Published in final edited form as:

Cryst Growth Des. 2009 ; 9(1): 348–351. doi:10.1021/cg800509f.

Postcrystallization Analysis of the Irreproducibility of the Human Intrinsic Factor—Cobalamin Complex Crystals

N. Sukumar^{*,†}, F. S. Mathews[‡], M. M. Gordon[§], S. E. Ealick^{†,||}, and D. H. Alpers[§]

NE-CAT, Building 436, Argonne National Laboratory, Argonne, Illinois 60439, Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, Missouri 63110, Division of Gastroenterology, Washington University School of Medicine, St. Louis, Missouri 63110, and Department of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853

[†]Argonne National Laboratory.

[‡]Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine.

[§]Division of Gastroenterology, Washington University School of Medicine.

^{||}Cornell University.

Abstract

Approximately 15% (w/w) of human intrinsic factor (IF) is comprised of carbohydrate side chains, making crystallization problematic. In addition, IF is sensitive to proteolysis. To understand the role of these factors in crystallization, we carried out dynamic light scattering studies and assessed their correlation with crystallization. The packing of the IF—cobalamin complex and the known properties of the protein in solution were also analyzed to explore the irreproducibility of the IF—cobalamin complex crystals and the difficulty in obtaining apo-IF crystals suitable for crystallographic analysis. The results indicate that although glycosylation may in general be inhibitory for crystallization, time-dependent proteolysis appears to play a much more important role in the process of crystallization of IF. Thus, the presence of cobalamin and of domain fragments that can form incomplete dimers lacking one of two β -domains appears to promote the crystallization of IF.

Introduction

Production of suitable crystals is frequently a bottleneck in protein crystallography. This is especially true in the case of glycoproteins because of their peculiar properties, such as the high polarity of solvent-exposed hydroxyl groups, variability in composition, and the flexibility of the carbohydrate linkage.¹ These factors lead to heterogeneity and positional disorder of the attached glycan chains.

The process of crystallization can be explained in terms of a phase diagram,² which can be divided into four regions: (a) the high-precipitation zone, (b) the moderate-precipitation zone (nucleation zone), (c) the metastable zone, for which the crystals are stable and may grow larger without further nucleation, and (d) the undersaturation zone, where the protein remains fully dissolved and will never crystallize. The important stages, nucleation and growth, which correspond to the moderate-precipitation and metastable zones, occur under conditions where

the protein is close to or above saturating concentrations. The interactions of the aqueous solvent with the protein and other solutes affect the role of water in the crystallization process.³ In situations where crystals fail to grow larger (i.e., cannot reach the metastable state from the nucleation state), several seeding techniques such as macroseeding, microseeding, etc., can promote further growth.⁴

Intrinsic factor (IF) is a glycoprotein and one of the three cobalamin (Cbl or vitamin B₁₂) transporting proteins present in mammals.^{5,6} The function of IF is to promote absorption of cobalamin in the ileum by specific receptor-mediated endocytosis.⁷ Recently, the crystal structure of IF⁸ was reported.

Several batches of IF and the IF—Cbl complex were subjected to crystallization experiments, but only occasionally were suitable crystals of the IF—Cbl complex obtained (Figure 1 of the Supporting Information). In spite of repeated attempts, apo-IF crystals could never be grown to an optimal size, although microcrystals formed occasionally. In this paper, an *a posteriori* postcrystallization analysis was carried out, and the results were compared with precrystallization dynamic light scattering (DLS) studies to explore the reasons behind the irreproducible behavior of IF.

Experimental Section

Protein Production

Human IF was previously cloned, sequenced, and expressed in *Pichia pastoris* where the yield is much higher than in the baculovirus expression system.^{9,10} Analysis of the elution fractions by SDS—PAGE demonstrated that the protein was purified to homogeneity and the IF was intact with an apparent molecular mass of ~55 kDa, which includes ~15% carbohydrate (w/w).¹¹ The IF—Cbl complex was prepared via addition of cobalamin to apo-IF in a 3:1 molar ratio followed by dialysis of the IF—Cbl complex against 10 mM sodium monobasic/potassium dibasic phosphate buffer (pH ~6.8) to remove the excess cobalamin.⁸ All of these preparations contained only a single protein band of ~55 kDa.

Precrystallization Dynamic Light Scattering

In DLS measurements, monochromatic light is passed through a sample to study the fluctuations in the scattered light intensity arising from the Brownian motion of the protein molecules in a small volume. Analysis of the data provides an estimate of the translational diffusion coefficient of the protein molecule in solution from which the hydrodynamic radius can be obtained using the Stokes—Einstein relationship.^{12,13} Subsequently, the hydrodynamic molecular mass of the sample can be calculated by comparing the translational diffusion coefficient to those of proteins with known molecular masses.

DLS studies on human IF and the IF—Cbl complex were carried out at 4 and 18 °C using an in-house DynaPro instrument (Protein Solution, Wyatt Technology, Santa Barbara, CA) operating at a wavelength of 655.7 nm. All analyses were carried out using DYNAMICS (Protein Solution, Wyatt Technology). The protein concentration for these measurements ranged from 1 to 5 mg/mL. Both IF and the IF—Cbl complex were filtered with a disposable 0.02 µm pore size Anotop-10 inorganic membrane filter (Whatman) just prior to the DLS measurement.

Results

IF Crystallization

Despite its highly glycosylated nature, extensive crystallization trials of human IF resulted in pink-colored crystals of the IF—Cbl complex using the sitting-drop vapor diffusion method at room temperature.⁸ Optimal conditions were determined to be 10% PEG20000, 100 mM MES (pH 6.0), 20 mM CaCl₂, and 9 mM BaCl₂. The crystals grew as clusters with rounded edges (Figure 1 of the Supporting Information) and formed in 4–6 weeks. For data collection, crystals with approximate dimensions 0.4 mm × 0.15 mm × 0.08 mm were separated out from the clusters using Micro-Tools (Hampton Research, Aliso Viejo, CA). The IF—Cbl crystals were difficult to produce, and over time, several batches of IF, with and without added cobalamin, were prepared for crystallization. Simultaneously, each batch was subjected to a DLS study.

Dynamic Light Scattering

The DLS experiments were conducted prior to crystallization using intact protein samples expressed in *P. pastoris*. The DLS data for apo-IF revealed a monodispersed distribution. For some of the batches, the hydrodynamic radius (R_h) was ~3.9 nm, corresponding to a molecular mass of ~75 kDa (Figure 1a). In spite of its monodisperse state, crystals of apo-IF of suitable size for X-ray crystallographic studies were never obtained from these preparations, and seeding with microcrystals was unsuccessful. In the rest of the batches of apo-IF, the sample aggregated into much higher molecular mass peaks and the R_h value varied from batch to batch. Analysis of the DLS data for the IF—Cbl complex revealed a bimodal distribution. For the batches that gave crystals, a peak at R_h of ~4.8 nm (molecular mass of ~125 kDa) was observed along with another strong peak at very much higher R_h values (~12 nm) (Figure 1b). Since the molecular mass of an IF monomer is ~55 kDa (including ~15% carbohydrate), the R_h ~ 3.9 nm apo-IF peak would be monomeric and the R_h ~ 4.8 nm IF—Cbl peak would be dimeric, assuming that the DLS-derived molecular masses systematically differ from the actual molecular masses by ~20%. This result indicates that some of the IF—Cbl complexes formed dimers while the remainder aggregated to form a much higher molecular mass species. A monomeric state for the intact apo-IF and a dimeric state for the intact IF—Cbl complex are consistent with solution studies previously reported for human IF expressed in *Arabidopsis thaliana* (see below).

Overall IF Structure

The structure of the IF—Cbl complex (Protein Data Bank entry 2PMV) was determined at 2.6 Å resolution.⁸ The IF—Cbl complex is a two-domain protein in which the Cbl molecule binds at the interface of the α - and β -domains. The final model contains two “complete” IF—Cbl molecules each containing an α -domain and a β -domain with a molecule of Cbl bound between them and, surprisingly, two “incomplete” molecules consisting of only the α -domain and no bound Cbl molecule. Each complete molecule ($\alpha\beta$) makes extensive contacts with an incomplete molecule (α only) to form an incomplete dimer ($\alpha_2\beta$).⁸ The two complete molecules lacked significant electron density between residues 274 and 288, and the incomplete molecules terminated at residue 273, suggesting strongly that the crystalline IF—Cbl complex consisted exclusively of a uniform population of α - and β -fragments. The average B factors of the α -domain fragments of complete and incomplete molecules are 46 and 49 Å², respectively, while it is 75 Å² for the β -domain fragments of complete molecules. The higher average B factor of the β -domain compared to that of the α -domain indicates a greater mobility for the β -domain.⁸ A glycosylation site was identified at residue Asn395 of the β -domain, and two molecules of *N*-acetylglucosamine were modeled into the electron density of each chain. Additional incomplete density extended beyond the second sugar, suggesting that more sugar molecules are present.

IF Proteolysis

Despite its resistance to luminal proteases in vivo,^{11,14} wild-type IF can be cleaved in the ileum into independent α - and β -domains by cathepsin L, an intracellular protease,¹⁵ and in vitro by plant proteases during expression of *A. thaliana*-expressed IF.¹⁶ The SDS—PAGE analysis carried out at the end of the purification for the IF expressed in *A. thaliana* indicated that approximately one-third of protein was cleaved into fragments of ~20 and ~30 kDa while the remaining two-thirds of the protein was intact with a molecular mass of 50 kDa.¹⁷ Biochemical studies on *A. thaliana*-expressed IF in solution indicated that independent α - and β -domain fragments of IF can still associate and form a complex when bound to Cbl.¹⁶ In addition, full-length molecules were found to form dimers in solution, although the cleaved components did not. In this study, IF was expressed in *P. pastoris*, which is known to contain proteases, and the IF—Cbl crystals took ~4–6 weeks to grow. The long incubation time at the high protein concentration (~10 mg/mL) during crystallization apparently resulted in cleavage of IF into its α - and β -domains in those preparations. The inability to detect intact IF molecules in the crystal structure of IF suggests that fairly extensive proteolysis occurred during crystal formation, even with only low levels of protease present.⁸ Also, it has been shown that the α -domain exhibits a lower affinity for Cbl compared to the β -domain, although the α -domain is essential for the retention of bound Cbl.¹⁸ During IF maturation, Cbl appears to bind first to the β -domain, after which the ligand binding site is formed by association of the α -domain.

Crystal Packing and Surface Contacts

A detailed analysis of the IF—Cbl complex (Figure 2) reveals that the environment of the α -domain fragments of the incomplete molecules within the crystal lattice is crowded, and incorporation of the β -domains would be prevented by severe lattice contacts. On the other hand, both domains of the complete IF molecules are easily accommodated in the lattice and the glycosyl moieties covalently linked to Asn395 of the complete molecule are pointed outward, making no specific intermolecular interactions. As observed in other glycoproteins,¹ the glycosyl chains are highly disordered, implying a high flexibility. While only two sugar residues were modeled for each β -subunit, a large volume of very weak density near the glycosylation site indicated the possibility that more sugars could be accommodated. As a result, a substantial fraction of the potential crystal contact-forming protein surface is obscured.

Table 1 lists the surface area buried upon crystal formation. The crystal contacts of the IF—Cbl complex were analyzed using AREAIMOL¹⁹ and CCP4mg²⁰ of the CCP4 package.²¹ The analysis shows that the α -domains of the incomplete molecules are the dominant components of the packing interface, contributing ~54% of all surface contact area compared to ~36–40% contributed by the α -domain of complete molecules. However, the rmsd between equivalent atoms of α -domains of complete and incomplete molecules is ~0.48 Å,⁸ which indicates that both the domains adopt similar conformations. In contrast, the β -domains of the complete molecules make few interactions and contribute only ~4–10% of the total contact area. This is consistent with the crystal packing where the β -domains are adjacent to a large void between symmetry-related molecules that could accommodate the additional bulk of the covalently bound polysaccharide at position Asn395.

Within the asymmetric unit, the α -domains of complete and incomplete molecules, related by a local 2-fold axis, are stacked on one another. The buried surface area is ~1200 Å² per molecule, and 24 hydrogen bonds exist between the domains. The two complete molecules in the asymmetric unit make contact through their α -domains; the buried surface area between them is ~500 Å² per molecule with one connecting salt bridge. With one exception, the contacts of the two dimers with the other symmetry-related molecules are similar in nature, with buried surface areas of ~300–500 Å² and one or two connecting hydrogen bonds. The exception is the contacts formed between α -domains of incomplete molecules with seven or eight hydrogen

bonds per contact, although the buried surface area is still $\sim 400\text{--}500 \text{ \AA}^2$ per molecule, similar to those of other intermolecular contacts.

Discussion

Uses of Dynamic Light Scattering

DLS is a useful tool for understanding the crystallization process, especially for finding a suitable nucleation condition. Interestingly, in the case of IF, the monodispersed apo-IF did not yield crystals, while the polydispersed IF—Cbl complex did. This indicates that the initial DLS study alone may not have been sufficiently predictive of success because the state of the protein in the crystallization droplet changed as a result of proteolysis during the several weeks of crystal growth. In the case of apo-IF, while the SDS—PAGE analysis carried out after purification always indicated intact protein with a molecular mass of $\sim 55 \text{ kDa}$, proteolyzed domains were produced in sufficient but small amounts in many preparations to be detected by protein gels. Presumably, this proteolysis continued during crystal formation to different degrees over a period of time. The packing analysis on the IF—Cbl complex confirms this fact. As the concentration of proteolyzed domains increased, the state of the protein could have changed from monodisperse to polydisperse, especially in the absence of Cbl, which otherwise could join with the cleaved α - and β -domains to form a complex. This fairly complete proteolysis and the presence of N-linked glycosylation sites may have contributed to the inability to obtain crystals of apo-IF. A systematic DLS study using samples from the drops in the crystallization setup at regular time intervals might provide a better indication of the state of the protein during crystallization; however, such experiments would be difficult to perform because of the very small volumes involved.

Factors Affecting IF Crystallization

In the presence of Cbl, the separated α - and β -domains can associate to form a complex. In almost every batch of IF—Cbl complex with a DLS peak at $\sim 125 \text{ kDa}$, small crystals were observed, indicating that nucleation had occurred. Given an equilibrium between Cbl-bound and unbound species, subunits with both the α - and β -domains are stabilized by one set of crystal packing interactions, while subunits with only the α -domain present are stabilized by a different set. As the crystals grow, a supply of free α -domains must be available for incorporation into the crystal lattice. This suggests that the main limitation to growing larger crystals lies in the metastable state where nucleation rarely occurs and crystals normally grow to larger sizes.² To form an ordered crystal lattice, nucleation and growth would require selective incorporation of the Cbl-bound species. In this study involving the IF—Cbl complex, seeding was not successful, suggesting the presence of both Cbl-bound and unbound species under growth conditions.

The large dimer interface and hydrogen bonding network between IF—Cbl molecules in the crystal structure appear to stabilize a dimeric molecule in solution under the initial crystallization conditions as indicated by the DLS measurements. Subsequent proteolysis of the linker between domains would reduce the stability of the full-length cleaved dimers, as indicated by earlier studies of proteolytically separated domains in solution.¹⁶ This suggests a plausible model for crystal growth of the IF—Cbl complex that is consistent with the observations presented here. The packing arrangement observed in the IF—Cbl crystals does not favor growth or nucleation of cleaved full-length molecules because of the severe lattice contacts created by the additional β -domain. However, nucleation could occur if several cleaved molecules minus their β -domains come together to form a nascent lattice as we found in the crystal structure. Crystal growth could then follow through incorporation into the nucleated crystal of additional complete molecules that dissociate spontaneously from dimers. Such growth would be favored by the lattice forces that stabilize binding of the isolated α -

domains of the incomplete molecules through a greater surface area and a number of hydrogen bonds larger than the number available for other intermolecular contacts in the crystal lattice. It is possible that glycosylation of the β -domain poses an additional complication due to chemical heterogeneity and conformational flexibility of the carbohydrate moiety. However, in the case of the IF—Cbl complex, though the packing analysis indicates that a large number of sugars could be present, glycosylation per se appears to be a less important factor in inhibiting crystallization.

Concluding Remarks

Time-dependent proteolysis such as that observed for IF may have general applicability to protein crystallization when there is a problem obtaining crystals of optimum size. Moreover, the importance of proteolysis may be overlooked when other problems such as glycosylation exist. For example, although it was observed that one-third of IF expressed in *A. thaliana* was cleaved by the end of the purification, the difficulty in obtaining crystals was thought to be due to glycosylation.^{22,23} The results of this study are offered as a reminder that proteolysis may be an important factor in cases of problematic crystal growth.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was carried out at the NE-CAT facility at the Advanced Photon Source (APS), supported by Grant RR-15301 from the National Center for Research Resources at the National Institutes of Health (NIH) and Grants DK-33487 (D.H.A.), DK-56342 (D.H.A.), and GM20530 (F.S.M.) from the NIH. Use of the APS is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Science, under Contract DE-AC02-06CH11357.

References

- (1). Bush CA, Martin-Pastor M, Imberty A. Annu. Rev. Biophys. Biomol. Struct 1999;28:269–293. [PubMed: 10410803]
- (2). Chayen NE. Curr. Opin. Struct. Biol 2004;14:577–583. [PubMed: 15465318]
- (3). Wiener MC. Curr. Opin. Colloid Interface Sci 2001;6:412–419.
- (4). Stura, EA. Crystallization of Nucleic Acids and Proteins: A Practical Approach. Vol. 2nd. Ducruix, A.; Giege, R.; Hames, BD., editors. Oxford University Press; New York: 1999. p. 177-208. The Practical Approach Series
- (5). Banerjee, R. Chemistry and Biochemistry of B12. John Wiley and Sons; New York: 1999.
- (6). Brown KL. Chem. Rev 2005;105:2075–2149. [PubMed: 15941210]
- (7). Kapadia CR, Serfilippi D, Voloshin K, Donaldson RM Jr. J. Clin. Invest 1983;71:440–448. [PubMed: 6826716]
- (8). Mathews FS, Gordon MM, Chen Z, Rajashankar KR, Ealick SE, Alpers DH, Sukumar N. Proc. Natl. Acad. Sci. U.S.A 2007;104:17311–17316. [PubMed: 17954916]
- (9). Wen J, Kinnear MB, Richardson MA, Willetts NS, Russell-Jones GJ, Gordon MM, Alpers DH. Biochim. Biophys. Acta 2000;1490:43–53. [PubMed: 10786616]
- (10). Gordon MM, Russell-Jones G, Alpers DH. Methods Enzymol 1997;281:255–261. [PubMed: 9250989]
- (11). Dieckgraefe BK, Seetharam B, Banaszak L, Leykam JF, Alpers DH. Proc. Natl. Acad. Sci. U.S.A 1988;85:46–50. [PubMed: 3422425]
- (12). Ferre-D'Amare, AR.; Burley, SK. Methods in Enzymology. Carter, CW., Jr.; Sweet, RM., editors. Vol. 276. Academic Press; New York: 1997. p. 157-166.

- (13). Borgstahl, GEO. Methods in Molecular Biology: Macromolecular Crystallization Protocols Volume I. Preparation and Crystallization of Macromolecules. Double, S., editor. Vol. 363. Humana Press Inc.; Totowa, NJ: 2007. p. 109-129.
- (14). Mancia F, Keep NH, Nakagawa A, Leadlay PF, McSweeney S, Rasmussen B, Bosecke P, Diat O, Evans PR. Structure 1996;4:339–350. [PubMed: 8805541]
- (15). Gordon MM, Howard T, Becich MJ, Alpers DH. Am. J. Physiol 1995;268:G33–G40. [PubMed: 7840205]
- (16). Fedosov SN, Fedosova NU, Berglund L, Moestrup SK, Nexø E, Petersen TE. Biochemistry 2004;43:15095–15102. [PubMed: 15554717]
- (17). Fedosov SN, Laursen NB, Nexø E, Moestrup SK, Petersen TE, Jensen EO, Berglund L. Eur. J. Biochem 2003;270:3362–3367. [PubMed: 12899693]
- (18). Fedosov SN, Fedosova NU, Berglund L, Moestrup SK, Nexø E, Petersen TE. Biochemistry 2005;44:3604–3614. [PubMed: 15736970]
- (19). Lee B, Richards FM. J. Mol. Biol 1971;55:379–400. [PubMed: 5551392]
- (20). Potterton L, McNicholas S, Krissinel E, Gruber J, Cowtan K, Emsley P, Murshudov GN, Cohen S, Perrakis A, Noble M. Acta Crystallogr 2004;D60:2288–2294.
- (21). Collaborative Computational Project Number 4. Acta Crystallogr 1994;D50:760–763.
- (22). Wuerges J, Garau G, Geremia S, Fedosov SN, Petersen TE, Randaccio L. Proc. Natl. Acad. Sci. U.S.A 2006;103:4386–4391. [PubMed: 16537422]
- (23). Wuerges J, Geremia S, Randaccio L. Biochem. J 2007;403:431–440. [PubMed: 17274763]

Figure 1. DLS analysis of IF. Representative diagram of the distribution of the hydrodynamic radius vs amplitude of (a) IF and (b) the IF—Cbl complex. The mean of the population is represented by a pink square.

Figure 2.

Crystal packing of the IF—Cbl surface diagram in the unit cell viewed down the b axis. The surface diagram of α - and β -domains of complete molecules of IF are colored gold and red, respectively, while the incomplete molecules are colored blue. The cobalamin and N -acetylglucosamine residues are shown in ball-and-stick format and colored green and tan, respectively. This diagram was produced using CCP4mg.²⁰

Table 1**Packing and Surface Contact Analysis**

total solvent-accessible area of crystallographic molecules ($\text{\AA}^2$)	53968
total solvent-accessible area by including the symmetry-related atoms ($\text{\AA}^2$)	49573
difference in area due to the presence of the symmetry-related atoms ($\text{\AA}^2$)	-4395
area difference in individual domains ($\text{\AA}^2$)/percentage of total contact surface ^a	
α -domain of complete molecule (molecule A)	-771/17.6%
β -domain of complete molecule (molecule A)	-228/5.2%
α -domain of incomplete molecule (molecule A)	-1177/26.8%
α -domain of complete molecule (molecule B)	-916/20.8%
β -domain of complete molecule (molecule B)	-102/2.3%
α -domain of incomplete molecule (molecule B)	-1200/27.8%
area difference between incomplete dimer ($\text{\AA}^2$)	500 (approximately)
no. of hydrogen bonds between α -domains of incomplete dimer	24
no. of hydrogen bonds between symmetrically related α -domains of incomplete molecule	8

^aTwo crystallographically independent molecules present in asymmetric units.