Anomalies in low resolution crystal structures from the PDB resolved with Rosetta glycan refinement.
A. Mannose shown in common ring conformations that can be interconverted rank by energy. Glycan sugar conformations should always be modeled as chairs unless there is strong evidence to the contrary. **B.** Two anomeric forms are shown, alpha with the glycosidic oxygen and the C5 carbon in trans, and beta in cis. The anomeric carbon is circled in red. **C.** Fucose 507 of PDB entry 5nsc has an incorrect anomeric connection in the input (magenta), and this is resolved in the refined model (blue). **D.** Fucose 507 of PDB entry 5k65 is in a high energy ring conformation, and this is resolved in the refined model. **E.** Asp 297 of chain B in entry 5k65 fails to form a glycosidic bond to the N-acetyl glucosamine. This is resolved in the output model, and the connection can be seen in the density after rephasing. All density maps are shown at a threshold of 1.
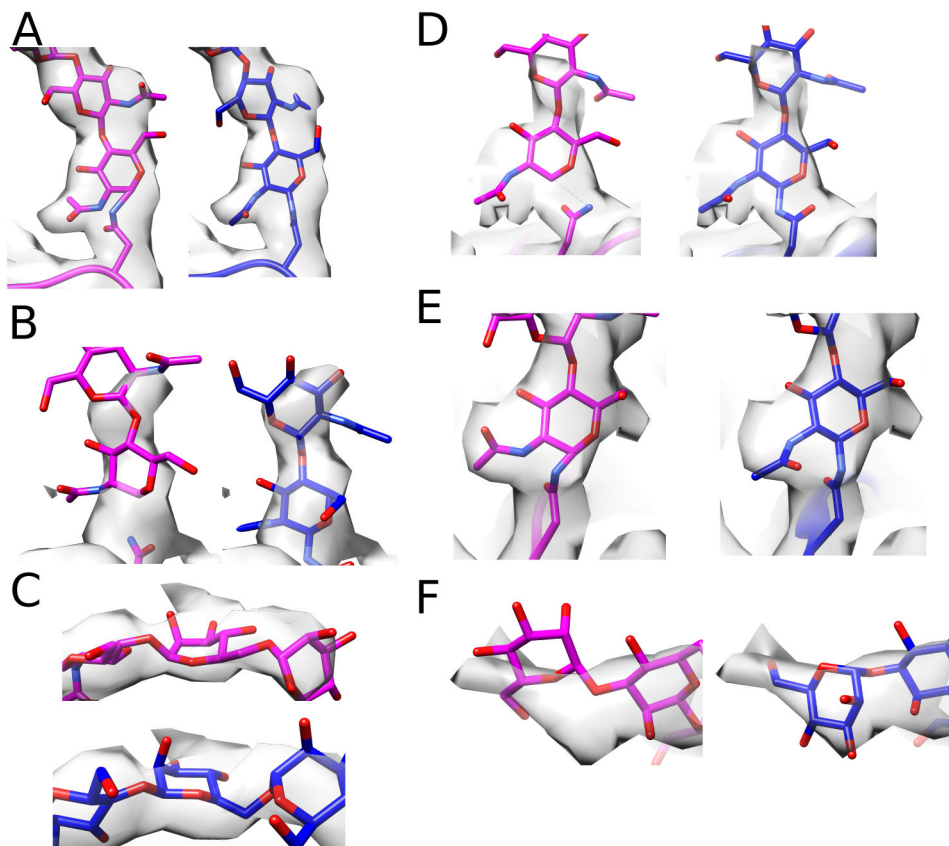
Figure 2. Correcting high energy glycans in cryoEM structures
A. The unfavorable glycosidic bond between asparagine 241 and N-acetyl glucosamine 1404 of NL63 (magenta) is resolved in the refined model (blue). **B.** The poor fit to the density of the glycan chain and disconnected glycosidic bond of asparagine 1174 in the NL63 input model is resolved during refinement. **C.** The high energy, envelope, ring conformation of mannose 1428 of NL63, center, is resolved during refinement. **D.** NAG 1301 of HIV does not form a proper glycosidic bond, and the glycans of the chain do not fit the density in the input (magenta). These issues are resolved in the refined model (blue). **E.** NAG 1386 of HIV has a unfavorable glycosidic bond angle (magenta) which is resolved in the refined model (blue). **F.** MAN 1200 of HIV fits the density poorly and the glycosidic bond angle is unfavorable (magenta). In the output model (blue) these issues are resolved.

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Protein Structure	Walls et al (2016)	PDBID: 5SZA
Protein Structure	Schwarzenbacher et al. (1999)	PDBID: 1C1Z
Protein Structure	Modis et al. (2005)	PDBID: 1UZG
Protein Structure	Kanai et al. (2006)	PDBID: 2I69
Protein Structure	Favuzza et al. (2007)	PDBID: 5EZJ
Protein Structure	Lin et al. (2017)	PDBID: 5GZ4
Protein Structure	Qin et al. (2017)	PDBID: 5H9Y
Protein Structure	Lobner et al. (2017)	PDBID: 5K65
Protein Structure	Wu et al. (2017)	PDBID: 5LA4
Protein Structure	Goepfert et al. (2017)	PDBID: 5N92
Protein Structure	De Nardis et al. (2017)	PDBID: 5NSC
Protein Structure	Gorelik et al. (2017)	PDBID: 5VEM
Protein Structure	Cingolani et al. (2017)	PDBID: 5WBE
Electron Density	Walls et al. (2016)	EMDB: 8331
Software and Algorithms		
Rosetta	(Labonte et al., 2017)	N/A
Phenix	(Terwilliger et al., 2012)	N/A
Privateer	(Agirre et al., 2015b)	N/A

Table 1.
Anomalies in deposited structures of glycoprotein conjugates.

This table shows the resolution of the experimental data as well as the number of incorrect anomers and high energy ring conformations, as reported by Privateer, for each structure in our benchmark set of crystal stuctures before and after refinement with the three different methods, Phenix refinement alone, Phenix-Privateer (when high energy ring conformations are present in the input), and Phenix-Rosetta glycan refinement. Cells marked na are those for which privateer constraints are not generated as no high energy conformations are detected in the input.

PDB ID	# of Glycans	Reported Resolution (Å)	Original Model		Phenix Refined		Phenix + Privateer		Rosetta Model	
			Wrong Anomer	High Energy Ring	Wrong Anomer	High Energy Ring	Wrong Anomer	High Energy Ring	Wrong Anomer	High Energy Ring
1c1z	11	2.87	2	2	0	4	0	2	0	0
1uzg	13	3.5	0	0	0	0	na	na	0	0
2i69	3	3.11	0	2	0	0	0	0	0	0
5ezj	4	1.95	0	0	0	0	na	na	0	0
5gz4	12	2.55	0	2	0	2	0	0	0	0
5h9y	15	1.97	0	8	1	6	0	3	0	0
5k65	12	2.5	1	1	1	1	1	1	0	0
5la4	7	1.9	0	0	0	0	na	na	0	0
5n92	3	2.3	0	1	0	0	0	0	0	0
5nsc	14	2.3	1	0	0	2	na	na	0	0
5vem	18	2.6	0	1	0	1	0	0	0	0
5wbe	21	2.75	0	6	0	5	0	5	0	0
Total	**133**		**4**	**23**	2	20	1	11	**0**	**0**