Cyclic enterobacterial common antigen: Potential contaminant of bacterially expressed protein preparations

Paul J.A. Erbel[a], Ronald Seidel[b], Scott E. Macintosh[c], Lisa N. Gentile[c,*], Juan C. Amor[d], Richard A. Kahn[d], James H. Prestegard[d], Lawrence P. McIntosh[c] & Kevin H. Gardner[a,**]

[a]*Departments of Biochemistry and Pharmacology, University of Texas Southwestern Medical Center, Dallas TX 75390-9038, U.S.A.;* [b]*Complex Carbohydrate Research Center, University of Georgia, Athens GA 30602-4712, U.S.A.;* [c]*Department of Biochemistry and Molecular Biology, Department of Chemistry, and the Protein Engineering Network of Centres of Excellence, University of British Columbia, Vancouver BC V6T 1Z3, Canada;* [d]*Department of Biochemistry, Emory University School of Medicine, Atlanta GA 30322, U.S.A.*

Abstract

We have previously reported the identification of the cyclic enterobacterial common antigen (ECA$_{CYC}$) polysaccharide in *E. coli* strains commonly used for heterologous protein expression (PJA Erbel et al., *J. Bacteriol.* **185** (2003): 1995). Following this initial report, interactions among several NMR groups established that characteristic *N*-acetyl signals of ECA$_{CYC}$ have been observed in ^{15}N-^{1}H HSQC spectra of samples of various bacterially-expressed proteins suggesting that this water-soluble carbohydrate is a common contaminant. We provide NMR spectroscopic tools to recognize ECA$_{CYC}$ in protein samples, as well as several methods to remove this contaminant. Early recognition of ECA-based NMR signals will prevent time-consuming analyses of this copurifying carbohydrate.

Introduction

Despite advances in equipment and methodology, most modern biological NMR spectroscopic techniques require samples with relatively large quantities ($\geq$ milligrams) of protein for optimal sensitivity. Protein preparation on this scale is most often achieved using bacterial expression systems, as the necessary molecular biology techniques are easily implemented and this route can economically generate samples that are enriched with NMR-relevant isotopes such as ^{2}H, ^{13}C and ^{15}N. Once proteins are expressed in this manner, they are usually purified away from

a biochemically-complex cell lysate mixture using a combination of chromatographic methods. During these procedures, the purity of a sample is typically monitored using only protein-directed techniques (e.g., SDS-PAGE gel electrophoresis, absorbance at 280 nm, or activity assays), neglecting the detection of contaminants consisting of other classes of biomolecules.

A shortcoming of this route has been highlighted by recent experiences with a common bacterial carbohydrate contaminant present in different protein samples prepared in each of our laboratories. For two of us (PJAE, KHG), our attention was drawn to this problem by the presence of unusual peaks in the amide region of the ^{15}N-^{1}H HSQC spectra of samples of the C-terminal PAS domain of human HIF-2α (HIFd) expressed in *E. coli* (Erbel et al., 2003). Subsequent purification and analysis revealed that these signals originated from the N-acetyl moieties of a water-

---

*Present address: Department of Chemistry, Western Washington University, Bellingham, WA 98225, U.S.A.

**To whom correspondence should be addressed. Department of Biochemistry, University of Texas Southwestern Medical Center 5323 Harry Hines Blvd., Dallas, TX 75390-9038, U.S.A. E-mail: Kevin.Gardner@UTSouthwestern.edu

soluble form of the enterobacterial common antigen (ECA$_{CYC}$), a cyclic polysaccharide that had not been previously characterized in *E. coli*. After completing this analysis, interactions among our groups established that each of us had observed ECA$_{CYC}$ signals in preparations of several different proteins suggesting that this type of contamination may be observed more broadly than each of us had initially anticipated. In light of this, we present a summary of our experiences to provide the community with tools to both recognize and eliminate ECA$_{CYC}$ contamination of bacterially-expressed protein preparations.

ECA background

ECA has been identified in several forms, all of which contain a common carbohydrate moiety comprised of the trisaccharide repeat unit ->3)-α-D-Fuc4NAc-(1->4)-β-D-ManNAcA-(1->4)-α-D-GlcNAc-(1-> where Fuc4NAc, ManNAcA, and GlcNAc respectively denote 4-acetamido-4,6-dideoxy-D-galactose, *N*-acetyl-D-mannosaminuronic acid, and *N*-acetyl-D-glucosamine (Figure 1A) (Lugowski et al., 1983). The C-6 position of GlcNAc is commonly found to be at least partially O-acetylated as well (Bruix et al., 1995; Erbel et al., 2003; Lugowski et al., 1983; Staaf et al., 2001). The best-characterized form of ECA, ECA$_{PG}$, is a glycolipid found at the cell surface of all gram-negative enteric bacteria. In ECA$_{PG}$, individual ECA polysaccharide chains are covalently linked to diacylglycerolphosphate via a glycosidic linkage to the terminal GlcNAc residue, anchoring these chains into the outer cell membrane (Rick and Silver, 1996).

While membrane-associated ECA$_{PG}$ can be easily removed from protein preparations by centrifugation, an alternative water-soluble, cyclic form (ECA$_{CYC}$) poses a more troublesome issue for purification of soluble proteins. ECA$_{CYC}$ has been found in cell extracts of several bacterial strains (Rick and Silver, 1996), including the *E. coli* B and K-12 strains that are commonly used for heterologous gene expression (Erbel et al., 2003). In contrast to ECA$_{PG}$, ECA$_{CYC}$ lacks the aglycan and instead links the reducing end to Fuc4NAc, forming a 2.4 kDa cyclic carbohydrate containing four trisaccharide repeats in *E. coli* (Figure 1) as shown by negative-ion mode ESI-MS (Erbel et al., 2003). Both ECA$_{PG}$ and ECA$_{CYC}$ appear to be quite abundant in bacteria, constituting a combined 0.4% of the cellular dry weight of *E. coli* K-12 (Erbel et al., 2003).

NMR characteristics of ECA$_{CYC}$

Given that the ^{15}N-^{1}H signals from the N-acetyl moieties of ECA$_{CYC}$ all fall within the spectral region characteristic of protein amides, the presence of this contaminant may initially be attributed to a polypeptide species. However, ECA$_{CYC}$ contamination can be established through several routes: (i) The trisaccharide repeat yields a distinct pattern of three ^{15}N-^{1}H signals (Figure 1B), which may be narrower or more intense than the corresponding protein peaks (suggesting ECA$_{CYC}$ concentrations in a final sample of at least $\sim$ 10 μM given typical protein concentrations of $\geq$ 100 μM and recognizing the effects of differential relaxation). In some cases, we have observed multiple peaks for at least two of these signals, which we attribute to sample heterogeneity (e.g., differential levels of O-acetylation) (Erbel et al., 2003). (ii) The ^{15}N-^{1}H signals of this oligosaccharide have relaxation properties indicative of a $\sim$ 2 kDa molecule (i.e., ^{15}N T$_1$ $\sim$ 650 ms, T$_2$ $\sim$ 390 ms, and ^{15}N\{^{1}H\} NOE $\sim$ −0.2 at 30 °C). (iii) Triple resonance experiments, such as CBCA(CO)NH, recorded on U-^{15}N, ^{13}C ECA$_{CYC}$ samples yield correlations from the amides to methyl ($\sim$ 25 ppm) and carbohydrate ($\sim$ 70 and 100 ppm) carbons with shifts distinct from those expected for a protein (Erbel et al., 2003). (iv) ^{13}C-labeled ECA$_{CYC}$ can be readily differentiated from a protein by ^{13}C-^{1}H HSQC spectroscopy due to the characteristic chemical shifts of its carbohydrate groups, particularly the anomeric carbon signals near 100 ppm. An overview of the chemical shift information of *E. coli* ECA$_{CYC}$ is shown in Table 1.

Common aspects of ECA$_{CYC}$ contamination of bacterially-expressed proteins

ECA$_{CYC}$ was found in preparations of HIFd (residues 240–350, MW 13.2 kDa, pI$_{calc}$ = 6.2), two Pointed domain-containing fragments of the Ets transcription factor family member GA-binding protein α, denoted as GABPα$^{(138−254)}$ (residues 138–254, MW 13.7 kDa, pI$_{calc}$ = 4.9) and GABPα$^{(168−254)}$ (residues 168–254, MW 10.3 kDa, pI$_{calc}$ = 5.2) (MacIntosh, 2001; Mackereth et al., 2002) and an N-terminally truncated form of the human small GTPase ADP-Ribosylation Factor 1 (Δ17Arf1, residues 18–181; MW 18.8 kDa, pI$_{calc}$ = 5.6) (Kahn et al., 1992). The purification schemes for all proteins were based on anion exchange
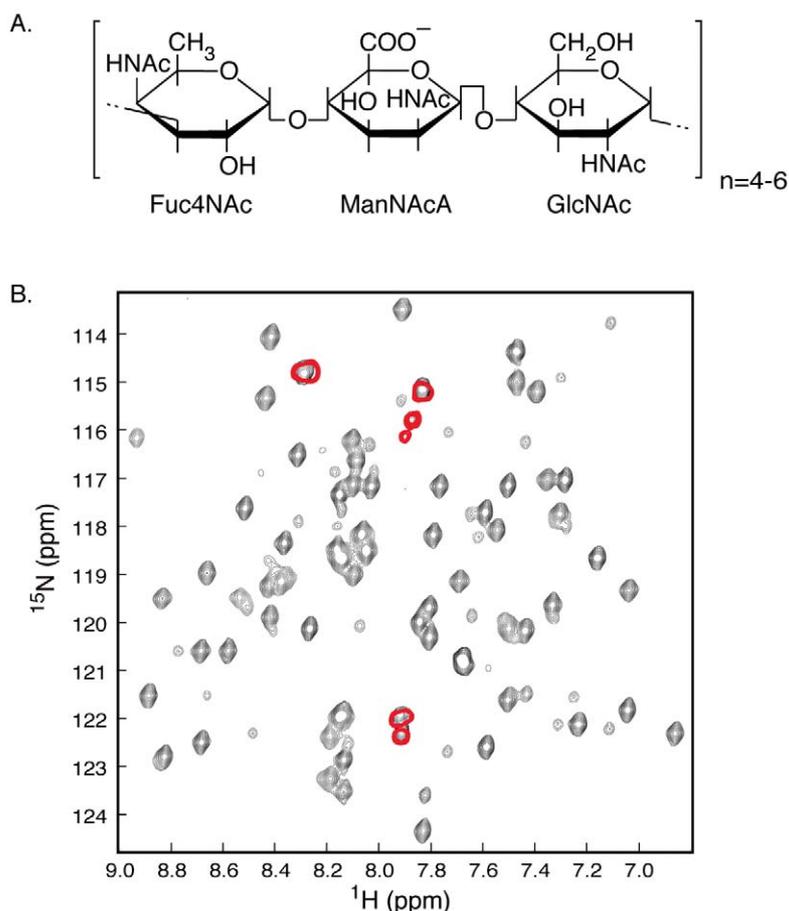
Figure 1. (A) Covalent structure of trisaccharide repeat of ECA. For *E. coli* ECA$_{CYC}$, n = 4, although n = 5 and 6 forms have been characterized in other strains. Abbreviations for the monosaccharide units are: Fuc4NAc (4-acetamido-4.6-dideoxy-α-D-galactose), ManNAcA (N-acetyl-β-D-mannosaminuronic acid) and GlcNAc (N-acetyl-α-D-glucosamine). (B). An example of ECA$_{CYC}$ contamination as observed by ^{15}N-^{1}H HSQC spectroscopy. Black contours: Spectrum of Q-Sepharose purified GABPα$^{(168-254)}$; red contours: Spectrum of purified ^{15}N-labeled ECA$_{CYC}$. All spectra are recorded at 500 MHz.

chromatography (Source Q15, Fast Flow Q Sepharose and High Q resin) and a subsequent gel filtration step (Superdex75 and Sephacryl S-100).

Unfortunately, the process of separating ECA$_{CYC}$ from expressed proteins using these approaches is made difficult by the charge and unusual shape of this carbohydrate. Anion exchange procedures are complicated by the negative charge of the ManNAcA units in ECA$_{CYC}$, giving it a modest binding affinity for Q-type resins (elution < 200 mM NaCl). This is further complicated by the heterogeneity of ECA$_{CYC}$ (e.g. O-acetylation), producing a broad elution profile and correspondingly poor separation. While this straightforwardly explains the copurification of ECA$_{CYC}$ through this step, the failure of size exclusion chromatography to separate a 2.4 kDa carbohydrate from several proteins larger than 10 kDa is rather surprising. Other methods based on molecular size, including dialysis and centrifugal/pressure-driven ultrafiltration devices with filters up to 10 kDa, were similarly unsuccessful at removing ECA$_{CYC}$ from our samples. We suggest that these difficulties are likely caused by the unusual shape of ECA$_{CYC}$, which adopts a disc-like shape with a diameter of approximately 20 Å on the long axis (Staaf et al., 2001).

Methods to eliminate ECA$_{CYC}$

Our groups have independently developed two general approaches to remove ECA$_{CYC}$ from the protein preparations, the first of which is to reduce or eliminate biosynthesis of this species. The most robust

Table 1. ^{1}H- and ^{13}C-NMR chemicals shifts (δ in ppm) for *E. coli* ECA$_{CYC}$

	GlcNAc	ManNAcA	Fuc4NAc
H-1	4.97	4.86	5.12
H-2	3.98	4.53	3.81
H-3	3.94	4.06	4.07
H-4	3.77	3.78	4.33
H-5	4.04	3.82	4.26
H-6A	3.84		
H-6B	3.79		
CH$_3$			1.05
NAcH	2.02	2.08	2.08
OAcH	2.08		
C-1	94.1	99.5	101.9
C-2	54.1	54.0(58.2)	68.6
C-3	70.4(68.8)	73.3	72.6
C-4	79.7	78.8	50.7(51.6)
C-5	71.3	78.7	67.6(67.2)
C-6	61.1		
CH$_3$			16.6
NAcH	23.2	23.1	23.0
OAcH	17.7		
H(N)[a]	7.99	8.30(8.36)	7.87(7.91,7.98)
[15]N(H)[a]	122(122.4)	114.8	115.2(115.8,116.0)
[13]C(O)[b]	175.0	176.4	175.2(175.6,174.8)

[1]H reference DSS = 0.007 ppm.
[13]C reference DSS = −1.84 ppm.
[a]Measured in a ^{15}N labeled sample H$_2$O.
[b]Measured in a ^{13}C labeled sample in the presence of HIFd in H$_2$O at 30 °C.

way to achieve this is to use mutant bacterial strains that are defective in ECA biosynthesis. Three such strains, containing mutations in the *wecA*, *wecF*, and *wecG* genes, have been constructed in a BL21(DE3) background, allowing in each the expression from T7-promoter driven plasmids (Erbel et al., 2003). These strains do not synthesize ECA$_{PG}$ or ECA$_{CYC}$, but still show normal growth and protein overexpression behavior and are available on request.

As an alternative strategy for strains already in common use among NMR laboratories, one can also reduce the amount of ECA$_{CYC}$ present by the growth conditions used for protein overexpression. Spectra of Δ17Arf1 purified from cultures after 3, 7 or 14–16 h post induction at 37 °C (Figure 2) show that the amount of ECA$_{CYC}$ appears to roughly correlate with the length of induction. This trend does not appear to be strictly linear, given the dramatic increase in ECA$_{CYC}$ abundance in the 14 h spectra compared

to either of the earlier spectra. Additionally, comparison of spectra obtained from Δ17Arf1 samples expressed in either M9 or MOPS media (Neidhardt et al., 1974) supplemented with 1g/L [15]NH$_4$Cl, 4g/L glucose and a multivitamin mix (BRL), show that significantly higher quantities of ECA$_{CYC}$ were found in the M9 sample (Figure 2). Therefore we conclude that the choice of media and the length of time for protein expression have an effect on the amount of ECA$_{CYC}$ present in the final sample.

The second general approach relies on improved chromatographic techniques to ensure better separation of protein from ECA$_{CYC}$. In the case of HIFd, anion exchange chromatography in buffers with lower ionic strength (achieved by replacing sodium phosphate with Tris) specifically strengthened the affinity of HIFd for Source 15Q resin without affecting the binding of ECA$_{CYC}$. This allowed the separation of these molecules from one another (Erbel et al., 2003). Alternatively, GABPα$^{(138−254)}$ could be separated from ECA$_{CYC}$ by an additional reverse phase HPLC step using a Perspective Biosystems Porous RP 1080 column and a linear gradient of increasing acetonitrile in water with 0.1% TFA. Where practical, affinity purification methods also offer a general solution to removing ECA$_{CYC}$. For example, a His$_6$-tagged version of GABPα$^{(168−254)}$ is purified free of ECA$_{CYC}$ by passage through Ni^{2+}-charged NTA columns.

It is important to note that the chemical shifts of GABPα fragments and HIFd are identical before and after removal of ECA$_{CYC}$ by HPLC or metal-affinity chromatography, indicating that the carbohydrate does not specifically bind these proteins with any appreciable affinity.

Discussion

Here we have presented several independent examples where ECA$_{CYC}$ was found as an impurity in protein samples obtained from overexpression in *E. coli*. This suggests that ECA$_{CYC}$ could be a rather common contaminant of many proteins obtained from bacterial expression systems. Another literature report (Bruix et al., 1995) also described a linear and lipid-free form of ECA found in a sample of the chemotactic protein CheY that does not correspond to a known ECA form or biosynthetic intermediate. In retrospect, this might have originated from degradation of ECA$_{CYC}$ during purification, since these investigators identified a free reducing terminal Fuc4NAc residue in their purified
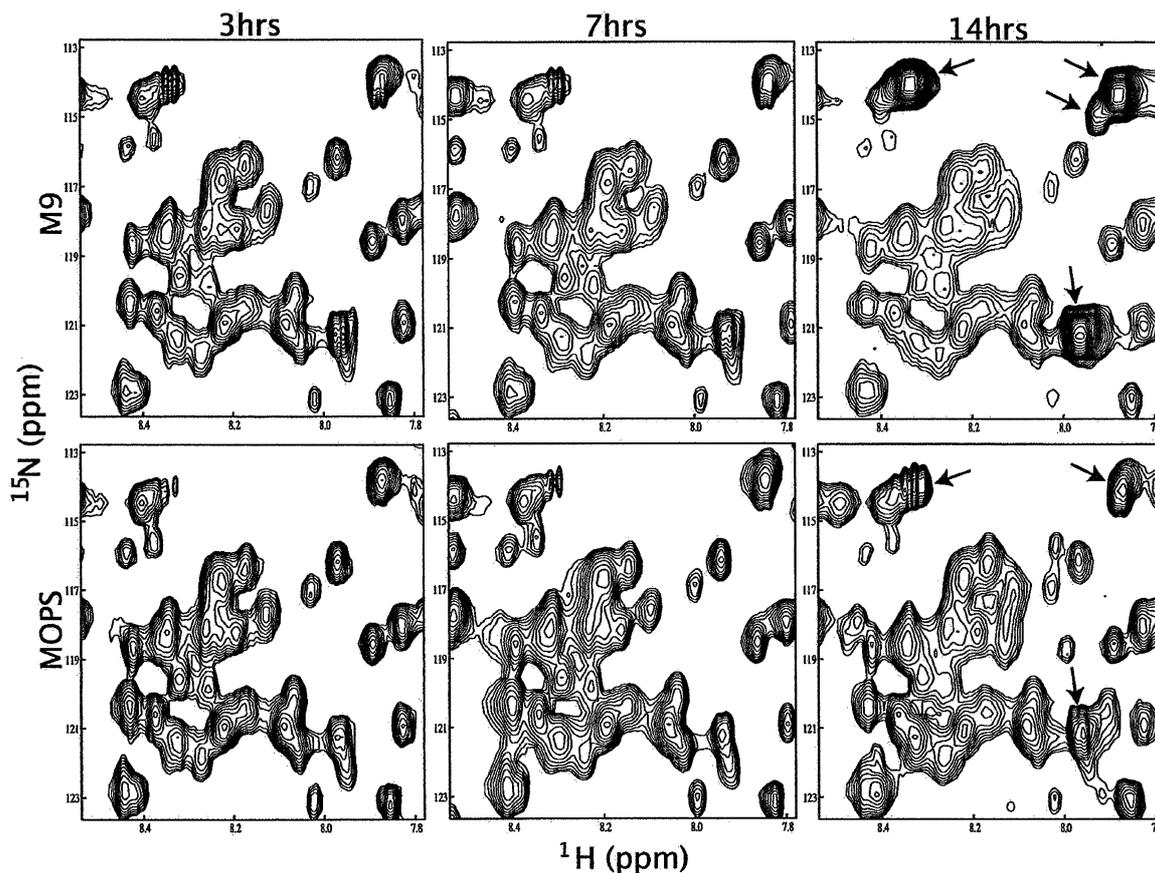
Figure 2. 800 MHz ^{15}N-^{1}H HSQC spectra of 0.5 mM Δ17Arf1 grown at 37 °C in either M9 (top) or MOPS minimal media (bottom) taken at 3, 7 or 14 hr after induction with 0.5 mM IPTG. Arrows in the 14 hr column indicate peaks originating from the ECA$_{CYC}$ amide signals.

preparations. However, GlcNAc is the potential reducing terminal amino sugar of ECA trisaccharides, this observation suggests the possibility that these polysaccharides resulted from degradation of ECA$_{CYC}$ that was originally present in the fractions containing CheY.

The NMR spectroscopic tools provided here to recognize ECA$_{CYC}$ in proteins prepared using standard *E. coli* expression systems, as well as methods to avoid or remove this contaminant, will be useful for a number of applications. Hopefully this will prevent time-consuming NMR analyses by early recognition of these ECA-based signals. This is particularly important for applications such as in the structural genomics field, where the co-purification or apparent binding of a ligand is often interpreted to provide clues to protein function (Parsons et al., 2003).

References

Bruix, M., Jimenez-Barbero, J. and Cronet, P. (1995) *Carbohydr. Res.*, **273**, 157–170.
Erbel, P.J., Barr, K., Gao, N., Gerwig, G.J., Rick, P.D. and Gardner, K.H. (2003) *J. Bacteriol.*, **185**, 1995–2004.

Kahn, R.A., Randazzo, P., Serafini, T., Weiss, O., Rulka, C., Clark, J., Amherdt, M., Roller, P., Orci, L. and Rothman, J.E. (1992) *J. Biol. Chem.*, **267**, 13039–46.

Lugowski, C., Romanowska, E., Kenne, L. and Lindberg, B. (1983) *Carbohydr. Res.*, **118**, 173–181.

MacIntosh, S.E. (2001) M.Sc. Thesis, University of British Columbia, Vancouver.

Mackereth, C.D., Scharpf, M., Gentile, L.N. and McIntosh, L.P. (2002) *J. Biomol. NMR*, **24**, 71–2.

Neidhardt, F.C., Bloch, P.L. and Smith, D.F. (1974) *J. Bacteriol.*, **119**, 736–747.

Parsons, L., Bonander, N., Eisenstein, E., Gilson, M., Kairys, V. and Orban, J. (2003) *Biochemistry*, **42**, 80–89.

Rick, P.D. and Silver, R.P. (1996) In Escherichia coli *and Salmonella: Cellular and Molecular Biology*, Vol. 1, Neidhardt, F.C. (Ed.), American Society for Microbiology, Washington, D.C., pp. 104–122.

Staaf, M., Hoog, C., Stevensson, B., Maliniak, A. and Widmalm, G. (2001) *Biochemistry*, **40**, 3623–8.