Structure
Ways & Means

All-Atom Ensemble Modeling to Analyze Small-Angle X-Ray Scattering of Glycosylated Proteins

Miklos Guttman,[1] Patrick Weinkam,[2] Andrej Sali,[2] and Kelly K. Lee[1,*]
[1]Department of Medicinal Chemistry, University of Washington, Seattle, WA 98195, USA
[2]Department of Bioengineering and Therapeutic Sciences, Department of Pharmaceutical Chemistry, and California Institute for Quantitative Biosciences, University of California, San Francisco, San Francisco, CA 94158, USA
*Correspondence: kklee@uw.edu
http://dx.doi.org/10.1016/j.str.2013.02.004

SUMMARY

The flexible and heterogeneous nature of carbohydrate chains often renders glycoproteins refractory to traditional structure determination methods. Small-angle X-ray scattering (SAXS) can be a useful tool for obtaining structural information of these systems. All-atom modeling of glycoproteins with flexible glycan chains was applied to interpret the solution SAXS data for a set of glycoproteins. For simpler systems (single glycan, with a well-defined protein structure), all-atom modeling generates models in excellent agreement with the scattering pattern and reveals the approximate spatial occupancy of the glycan chain in solution. For more complex systems (several glycan chains, or unknown protein substructure), the approach can still provide insightful models, though the orientations of glycans become poorly determined. Ab initio shape reconstructions appear to capture the global morphology of glycoproteins but in most cases offer little information about glycan spatial occupancy. The all-atom modeling methodology is available as a web server at http://salilab.org/allosmod-foxs.

INTRODUCTION

The covalent addition of carbohydrate groups to proteins is one of the most common posttranslational modifications, with over half of the human proteome containing some degree of glycosylation (Brooks, 2009). Glycosylation can stabilize proteins, shield the underlying protein from proteases, increase solubility, and often plays a critical role in cell-cell and host-pathogen interactions (Shental-Bechor and Levy, 2008). Despite their importance, glycoproteins as a class remain a challenging target for structural biology (Walsh, 2010). The high degree of flexibility of the glycans presented on the protein surface as well as the significant micro- and macroglycan heterogeneity exhibited within a purified population of glycoprotein (Imberty and Pérez, 2000) limits the efficacy of X-ray crystallography as a tool for their structural determination. Glycan modifications also limit the utility of nuclear magnetic resonance (NMR) spectroscopy due to detrimental effects on rotational correlation times and significant chemical shift overlap (Slynko et al., 2009).

Small-angle X-ray scattering (SAXS) has become a versatile tool for examining macromolecules in the native solution state. It can be used to study disordered systems and molecules with significant internal flexibility (Bernadó, 2010; Hammel, 2012; Jacques and Trewhella, 2010; Mertens and Svergun, 2010; Pelikan et al., 2009; Putnam et al., 2007; Rambo and Tainer, 2010). Two caveats add a layer of complexity for applying SAXS to study glycoproteins, namely, the different scattering contrast of the carbohydrate relative to the protein (Hammel et al., 2002) and the disproportionate layer of hydration associated with the sugars (Wittmann, 2007). Despite these limitations, the low-resolution morphologies of a number of glycoproteins determined by SAXS have been reported (Bernocco et al., 2003; Guttman et al., 2012; Lynn et al., 2005). While most of these studies are unable to directly visualize the influence or position of the glycans, in some instances the glycan positions appeared to be identifiable (Hammel et al., 2002). Rigid body refinement and molecular dynamics guided by SAXS data have been established as a method for obtaining more detailed models, but the current programs are not suited for taking glycosylation into account (Pelikan et al., 2009).

NMR studies of isolated carbohydrate chains and molecular dynamics have shown that significant internal flexibility exists within glycan chains (Woods et al., 1998). All-atom force fields have been developed for assessing the internal structures of sugars in silico (Frank and Schloissnig, 2010; Kirschner et al., 2008), and computational modeling has been useful for generating fully glycosylated structures using the nonglycosylated template (Bohne-Lang and von der Lieth, 2005; Renouf and Hounsell, 1995). However, experimental validation of these structural models has been limited.

Here, we establish the approach of assessing large ensembles of glycoprotein models against SAXS data to test the validity of the models. Concurrently, we examine the reliability of SAXS ab initio reconstruction for glycoproteins. Last, we demonstrate the accuracy of molecular weight (MW) determination from the SAXS intensity for glycoproteins. This report provides a guideline and documents limitations of SAXS applied to the study of glycoproteins.

Table 1. Molecular Mass Determinations for the Panel of Glycoproteins

	Mass Glycan (%)	Expected[a]	SLS[b]	DLS[c]	I(0) (water)	I(0) (stds)	SAXS MoW[d]	Autoporod[d]	N-PAGE	SEC[e]
Lysozyme	0	14.3	14.3	10.6 ± 0.8	12.7	10.9	8.6	11	ND	3.4
RNaseA	0	13.7	15.1	10.3 ± 1	15.9	ND	9.9/9.9	10.3/10	ND	10.7
RNaseB	9.1	15.1	15.3	13 ± 1.4	16.8	14.8	10.4/12.2	11.4/11.6	ND	13.5
a1AGP	40.2	36.1	35.2	46.1 ± 1.5	35.9	34.0	41.8/39.4	60.7/56.5	38.3	75.2
Fetuin	23.6	47.6	48	65 ± 2.1	46.9	42.7	49.4/44.9	65/65.2	75.8	111.1
17b IgG Fc	6.0	53.6	62.5	56.6 ± 2.1	57.9	60.2	51.1	47.1	Smear	46.2
BHA	17.2	203.7	184.7	204.6 ± 6.4	194.5	183.6	176.0	252.8	322	316.2

Molecular mass determinations are shown in kDa.
[a]MW calculated for the most abundant glycoforms.
[b]MW estimated from static light scattering. Uncertainty in this measurement is primarily dependent on the uncertainty in concentration determination, expected to be <15%.
[c]MW estimation (average ± SD) from dynamic light scattering assuming a spherical model.
[d]MW estimation from SAXSMoW using q range of 0–0.45 (Fischer et al., 2010). The second set of values for SAXSMow and Autoporod (Petoukhov et al., 2007) were obtained from the FPLC-SAXS data sets. Uncertainty is expected to be around 10%.
[e]MW estimated from size exclusion chromatography.

RESULTS AND DISCUSSION

Molecular Mass Determination of Glycoproteins

Many classic techniques for MW determination of proteins and complexes yield inaccurate estimates when applied to proteins with significant glycan content (Lewis and Junghans, 2000; Wen et al., 1996). In our analysis, the MWs estimated from size exclusion chromatography and native polyacrylamide gel electrophoresis (N-PAGE) are significantly off for proteins with high glycan content, in some cases by a factor of two (Table 1). MW estimates from dynamic light scattering, which assumed a spherical model, were overall more accurate than size exclusion chromatography (SEC) and N-PAGE, but the masses of several glycoproteins were still overestimated. This is probably the effect of flexible glycans contributing disproportionately to the observed hydrodynamic radius and thus the inferred globular MW. In contrast, accurate MW values are obtained by static light scattering (SLS) for every sample examined.

An accurate MW derived from SAXS measurements is an important check for ensuring that the data are usable for structural interpretation (Jacques and Trewhella, 2010). Several methods have been established to calculate the MW from SAXS data (Fischer et al., 2010; Mylonas and Svergun, 2007). The MW determined using the SAXS intensity at zero angle, I(0), relative to either water or a reference protein (in this case RNaseA), appears to be accurate for glycoproteins, provided that any contrast offset and differences in the partial specific volume are accounted for (Table 1). On the other hand, MW estimates of the glycoproteins from the Porod volume with autoporod (Petoukhov et al., 2007) were less accurate than those obtained from I(0) because, as expected, Porod's law and the volume approximation break down when applied to noncompact molecules with nonuniform scattering contrast (Mertens and Svergun, 2010; Porod, 1982). The more recently developed SAXSMoW calculation (Fischer et al., 2010) appears to provide more accurate MW determination than autoporod. We note that SAXSMoW calculations were far less accurate when the data were truncated below q = 0.45 $Å^{-1}$; therefore, data quality at the higher angles seems critical for accuracy.

From the current SAXS I(0) and SLS MW determination, it is clear that all of the proteins examined are monomeric in solution, except for immunoglobulin G (IgG) Fc (dimer) and bromelain cleaved hemagglutinin (BHA) (trimer). In contrast, the inaccurate masses inferred from size exclusion chromatography, native gels, and even autoporod would lead one to mistakenly conclude that several of these species are dimers. The unique case is the Fc portion of human IgG, which shows a lower value with I(0) and better accuracy with the Porod-based MW estimation. Unlike most glycoproteins where the glycans are at the periphery of the protein interacting with the solvent, the Fc domain has the glycans positioned at the center of the molecule, forming a part of the dimeric interface (Deisenhofer, 1981). The restrained nature of these glycans results in a more compact molecule for which the Porod-based MW approximation is suitable.

All-Atom Modeling with Well-Defined Templates

RNaseA/B

Bovine RNaseA and RNaseB serve as a useful comparison for examining the scattering effect of glycans as the two proteins differ only in that RNaseB has a single high-mannose type glycan chain at N34. Furthermore, both RNaseA and RNaseB have been shown by NMR spectroscopy and X-ray crystallography to have highly similar protein structures (Williams et al., 1987). SAXS analysis with online size exclusion chromatography resolved monomeric RNaseA (Figure S1A available online) yielding an Rg consistent with earlier reports (Mylonas and Svergun, 2007) and a Porod volume and Dmax in good agreement with those predicted from the NMR structure (Santoro et al., 1993) (Table S1). The SAXS pattern is also consistent with the theoretical scattering pattern calculated from the NMR ensemble of RNaseA with a χ of 1.47 ± 0.05 (Figure 1A).

The SAXS data for RNaseB showed a slightly larger Rg, Porod volume, and Dmax, as expected due to the additional glycan chain (Table S1). The Kratky plots of RNaseA and RNaseB are similar, indicating a comparable degree of overall compactness (Putnam et al., 2007) (Figure S3). Several structural ensembles of RNaseB were generated with a range of high-mannose type glycan chain sizes from three to nine mannose moieties to test

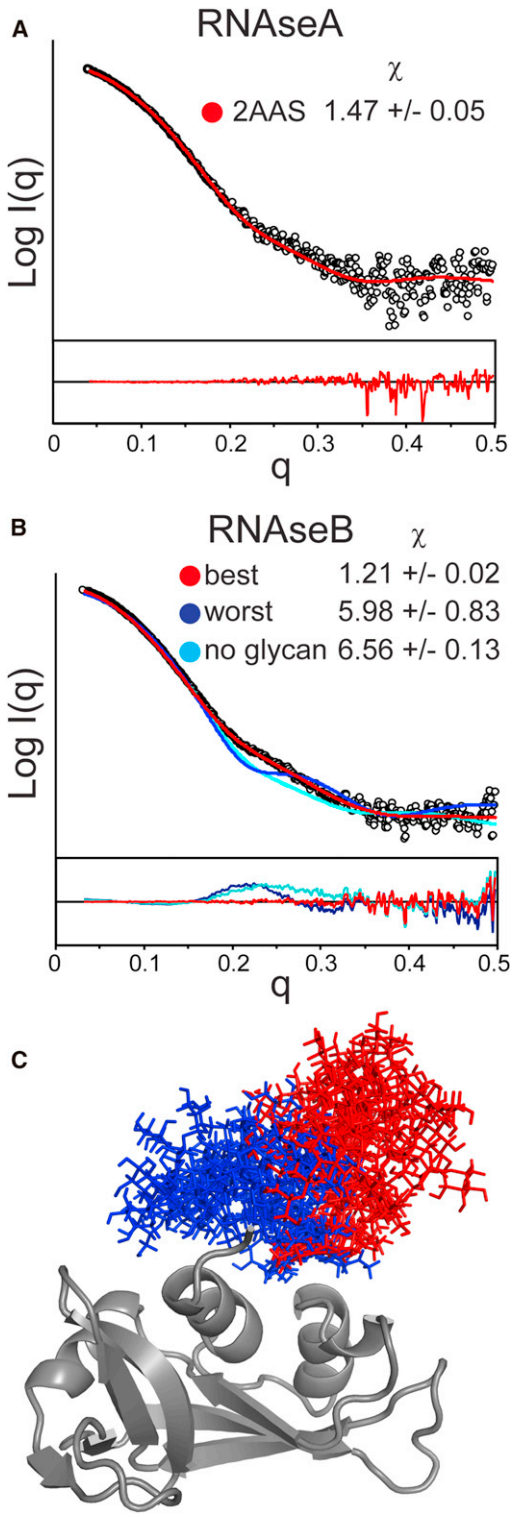

Figure 1. Comparison of Calculated SAXS Patterns from All-Atom Modeling with Experimentally Measured Data for RNAseA Unglycosylated and RNAseB Glycosylated

(A) SAXS profile of RNaseA (black circles) and the theoretical profile of the best-fit model from the NMR ensemble (2AAS) (red).

(B) SAXS profile of RNaseB (black circles) with fits for the best model (lowest χ, red), poorest models (highest χ, blue), and model of RNaseB without the glycan (cyan). Residual plots are shown below. Results of the top ten averages (±SD) of the fits are reported in the inset. FPLC-SAXS traces and Guinier plots are shown in Figures S1 and S3.

(C) Glycan positions are shown for the ten best (red) and poorest (blue) fitting models.

See also Figures S1 and S2.

the effect of glycan size on the fits to the scattering data. The single best fits and overall χ values from each ensemble show that the models with five mannose groups (Man_5) agree most closely with the scattering data (Table S2). This is already a promising result, as it is known that the Man_5 glycoform is the predominant glycoform in RNaseB, with Man_6 through Man_9 in much lower abundance (Guttman, 1997).

A comparison of the best- and poorest-fitting RNaseB structures with a Man_5 glycan indicates a clear difference in glycan orientation (Figures 1B and 1C). In the ten best-fitting models (χ of 1.21 ± 0.02), the glycan chain protrudes outward from the protein, whereas in the models with the worst fits (χ of 5.98 ± 0.83), the glycan chain is localized close to the protein surface. The glycan orientation revealed from the best-fitting models is consistent with crystallographic studies, which concluded that the glycan is relatively disordered and faces into a solvent void in the crystal lattice (Williams et al., 1987).

Since several conformations may exist in solution, especially for flexible systems, a single structural model may not accurately fit the measured SAXS data. The minimal ensemble search (MES) has been developed to find a combination of the smallest possible number of structures that recapitulate the observed SAXS pattern (Pelikan et al., 2009). For the present analyses, MES was attempted to improve the fit, incorporating up to five models from each ensemble. For RNaseB, MES led to no improvement in the fit over the single best model. Taken together, this indicates that, while there is probably high internal flexibility within each glycosidic bond, the spatial occupancy of the full glycan chain is relatively limited. One possible explanation is the carbohydrates can achieve more favorable hydration with the glycan protruding away from the protein. Alternatively, weak interactions may constrain certain dihedrals at the base of the glycan, constraining the overall orientation. Even if some glycan dihedrals were rigid, the current methodology is of insufficient resolution for probing such details.

IgG Fc

The papain-cleaved constant domain (Fc) of human IgG was also isolated and examined. This molecule is a homodimer of the Fc_2 and Fc_3 domains with a single N-linked glycan on each subunit. The crystal structure of IgG Fc includes the coordinates for the entire glycan chain, which form contacts at the cavity between the two Fc_2 domains (Deisenhofer, 1981). Initial models of the Fc domains based on the full coordinates in the crystal, including the termini absent in the crystal, resulted in a χ of 1.77 ± 0.04. Ensembles were generated to probe a diverse occupancy of the glycan chains. The resulting models with the glycans deviating from the dimeric interface showed poorer agreement with the measured SAXS data with χ of 2.17 ± 0.21.

A wide array of fragment antigen binding (Fab) domains of IgG have been structurally characterized and have shown distinct "elbow" angles (angle between Fv and Fc_1). In a recent study, these elbow motions were a requirement for Fab structures to

obtain good fits to SAXS data (Schneidman-Duhovny et al., 2012). As the IgG Fc has a similar hinge between the second and third constant domains (Fc_2 and Fc_3), a structural ensemble was generated allowing hinge motions, while restraining glycans to the crystal coordinates. Agreement with the SAXS data was significantly improved with a χ of 1.42 ± 0.03 (Figure 2A; Table S2). Thus, for this glycoprotein, the sampling of the N-terminal extension and elbow angle, rather than glycan positions, was critical for generating accurate models. Although the sampling of the ten residue N-terminal extension is broad, among best-fitting models it consistently protrudes away from the protein, rather than near the surface. The best models show some variance in the elbow angle, but MES did not lead to an appreciable improvement in the fit (χ of 1.35 versus 1.38 for the single best model). This may highlight one limitation of SAXS, namely, that several distinct structures can yield very similar SAXS patterns. Despite this caveat, the analysis indicates that IgG Fc is not fixed in the angle suggested by the crystal structure.

a1AGP

A step up in complexity, alpha-1-glycoprotein (a1AGP) provided another useful test of our all-atom modeling approach since it bears five complex type glycans and the nonglycosylated protein structure has been determined (Schönfeld et al., 2008). The fully glycosylated form has previously been crystallized, but diffraction was poor, presumably due to the relatively disordered glycans (McPherson et al., 1984). Since the glycans don't seem to alter the tertiary structure (Friedman et al., 1986), models were generated to sample glycan conformations, without any flexibility permitted for the protein beyond the modified Asn residues and the N-terminal tail that was absent from the crystal structures. The resulting structural ensemble of a1AGP with glycans included several structures that were in reasonable agreement with the SAXS data (χ of 1.62 ± 0.06) (Figure 2B; Table S2).

To test the effect of the glycan size on the resulting fits, the same experimental data were fit to an additional ensemble omitting all sialic acids (NeuAc) at the end of each glycan chain. Modeling without NeuAc generated models with significantly poorer fits than with NeuAc (2.27 ± 0.27 versus 1.62 ± 0.06), with each ensemble sampling 200 models. The calculated Rg from the models without NeuAc (23.93 ± 0.25) is also less consistent with the SAXS data than when NeuAc was included (25.56 ± 0.20; Rg obtained from SAXS was 26.5 ± 0.06). Therefore, similar to the RNaseB results, including the proper glycan type/size is important for the accuracy of the resulting models.

While the modeling approach generates models consistent with the SAXS data, comparisons of the best-fit models reveal no apparent spatial convergence for any of the glycan chains. MES resulted in an improvement in the fit (χ 1.32 versus 1.52 for the best single model), indicating that several glycan conformations more accurately describe the solution structure of this system. Unlike with RNaseB, some or all of the glycan chains in a1AGP may, in fact, have broad spatial sampling. Additionally, the higher number of variables (five glycan chains and an unresolved N terminus) will limit the likelihood of capturing a single conformation that most accurately fits the SAXS pattern, even with exhaustive ensemble sizes. However, the fact that good fits are achieved with the glycosylated models indicates that the protein structure in the nonglycosylated crystal structure is consistent with that in the fully glycosylated protein.

BHA

Influenza hemagglutinin is a trimeric viral surface protein that catalyzes host-virus membrane fusion and contains several glycans to shield itself from the host immune response. Although the crystal structure of bromelain-released influenza hemagglutinin (BHA) has been solved in its glycosylated form, the majority of carbohydrate was not resolved, presumably due to inherent flexibility and disorder (Wilson et al., 1981). Of the 200 all-atom models generated including the full 21 glycans (per trimer), the best ten were consistent with the SAXS data with a χ of 1.77 ± 0.03 (Figure 2C). There was no clear spatial occupancy bias for the glycans between the best- and worst-fitting models, and MES for this system led to no significant improvement in the goodness of fit over the best individual model (χ 1.71 versus 1.73). As glycans chains decorate most of the protein surface, changes in any single glycan may have only minor effects on the overall fit. Therefore, for a system essentially coated in glycan chains, our all-atom modeling approach can assess whether or not the overall structure is consistent with SAXS data but reveals little about the specific disposition of individual glycans.

Modeling with Little Structural Information: Fetuin

Since no structural data exist for bovine fetuin, it serves as a good system to test the limits of the all-atom modeling approach for SAXS analysis. Along with the three complex type N-linked glycans, the protein also contains four sites of O-linked glycosylation (Figure S2). The Kratky plot indicates that fetuin is mostly a well-ordered globule with some disordered portions (Figure S3). Based on sequence similarity, the protein has two cystatin domains followed by a 100 residue C-terminal portion. Several algorithms (DISEMBL, DisProt, PrDOS) predict that much of the C-terminal region (residues 240–320) is intrinsically disordered (Ishida and Kinoshita, 2007; Linding et al., 2003; Sickmeier et al., 2007). HS-1, the human homolog of fetuin, is also known to be proteolytically cleaved at the residues 280–317 (Nawratil et al., 1996). The full sequence of fetuin was submitted to I-TASSER along with disulfide restraints for in silico structure prediction (Roy et al., 2010; Zhang, 2008). Low C scores of −4.54 and −4.95 for the top two models were achieved with a relatively low clustering density (0.0049 and 0.0033), indicating poor confidence in the models (predicted to be 0.24 ± 0.07 [TM score] and 18.1 ± 2.4 Å [rmsd]). Nevertheless, the coordinates from each of the five models were used as a template for generating glycosylated ensembles to compare against the SAXS data.

Two out of the five ensembles produced structural models that were in loose agreement with the SAXS data (χ of 2.06 ± 0.12 and 1.95 ± 0.16) (Figure 2D; Table S2). Since these templates have significantly different protein structures, namely, a completely different fold within the N-terminal cystatin domain, this presents a limitation of the current methodology. With so many degrees of freedom, it is possible to generate an array of distinct structures with similar consistency to the scattering data. However, all of the starting models without glycans had terrible fits to the SAXS data (χ > 15). Therefore, the incorporation of the glycans into this type of analysis is essential. With additional restraints, this approach could generate a more reliable model of this glycoprotein.

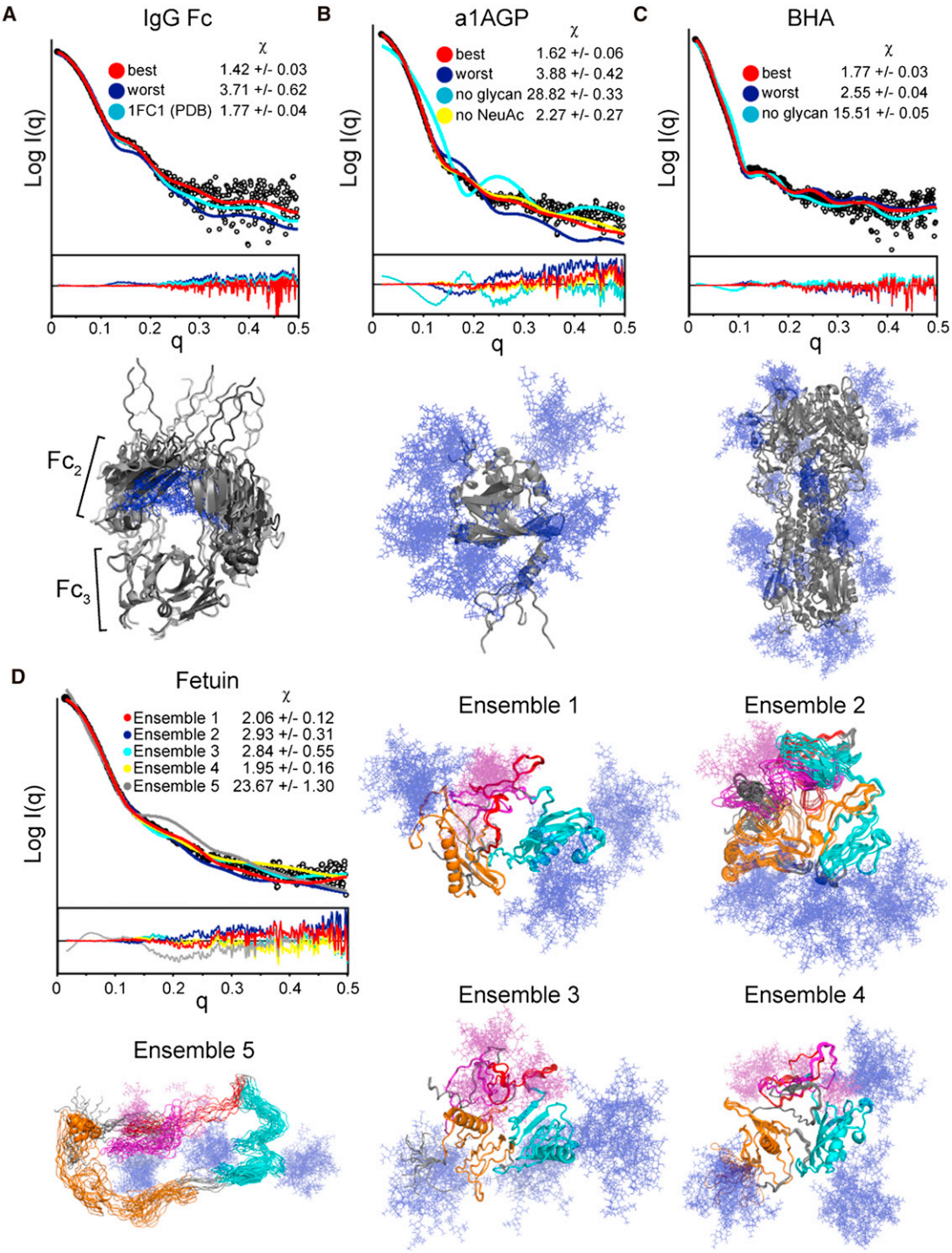

Figure 2. Comparison of Calculated SAXS Patterns from All-Atom Modeling with Experimentally Measured Data

SAXS profile and model fits for IgG Fc (A), a1AGP (B), BHA (C), and fetuin (D) with residual plots shown below. The resulting χ values for each set are reported in the insets and Table S2.

(A–C) Theoretical profiles of the best (red), worst (blue), and model without glycans (cyan) are fit against the SAXS data. Overlays of the top five models of each ensemble are shown below the data sets as gray cartoons with glycans depicted as blue sticks.

(D) SAXS profile of fetuin with theoretical fits for the glycosylated models using the top five structures from I-TASSER as templates. The top ten best-fitting models are shown for each of the five templates. The two predicted cystatin domains (8–118 and 132–230) are shown in orange and cyan, respectively. The region spanning 237–273, predicted to be intrinsically disordered is shown in red. The protease susceptible fragment 277–318 is shown in magenta. N-linked (blue) and O-linked (purple) glycan chains are shown as sticks.

See also Figures S1 and S2.

Ab Initio Reconstructions

The most common method of extracting structural information from SAXS data is the use of ab initio bead modeling to construct a shape envelope of the molecule consistent with the SAXS data (Petoukhov and Svergun, 2007). Typically, several bead models are calculated, aligned, and averaged to generate an average low-resolution model of the molecule. Although this approach can be useful for obtaining coarse-grain shape information, it is not well suited for examining flexible molecules (Bernadó, 2010; Hammel, 2012). Though often attached to an ordered protein substrate, glycans may contribute significant flexibility to a system. We tested the general accuracy of ab initio modeling for glycoproteins using two programs, DAMMIN and GASBOR.

Shape reconstruction was carried out for the RNaseA and RNaseB data sets to examine the overall effects of the single glycan chain. Both GASBOR and DAMMIN produced shape envelopes for RNaseA that were convergent with normalized spatial discrepancy (NSD) 0.63 ± 0.044 and 0.72 ± 0.01, respectively (values under 1 are considered to indicate consistency between the individual reconstructions). Docking the NMR structure into the final averaged shape reconstruction showed the expected shape and size for this protein (Figures 3A and 3B). The same approach for the glycosylated RNaseB also resulted in shape reconstructions with good convergence (NSD of 0.62 ± 0.038 and 0.82 ± 0.013). For DAMMIN, the composite, filtered shape envelope had a bulge proximal to the glycan base (Figure 3C). However, the composite GASBOR model showed only an overall larger envelope relative to RNaseA with no clear bulge (Figure 3D). Several of the individual GASBOR models showed glycan-like protrusions; however, the relative feature locations were highly variable. The effects of aligning and filtering the individual models appear to exclude the variable glycan density from the final averaged model. These results echo those reported for a study of Vitronectin, in which the carbohydrates also contributed little to the overall modeled shape (Lynn et al., 2005).

To test how ab initio reconstructions perform for a highly glycosylated system, we ran the same bead modeling approach on the a1AGP SAXS data. The final filtered shape envelope is clearly larger than would be expected for the protein alone; thus, the glycan density seems to be reconstructed at least to some degree (Figures 3E and 3F). Convergence of both GASBOR and DAMMIN ab initio shape reconstructions for a1AGP is weaker than that for RNaseA and RNaseB models (NSD 0.81 ± 0.04 and 1.50 ± 0.082, respectively), and it appears that the glycan positions are highly ambiguous. Therefore, beyond approximate shape of the molecule, little information on the glycan arrangement is present in the bead models for this system.

Last, we applied this approach to the fetuin data set, where it would potentially be the most useful in the absence of any structural data. The consistency between the models was moderate (NSDs 0.89 ± 0.04 and 1.55 ± 0.058 for DAMMIN and GASBOR, respectively). Both reconstruction programs revealed a similar overall elongated morphology with two lobes (Figures 3G and 3H). The best-fitting all-atom model of fetuin (from ensemble 4, χ 1.73) was docked into both composite envelopes. Although our all-atom model of fetuin is speculative, the overall shape is consistent with ab initio models, and therefore it may reflect the overall domain and glycan arrangement present in the solution structure.

In conclusion, even with all of the caveats limiting the ab initio approach for glycoproteins, it generates the expected size and a shape consistent with the protein substructure. For the examples studied here, the glycan positions cannot be effectively visualized. Instead, glycans appear to contribute bulk volume to the shape reconstructions or at best a bulge near the glycan attachment site. Since the scattering contrast of the carbohydrate is significantly different from that of the protein, small angle neutron scattering (SANS) with contrast matching should be an effective way to directly resolve glycans (Whitten and Trewhella, 2009). A similar technique is possible for contrast matching with SAXS; however, the high concentrations of solute required for effective contrast matching would render the solution conditions potentially nonphysiological and perturbing to protein structure (Putnam et al., 2007).

Conclusions

We have demonstrated that the native MW of glycoproteins can be accurately determined from SAXS, provided that the scattering contrast and partial specific volume are properly taken into account. Ab initio modeling can successfully determine the shape of glycoproteins but provides little information about the spatial occupancy of the glycans. For systems with known coordinates for the protein component, an all-atom modeling approach is a more effective way to interpret SAXS data, potentially revealing the positioning of the glycans or regions absent from the template structures. Although nearly all glycoproteins are expected to have some degree of glycan heterogeneity, modeling with the most abundant glycoform appears sufficient for generating structures in agreement with SAXS data. We have recently used this technique to identify the disposition of the large V1/V2 loops of HIV gp120, a highly glycosylated system (Guttman et al., 2012). With the new implementation of AllosMod (Weinkam et al., 2012) and FoXS (Schneidman-Duhovny et al., 2010) servers, ensembles of glycosylated conformations can be generated and fit against the scattering data within minutes, allowing rapid sampling and assessment of a wide range of glycoprotein conformations. The web server can be found at http://salilab.org/allosmod-foxs and has been set up to generate all types of O-linked and N-linked glycans found in mammalian systems. The approach will be valuable as SAXS data will be more readily available with the application of high-throughput methodologies (Hura et al., 2009).

EXPERIMENTAL PROCEDURES

Sample Preparation

Fetuin from calf serum, human a1AGP, and RNaseB from bovine pancreas were purchased from Sigma Aldrich. Lysozyme was from MP Biomedicals (Solon, OH) and bovine RNaseA was purchased from (Worthington Biochemicals). Each of these was repurified by size exclusion in PBS (20 mM sodium phosphate [pH 7.4], 150 mM NaCl, 0.02% sodium azide, 1 mM EDTA). The Fc portion of IgG 17b was prepared by treatment with papain (Thermo Scientific, Rockville, MD) and isolated by size exclusion chromatography. Hemagglutinin from X31 virus particles (Charles River, Wilmington, MA) was cleaved with bromelain (Sigma Aldrich) as described in Compans et al. (1970). Liberated BHA was separated from remaining particles by ultracentrifugation and purified by size exclusion chromatography (Superdex 200) in HBS

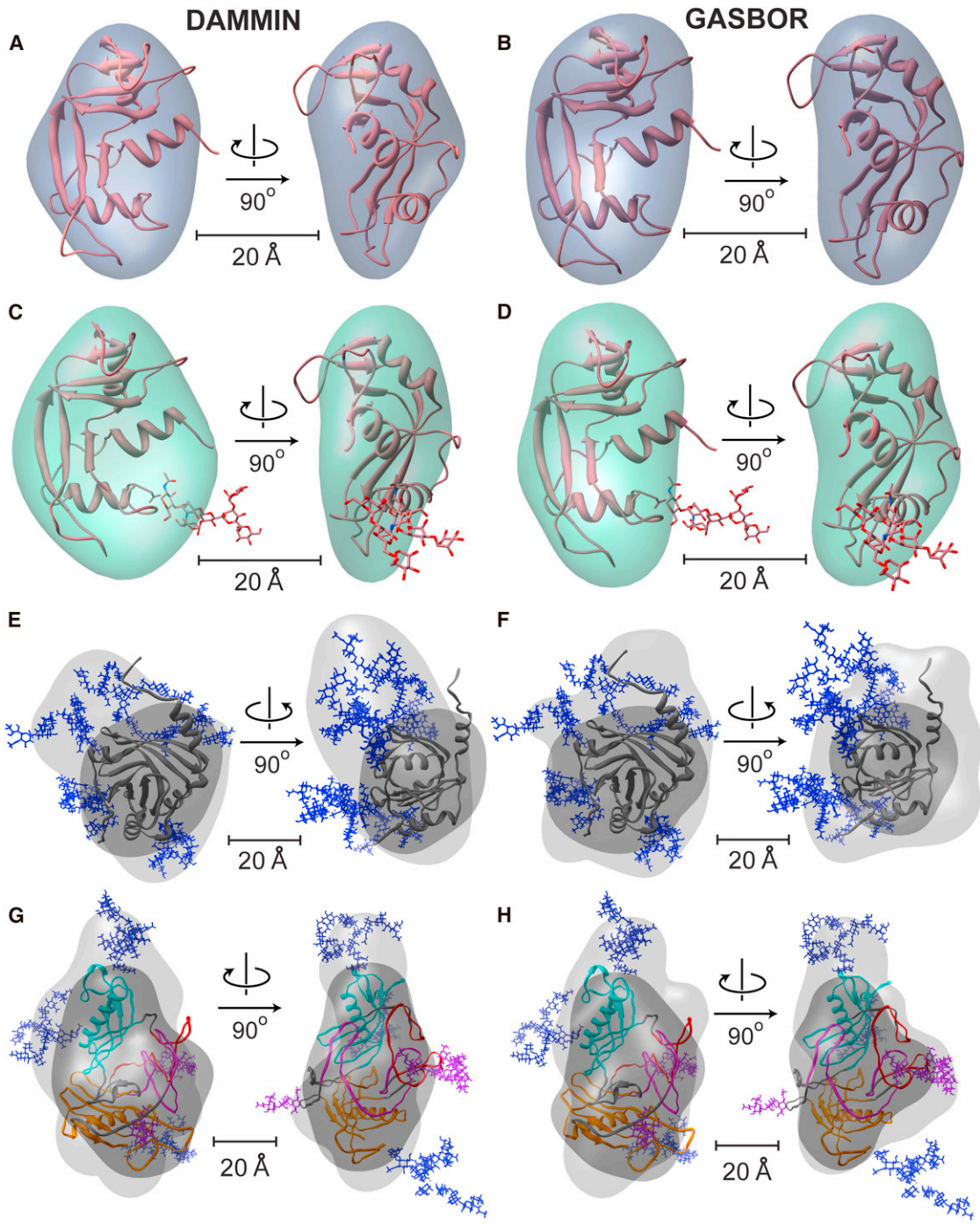

Figure 3. Ab Initio Shape Reconstruction of the Panel of Glycoproteins

Shape reconstructions of RNaseA (A and B), RNaseB (C and D), a1AGP (E and F), and fetuin (G and H) with averaged DAMMIN (left) and GASBOR (right) models shown. Fits are shown in Figure S3. The envelopes were scaled to the measured Porod volume (listed in Table S2). For each protein the best (lowest χ) all-atom model was docked into both enantiomers of the final ab initio shape envelopes using SITUS, only the better enantiomer is shown (Wriggers and Birmanns, 2001).

(A–D) All-atom models RNaseA and RNaseB are shown as cartoons with glycans as sticks.

(E–H) The shape reconstructions calculated for the full glycoprotein (light gray) and only the protein substructure alone (dark gray) are shown for a1AGP (E and F) a1AGP and fetuin (G and H) with N-linked glycans shown as blue sticks. The domains of fetuin are colored as described in Figure 2, with O-linked glycans shown as purple sticks.

See also Figure S3.

(10 mM HEPES [pH 7.4], 150 mM NaCl, 0.02% azide). Proteins were concentrated using microcentricon spin filters immediately before analysis (Amicon 10 kDa cutoff; Millipore, Billerica, MA). Sample purity was verified by SDS-PAGE and N-PAGE.

Dynamic Light Scattering

Dynamic and static light scattering measurements were performed at 20°C on a Dynapro Nanostar (Wyatt Technology, Santa Barbara, CA). Thirty acquisitions of 5 s were collected at a range of concentrations (0.1–1 mg/ml). Buffer viscosity was calculated with SedNterp (Laue et al., 1992). Data were fit using the Wyatt Dynamics analysis software assuming a spherical model and a differential index of refraction (dn/dc) of 0.185.

Small-Angle X-Ray Scattering

SAXS measurements were conducted on Beam Line 4-2 at the Stanford Synchrotron Radiation Laboratory (Smolsky et al., 2007). The focused 11 keV X-ray beam irradiated a thin-wall quartz capillary cell, which contained a sample aliquot, placed at 1.7 m upstream of the Rayonix MX 225HE detector (Evanston, IL). The detector pixel numbers were converted to the momentum transfer $q = 4\pi^*\sin\theta/\lambda$, where $2^*\theta$ is the scattering angle and λ the X-ray wavelength of 1.127 Å, using a silver behenate powder standard placed at the capillary position. A set of 16 consecutive 1 s X-ray exposures was made on each sample aliquot at 15°C with 15 μl oscillations (30 μl injected) between exposures to minimize radiation damage. Protein scattering data were processed by MarParse and scaled for the transmitted beam intensity integrated for each exposure and azimuthally averaged (Smolsky et al., 2007). Exposures were averaged with rejection criteria of 1.3 SDs relative to the initial exposure to omit frames exhibiting radiation damage or other artifacts. Buffer scattering data were processed in the same way and subtracted from corresponding protein scattering data.

All samples were collected over a broad concentration range (0.25–12 mg/ml). Data sets at the highest concentration free of self-association or interparticle interference as reflected by the linearity of the Guinier regions (QRg <1.3) were used for further analysis: lysozyme (1.5 mg/ml), RNaseA (1.5 mg/ml), RNaseB (1.5 mg/ml), a1AGP (0.75 mg/ml), fetuin (1.5 mg/ml), BHA (1.0 mg/ml), and IgG Fc (1.3 mg/ml). The 1D SAXS curves were processed using GNOM to determine approximate Rg and Dmax values for the P(r) plots, the pairwise distance distribution functions (Petoukhov and Svergun, 2007; Svergun, 1991, 1992).

MWs were calculated for each sample using the zero angle scattering intensity "I(0)" relative to water or RNaseA as a standard (Mylonas and Svergun, 2007). Partial specific volumes were calculated as the weighted average of the protein component 0.7425 (Mylonas and Svergun, 2007), and that based on the glycan from the known composition (Lewis and Junghans, 2000). Due to the higher contrast difference ($\Delta\rho$) of glycans, the difference in contrast between the reference protein (RNaseA) and each glycoprotein was factored using the equation:

$$MW_{prot} = \frac{MW_{std} \cdot \left(\frac{I(0)_{prot}}{C_{prot}}\right) \cdot \left(\overline{v}_{std} \cdot \Delta\rho_{std}\right)^2}{\left(\frac{I(0)_{std}}{C_{std}}\right) \cdot \left(\overline{v}_{prot} \cdot \Delta\rho_{prot}\right)^2},$$

where I(0) is the scattering intensity at zero angle of the standard (std) and protein (prot), C is the concentration, and v is the partial specific volume. The $\Delta\rho$ for glycoproteins was calculated as the difference of the weighted average of protein (4.47×10^{23} e/cm^3) and glycan (5.10×10^{23} e/cm^3) from the buffer (3.34×10^{23} e/cm^3). Molecular weights were also calculated on a relative scale using "SAXS MoW" (q up to 0.45 Å^{-1}) (Fischer et al., 2010), as well as from autoporod (Petoukhov et al., 2007).

Size exclusion directly online with SAXS (SEC-SAXS) was used to obtain higher quality data sets for RNaseA, RNaseB, a1AGP, and fetuin. One hundred microliters of each sample at 3–5 mg/ml were injected onto a highres Sepharose 200 column (GE Healthcare) with a flow rate of 50 μl/min in PBS. The flow was passed through the capillary cell and 1 s exposures were collected every 5 s. Rg and I(0) for each frame were batch analyzed using autoRg (Petoukhov et al., 2007), and frames with stable Rg values were merged in primus (Konarev et al., 2003). Absorbance, I(0), and Rg values for the elution profiles are shown in Figure S1. The Rg values near the peak shoulder showed no deviations, thus interparticle repulsion/association were not evident under these conditions.

All-Atom Modeling of Glycoproteins

Ensembles of all-atom structural models were generated in MODELER (Sali and Blundell, 1993) as an adaptation of previous work (Weinkam et al., 2012). Initial structures were generated by addition of glycan chains with ideal geometries derived from CHARMM (Guvench et al., 2008, 2009), followed by a 1 Å randomization of the full atomic coordinates. The structures are first relaxed with conjugate gradient steps followed by molecular dynamics equilibration at 300K followed by several rounds of simulated annealing. The procedure is repeated to produce 200 models for each protein (unless otherwise stated). The intraprotein interactions consist of bonded (for stereochemistry) and nonbonded (for excluded volume and atomic distance) terms used in standard MODELER. The atomic distance term is a sum of harmonic energy functions to restrain pairs of atoms relative to the template structure, and this term is omitted for atoms within two residues of a glycosylation site and for loops or termini absent in the template structure. Intraglycan interactions consist of a bonded term that is a sum of harmonic energy functions over all pairwise atomic distances within each sugar monomer. Bonds between sugar monomers move with proper dihedral rotation using distance and angle restraints. The full procedure can be accessed using our web server at http://salilab.org/allosmod. Details of each glycoprotein template are listed in the following subsections and site-specific glycan types are illustrated in Figure S3.

Next, the theoretical profile from each structural model was calculated and compared against the raw SAXS data using FoXS (Schneidman-Duhovny et al., 2010). As there is still no "best" established method for this type of comparison, we rely on the normalized discrepancy value (χ) to evaluate the goodness of fit (Putnam et al., 2007), as well as on a visual inspection of the fit. Since χ is heavily dependent on the error associated with the measurement, the metric is useful for comparing model SAXS patterns with experimentally measured data within a given set, but less so between sets with different experimental SAXS curves. For the current data sets, we consider a χ of 1.0–1.5 a good fit, whereas χ values above 2 indicate inconsistency between the model and the SAXS data.

We also examine the average and SD among the ten best-fitting models to assess the variance in χ and establish whether or not different modeling strategies result in significantly better fits to the same experimental profile. This type of comparison depends on adequate conformational sampling, which can be gauged by monitoring the best fit given increasingly larger samples. For each data set, the sampling was doubled until the best fit was not significantly better than previous data set (average of ten best fits minus one SD). In all cases except fetuin, 200 models were adequate by this criterion.

The ATSAS software package was used to predict Rg, Dmax, and envelope volumes for each structural model (Svergun et al., 1995). Proper glycan stereochemistry and linkage generated from MODELER were checked with PDBcare (http://www.glycosciences.de/tools/pdb-care/) and glycosidic linkage angles were analyzed by CARP (http://www.glycosciences.de/tools/carp) using the GlycoMaps database (Lütteke et al., 2005). Figures were made with PyMOL (DeLano, 2002) and Chimera (Pettersen et al., 2004).

RNaseB

The sequence of bovine ribonuclease (RNase) (uniprot P61823) was modeled utilizing coordinates from the crystal structure of RNaseB (Protein Data Bank [PDB] 1RBB) (Williams et al., 1987) as well as the NMR structure of RNaseA (PDB 2AAS) (Santoro et al., 1993).

IgG Fc

The IgG Fc portion of Ab 17b was modeled using the coordinates from the crystal structure of IgG Fc (PDB 1FC1) (Deisenhofer, 1981), together with the established disulfide connectivity for IgG. The coordinates for the complex type glycan with a single (β1-4)Gal on the (1-6)Man were included in the alignment. The hinge region (339–343) between the Fc_2 and Fc_3 domains was left unaligned to sample different elbow angles.

a1AGP

The sequence of a1AGP (uniprot P02763) was aligned to the sequence from the recently available unglycosylated crystal structure of a1AGP (PDB 3KQ0 and 3APU) (Schönfeld et al., 2008). Complex type tri and tetra-antennary glycans containing (α2-6) terminal sialic acids were introduced at positions (15, 38, 54, 75, and 85[tetra]) based on site-specific glycan studies (Imre et al., 2005; Treuheit et al., 1992).

Fetuin

The mature sequence (uniprot: P12763) was submitted to I-TASSER structure prediction (Roy et al., 2010; Zhang, 2008) with additional disulfide restraints based on previous findings (Araki et al., 1989). All five I-TASSER models were used as templates for MODELER runs, including triantennary sialylated complex glycans at Asn 81, 138, and 158 and O-linked glycans at S253, T262, S264, and S323, based on previous site specific studies (Edge and Spiro, 1987; Guttman, 1997; Royle et al., 2002) (Figure S2).

Influenza Hemagglutinin

The sequence of the bromelain released ectodomain of influenza hemagglutinin (BHA) (uniprot P03437) was threaded to the crystal structure (PDB 2HMG) (Wilson et al., 1981). High mannose and complex glycans were introduced at positions as shown in Figure S2. The glycosylation at each of the 7 Asn sites was based on predominant glycoforms observed from liquid chromatography-tandem mass spectrometry analysis of peptic fragments (unpublished data), which agree closely with earlier findings (Mir-Shekari et al., 1997).

Ab Initio Shape Reconstructions

The pairwise distance distribution function [P(r)] plot generated from GNOM (Svergun, 1992) was used for ab initio shape reconstruction with DAMMIN v5.3 and GASBOR22i (Svergun, 1999, 2001). The number of dummy residues used for GASBOR calculations was increased to account for the difference in scattering of glycans as described previously (Hammel et al., 2002). The bead models were aligned using SUPCOMB13 with enantiomers considered (Kozin and Svergun, 2001) and spatially filtered using DAMFILT (Volkov and Svergun, 2003). Plots of the resulting fits, Guinier plots, and P(r) plots are presented in Figure S3.

Convergence of the models was assessed by the normalized spatial discrepancy (NSD) showing values under 1, except for the GASBOR models for a1AGP and fetuin with NSDs of 1.50 ± 0.082 and 1.55 ± 0.058, respectively. Further refinement resulted in no improvement in χ values and had no significant effect on the resulting averaged models. GASBOR models showed a greater variation of goodness of fit (χ^2) compared to the DAMMIN models. Therefore, GASBOR models were ranked by goodness of fit (Svergun et al., 2001), and only the top quartile aligned and averaged. The resulting models were converted to a volume envelope using the program pdb2vol within the SITUS2.2 package (Wriggers and Birmanns, 2001). All-atom models were docked into both enantiomers of the volume envelope using the colores module within SITUS2.2. Ab initio reconstructions of nonglycosylated fetuin and a1AGP were performed starting with theoretical scattering pattern of the protein substructure adding the same relative error present in the glycoprotein SAXS data.

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.str.2013.02.004.

ACKNOWLEDGMENTS

We thank Tsutomu Matsui, Lester Carter, and the staff at SSRL for assistance with SAXS data collection. We are grateful to Dina Schneidman-Duhovny and Michal Hammel for guidance with all-atom modeling and Dr. Byron Hetrick and Dr. Seung Joong Kim for insightful discussions. 17b IgG was a kind gift from James Robinson and Peter Kwong. We also wish to thank Thomas Lutteke for generous assistance with glycan structure analysis. This work was supported by NIH grants F32-GM097805, R00-GM080352, R01-GM099989 (to K.K.L.), and R01-GM083960 (to A.S.). Data collection at SSRL was supported by Grant P41-RR001209 from the National Center for Research Resources.

Received: November 3, 2012
Revised: January 22, 2013
Accepted: February 11, 2013
Published: March 5, 2013

REFERENCES

Araki, T., Yoshioka, Y., and Schmid, K. (1989). The position of the disulfide bonds in human plasma alpha 2 HS-glycoprotein and the repeating double disulfide bonds in the domain structure. Biochim. Biophys. Acta *994*, 195–199.

Bernadó, P. (2010). Effect of interdomain dynamics on the structure determination of modular proteins by small-angle scattering. Eur. Biophys. J. *39*, 769–780.

Bernocco, S., Steiglitz, B.M., Svergun, D.I., Petoukhov, M.V., Ruggiero, F., Ricard-Blum, S., Ebel, C., Geourjon, C., Deleage, G., Font, B., et al. (2003). Low resolution structure determination shows procollagen C-proteinase enhancer to be an elongated multidomain glycoprotein. J. Biol. Chem. *278*, 7199–7205.

Bohne-Lang, A., and von der Lieth, C.W. (2005). GlyProt: in silico glycosylation of proteins. Nucleic Acids Res. *33*(Web Server issue), W214–W219.

Brooks, S.A. (2009). Strategies for analysis of the glycosylation of proteins: current status and future perspectives. Mol. Biotechnol. *43*, 76–88.

Compans, R.W., Klenk, H.D., Caliguiri, L.A., and Choppin, P.W. (1970). Influenza virus proteins. I. Analysis of polypeptides of the virion and identification of spike glycoproteins. Virology *42*, 880–889.

Deisenhofer, J. (1981). Crystallographic refinement and atomic models of a human Fc fragment and its complex with fragment B of protein A from Staphylococcus aureus at 2.9- and 2.8-A resolution. Biochemistry *20*, 2361–2370.

DeLano, W.L. (2002). The PyMOL Molecular Graphics System (San Carlos, CA, USA: DeLano Scientific).

Edge, A.S., and Spiro, R.G. (1987). Presence of an O-glycosidically linked hexasaccharide in fetuin. J. Biol. Chem. *262*, 16135–16141.

Fischer, H., de Oliveira Neto, M., Napolitano, H.B., Polikarpov, I., and Craievich, A.F. (2010). Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. J. Appl. Crystallogr. *43*, 101–109.

Frank, M., and Schloissnig, S. (2010). Bioinformatics and molecular modeling in glycobiology. Cell. Mol. Life Sci. *67*, 2749–2772.

Friedman, M.L., Wermeling, J.R., and Halsall, H.B. (1986). The influence of N-acetylneuraminic acid on the properties of human orosomucoid. Biochem. J. *236*, 149–153.

Guttman, A. (1997). Multistructure sequencing of N-linked fetuin glycans by capillary gel electrophoresis and enzyme matrix digestion. Electrophoresis *18*, 1136–1141.

Guttman, M., Kahn, M., Garcia, N.K., Hu, S.L., and Lee, K.K. (2012). Solution structure, conformational dynamics, and CD4-induced activation in full-length, glycosylated, monomeric HIV gp120. J. Virol. *86*, 8750–8764.

Guvench, O., Greene, S.N., Kamath, G., Brady, J.W., Venable, R.M., Pastor, R.W., and Mackerell, A.D., Jr. (2008). Additive empirical force field for hexopyranose monosaccharides. J. Comput. Chem. *29*, 2543–2564.

Guvench, O., Hatcher, E.R., Venable, R.M., Pastor, R.W., and Mackerell, A.D. (2009). CHARMM Additive All-Atom Force Field for Glycosidic Linkages between Hexopyranoses. J. Chem. Theory Comput. *5*, 2353–2370.

Hammel, M. (2012). Validation of macromolecular flexibility in solution by small-angle X-ray scattering (SAXS). Eur. Biophys. J. *41*, 789–799.

Hammel, M., Kriechbaum, M., Gries, A., Kostner, G.M., Laggner, P., and Prassl, R. (2002). Solution structure of human and bovine beta(2)-glycoprotein I revealed by small-angle X-ray scattering. J. Mol. Biol. *321*, 85–97.

Hura, G.L., Menon, A.L., Hammel, M., Rambo, R.P., Poole, F.L., 2nd, Tsutakawa, S.E., Jenney, F.E., Jr., Classen, S., Frankel, K.A., Hopkins, R.C., et al. (2009). Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). Nat. Methods *6*, 606–612.

Imberty, A., and Pérez, S. (2000). Structure, conformation, and dynamics of bioactive oligosaccharides: theoretical approaches and experimental validations. Chem. Rev. *100*, 4567–4588.

Imre, T., Schlosser, G., Pocsfalvi, G., Siciliano, R., Molnár-Szöllosi, E., Kremmer, T., Malorni, A., and Vékey, K. (2005). Glycosylation site analysis of human alpha-1-acid glycoprotein (AGP) by capillary liquid chromatography-electrospray mass spectrometry. J. Mass Spectrom. *40*, 1472–1483.

Ishida, T., and Kinoshita, K. (2007). PrDOS: prediction of disordered protein regions from amino acid sequence. Nucleic Acids Res. *35*(Web Server issue), W460–W464.

Jacques, D.A., and Trewhella, J. (2010). Small-angle scattering for structural biology—expanding the frontier while avoiding the pitfalls. Protein Sci. *19*, 642–657.

Kirschner, K.N., Yongye, A.B., Tschampel, S.M., González-Outeiriño, J., Daniels, C.R., Foley, B.L., and Woods, R.J. (2008). GLYCAM06: a generalizable biomolecular force field. Carbohydrates. J. Comput. Chem. *29*, 622–655.

Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J., and Svergun, D.I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. J. Appl. Crystallogr. *36*, 1277–1282.

Kozin, M.B., and Svergun, D.I. (2001). Automated matching of high- and low-resolution structural models. J. Appl. Crystallogr. *34*, 33–41.

Laue, T.M., Shah, B.D., Ridgeway, T.M., and Pelletier, S.L. (1992). Analytical Ultracentrifugation in Biochemistry and Polymer Science (Cambridge, UK: Royal Society of Chemistry).

Lewis, M.S., and Junghans, R.P. (2000). Ultracentrifugal analysis of molecular mass of glycoproteins of unknown or ill-defined carbohydrate composition. Methods Enzymol. *321*, 136–149.

Linding, R., Jensen, L.J., Diella, F., Bork, P., Gibson, T.J., and Russell, R.B. (2003). Protein disorder prediction: implications for structural proteomics. Structure *11*, 1453–1459.

Lütteke, T., Frank, M., and von der Lieth, C.W. (2005). Carbohydrate Structure Suite (CSS): analysis of carbohydrate 3D structures derived from the PDB. Nucleic Acids Res. *33*(Database issue), D242–D246.

Lynn, G.W., Heller, W.T., Mayasundari, A., Minor, K.H., and Peterson, C.B. (2005). A model for the three-dimensional structure of human plasma vitronectin from small-angle scattering measurements. Biochemistry *44*, 565–574.

McPherson, A., Friedman, M.L., and Halsall, H.B. (1984). Crystallization of alpha 1-acid glycoprotein. Biochem. Biophys. Res. Commun. *124*, 619–624.

Mertens, H.D., and Svergun, D.I. (2010). Structural characterization of proteins and complexes using small-angle X-ray solution scattering. J. Struct. Biol. *172*, 128–141.

Mir-Shekari, S.Y., Ashford, D.A., Harvey, D.J., Dwek, R.A., and Schulze, I.T. (1997). The glycosylation of the influenza A virus hemagglutinin by mammalian cells. A site-specific study. J. Biol. Chem. *272*, 4027–4036.

Mylonas, E., and Svergun, D.I. (2007). Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering. J. Appl. Crystallogr. *40*, s245–s249.

Nawratil, P., Lenzen, S., Kellermann, J., Haupt, H., Schinke, T., Müller-Esterl, W., and Jahnen-Dechent, W. (1996). Limited proteolysis of human alpha2-HS glycoprotein/fetuin. Evidence that a chymotryptic activity can release the connecting peptide. J. Biol. Chem. *271*, 31735–31741.

Pelikan, M., Hura, G.L., and Hammel, M. (2009). Structure and flexibility within proteins as identified through small angle X-ray scattering. Gen. Physiol. Biophys. *28*, 174–189.

Petoukhov, M.V., and Svergun, D.I. (2007). Analysis of X-ray and neutron scattering from biomacromolecular solutions. Curr. Opin. Struct. Biol. *17*, 562–571.

Petoukhov, M.V., Konarev, P.V., Kikhney, A.G., and Svergun, D.I. (2007). ATSAS 2.1 - towards automated and web-supported small-angle scattering data analysis. J. Appl. Crystallogr. *40*, s223–s228.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.

Porod, G. (1982). General Theory. In Small-Angle X-ray Scattering, O. Kratky and O. Gladder, eds. (London: Academic Press), pp. 17–51.

Putnam, C.D., Hammel, M., Hura, G.L., and Tainer, J.A. (2007). X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. Q. Rev. Biophys. *40*, 191–285.

Rambo, R.P., and Tainer, J.A. (2010). Bridging the solution divide: comprehensive structural analyses of dynamic RNA, DNA, and protein assemblies by small-angle X-ray scattering. Curr. Opin. Struct. Biol. *20*, 128–137.

Renouf, D.V., and Hounsell, E.F. (1995). Molecular modelling of glycoproteins by homology with non-glycosylated protein domains, computer simulated glycosylation and molecular dynamics. Adv. Exp. Med. Biol. *376*, 37–45.

Roy, A., Kucukural, A., and Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. Nat. Protoc. *5*, 725–738.

Royle, L., Mattu, T.S., Hart, E., Langridge, J.I., Merry, A.H., Murphy, N., Harvey, D.J., Dwek, R.A., and Rudd, P.M. (2002). An analytical and structural database provides a strategy for sequencing O-glycans from microgram quantities of glycoproteins. Anal. Biochem. *304*, 70–90.

Sali, A., and Blundell, T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. *234*, 779–815.

Santoro, J., González, C., Bruix, M., Neira, J.L., Nieto, J.L., Herranz, J., and Rico, M. (1993). High-resolution three-dimensional structure of ribonuclease A in solution by nuclear magnetic resonance spectroscopy. J. Mol. Biol. *229*, 722–734.

Schneidman-Duhovny, D., Hammel, M., and Sali, A. (2010). FoXS: a web server for rapid computation and fitting of SAXS profiles. Nucleic Acids Res. *38*(Web Server issue), W540–W544.

Schneidman-Duhovny, D., Kim, S.J., and Sali, A. (2012). Integrative structural modeling with small angle X-ray scattering profiles. BMC Struct. Biol. *12*, 17.

Schönfeld, D.L., Ravelli, R.B., Mueller, U., and Skerra, A. (2008). The 1.8-A crystal structure of alpha1-acid glycoprotein (Orosomucoid) solved by UV RIP reveals the broad drug-binding activity of this human plasma lipocalin. J. Mol. Biol. *384*, 393–405.

Shental-Bechor, D., and Levy, Y. (2008). Effect of glycosylation on protein folding: a close look at thermodynamic stabilization. Proc. Natl. Acad. Sci. USA *105*, 8256–8261.

Sickmeier, M., Hamilton, J.A., LeGall, T., Vacic, V., Cortese, M.S., Tantos, A., Szabo, B., Tompa, P., Chen, J., Uversky, V.N., et al. (2007). DisProt: the Database of Disordered Proteins. Nucleic Acids Res. *35*(Database issue), D786–D793.

Slynko, V., Schubert, M., Numao, S., Kowarik, M., Aebi, M., and Allain, F.H. (2009). NMR structure determination of a segmentally labeled glycoprotein using in vitro glycosylation. J. Am. Chem. Soc. *131*, 1274–1281.

Smolsky, I.L., Liu, P., Niebuhr, M., Ito, K., Weiss, T.M., and Tsuruta, H. (2007). Biological small-angle X-ray scattering facility at the Stanford Synchrotron Radiation Laboratory. J. Appl. Crystallogr. *40*, s453–s458.

Svergun, D.I. (1991). Mathematical Methods in Small-Angle Scattering Data Analysis. J. Appl. Crystallogr. *24*, 485–492.

Svergun, D.I. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J. Appl. Crystallogr. *25*, 495–503.

Svergun, D.I. (1999). Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. Biophys. J. *76*, 2879–2886.

Svergun, D., Barberato, C., and Koch, M.H.J. (1995). CRYSOL- a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. J. Appl. Crystallogr. *28*, 768–773.

Svergun, D.I., Petoukhov, M.V., and Koch, M.H. (2001). Determination of domain structure of proteins from X-ray solution scattering. Biophys. J. *80*, 2946–2953.

Treuheit, M.J., Costello, C.E., and Halsall, H.B. (1992). Analysis of the five glycosylation sites of human alpha 1-acid glycoprotein. Biochem. J. *283*, 105–112.

Volkov, V.V., and Svergun, D.I. (2003). Uniqueness of ab-initio shape determination in small-angle scattering. J. Appl. Crystallogr. *36*, 860–864.

Walsh, G. (2010). Biopharmaceutical benchmarks 2010. Nat. Biotechnol. *28*, 917–924.

Weinkam, P., Pons, J., and Sali, A. (2012). Structure-based model of allostery predicts coupling between distant sites. Proc. Natl. Acad. Sci. USA *109*, 4875–4880.

Wen, J., Arakawa, T., and Philo, J.S. (1996). Size-exclusion chromatography with on-line light-scattering, absorbance, and refractive index detectors for studying proteins and their interactions. Anal. Biochem. *240*, 155–166.

Whitten, A.E., and Trewhella, J. (2009). Small-angle scattering and neutron contrast variation for studying bio-molecular complexes. Methods Mol. Biol. *544*, 307–323.

Williams, R.L., Greene, S.M., and McPherson, A. (1987). The crystal structure of ribonuclease B at 2.5-A resolution. J. Biol. Chem. *262*, 16020–16031.

Wilson, I.A., Skehel, J.J., and Wiley, D.C. (1981). Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 A resolution. Nature *289*, 366–373.

Wittmann, V. (2007). Glycopeptides and Glycoprotein; Synthesis, Structure, and Application, *Volume 267* (Berlin, Heidelberg, New York: Springer).

Woods, R.J., Pathiaseril, A., Wormald, M.R., Edge, C.J., and Dwek, R.A. (1998). The high degree of internal flexibility observed for an oligomannose oligosaccharide does not alter the overall topology of the molecule. Eur. J. Biochem. *258*, 372–386.

Wriggers, W., and Birmanns, S. (2001). Using situs for flexible and rigid-body fitting of multiresolution single-molecule data. J. Struct. Biol. *133*, 193–202.

Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. BMC Bioinformatics *9*, 40.