1H, 13C, and 15N Backbone assignments of the human brain and acute leukemia cytoplasmic (BAALC) protein

Andras Lang1 · Amit Kumar1 · Jan Jirschitzka2 · Frank Bordusa3 · Oliver Ohlenschläger1 · Christoph Wiedemann3

Received: 18 February 2020 / Accepted: 26 March 2020 / Published online: 2 April 2020 © The Author(s) 2020

Abstract

The brain and acute leukemia cytoplasmic (BAALC; UniProt entry Q8WXS3) is a 180-residue-long human protein having six known isoforms. BAALC is expressed in either hematopoietic or neuroectodermal cells and its specific function is still to be revealed. However, as a presumably membrane-anchored protein at the cytoplasmic side it is speculated that BAALC exerts its function at the postsynaptic densities of certain neurons and might play a role in developing cytogenetically normal acute myeloid leukemia (CN-AML) when it is highly overexpressed by myeloid or lymphoid progenitor cells. In order to better understand the physiological role of BAALC and to provide the basis for a further molecular characterization of BAALC, we report here the 1H, 13C, and 15N resonance assignments for the backbone nuclei of its longest hematopoietic isoform (isoform 1). In addition, we present a 1HN and 15NH chemical shift comparison of BAALC with its shortest, neuroectodermal isoform (isoform 6) which shows only minor changes in the 1H and 15N chemical shifts.

Keywords

Resonance assignments · Heteronuclear NMR · Brain and acute leukemia cytoplasmic protein (BAALC) · Intrinsically disordered protein (IDP) · Neuroectodermal and hematopoietic cell function

Biological context

The brain and acute leukemia cytoplasmic (BAALC; UniProt entry Q8WXS3) is a human protein of 180 amino acids. Eight alternatively spliced transcripts of BAALC were detected and five of them are described to form stable isoforms (isoform 1–3, 5 and 6) (Tanner et al. 2001). The remaining three splice variants encode the same predicted 80-amino-acid protein (isoform 4). These isoforms are expressed in cells of hematopoietic or neuroectodermal origin. The respective gene is located on chromosome 8q22.3 and is highly conserved in mammals and rodents (Fig. 1) but absent in lower organisms (e.g. Caenorhabditis, Drosophila). Currently, the function of BAALC is not fully characterized, but studies indicate a high clinical significance in pathological processes from several leukemias [Acute Lymphoblastic Leukemia (Kuhnl et al. 2010) and Acute Myeloid Leukemia (Baldus et al. 2003b; Bienz et al. 2005; Marcucci et al. 2005) to trisomy 8/Warkany syndrome 2 (Hemsing et al. 2019)].

In bone marrow, the BAALC gene was shown to be expressed mainly in hematopoietic progenitor cells and to be down-regulated during their differentiation (Baldus et al. 2003a). Its over-expression is strongly correlated in cytogenetically normal acute myeloid leukemia (CN-AML) (Weber et al. 2014; Zhou et al. 2015), that is more prevalent with progressing age, and associates with poor outcome questioning the correlation as a pure coincidence. Leukemia, in general, is a very heterogeneous hematological disease due to clonal proliferation of various undifferentiated progenitor cells. Therefore, understanding the signalling leading to myeloid (and/or lymphoid) progenitor cell proliferation and differentiation is indispensable to obtain a deeper understanding of leukemogenesis. Interestingly, BAALC over-expression also positively correlates with the MNI expression level (Heuser et al. 2012) but there is likely a common upstream regulatory mechanism. It was shown that BAALC does not enhance self-renewal of hematopoietic
progenitor cells, but inhibits differentiation by desensitizing these
cells to all-trans retinoic acid-induced proliferation arrest and dif-
erentiation, although, less effectively than MN1 does. Interest-
ingly, CEBPA, one of the retinoic acid receptor target genes, is
also a target of RUNX1 (alternatively AML1) (Friedman 2015).
RUNX1 is a transcription factor important for hematopoietic cell
development during embryogenesis (Tober et al. 2016) and as a
hybrid protein formed by fusions of AML1 and ETO, a genetic
aberration leading to the acute myeloid leukemia subtype M2
(Lin et al. 2017). In addition, RUNX1 can markedly increase
the BAALC expression level if a certain SNP is located in the
BAALC regulatory region (Eisfeld et al. 2014). The guanine-
thymine exchange in this allele creates a binding site for the acti-
vating RUNX1 and predisposes the carrier to enhanced myeloid
leukemogenesis. Thus, a high BAALC level is a risk factor for
leukemogenesis. Whether the high BAALC expression is reason
or consequence of this CN-AML subtype, can be clarified only
by further investigations of the underlying molecular mechanism.

In rat, BAALC is membrane-anchored at its N-terminus
via myristoylation at Gly2 and palmitoylation at Cys3
(Wang et al. 2005) and, due to its identical sequence, pre-
sumably also in human. Whereas the BAALC gene is studied
regarding myeloid leukemia, its gene product, the BAALC
protein, was neither characterized by biophysical nor bio-
chemical methodology. This study presents 1H, 15N and 13C
backbone resonance assignments to provide the basis for an
atom-based structural view on the BAALC protein and its
interactions employing high-resolution NMR spectroscopy.

Methods and experiments

Protein expression and purification

The full human E. coli codon optimized BAALC (isoform 1) gene
(Tanner et al. 2001) was cloned into a pET28a plasmid using NdeI
and XhoI restriction enzymes. The plasmid was modified using
gene-tailor mutagenesis PCR using Platinum Taq DNA Polymer-
ase (Invitrogen) to replace the thrombin by a TEV enzyme cut
site between the His6-tag and the target gene. Therefore, the final
protein contains an extra Gly residue at the N-terminus leading
to the 181-residue-long protein. The 25 μl final reaction mixture
contained 1 × High Fidelity buffer, 1 mM MgSO4, 0.2 mM dNTP,
0.6–0.6 mM forward and backward primers ~ 6.5 ng template
(BAALC in pET28a between NdeI/XhoI) and 0.5 units of DNA
polymerase. The following primers are used:

\[ \text{fwd } 5' \text{ AGC AGC GGC CTG GTG CCG CGC GAA AAC CTA TAT TTT CAG GCC ATG 3'} \]

\[ \text{rev } 3' \text{ GTA GTG GTA GTG TCG TCG CCG GAC CAC GGC GCG 5'} \]

After the initial 2 min-long denaturation, the reaction had
20 repetitions such as denaturation at 94 °C for 30 s, anneal-
ing at 65 °C for 30 s and extension at 68 °C for 6 min before
the final round of extension at 68 °C for 10 min. The 10 μl
purified DNA (PCRapate kit) was treated with 20 units DpnI
in 1 × CutSmart buffer at 37 °C for 1 h to remove methylated
DNA. The purified DNA (PCRapate kit) was transformed (~125 ng) into 50 µl DH5α E. coli competent cells using heat shock (10 min on ice and 45 s at 42 °C) and left on a LB kanamycin plate O/N at 37 °C. 147 ng/µl plasmid was purified from colonies and its insert was confirmed by DNA sequencing (Eurofins Genomics).

40–50 ng plasmid was used for heat shock transformation into 25 µl BL21(DE3) cells. Colonies were grown in 500 ml LB medium at 37 °C until OD600 reached 1.5. Cells were pelleted at 4800 rpm for 15 min using a Sorvall H6000A swinging bucket rotor (i.e. ~6700×g). The pellet was resuspended in 1 l sterile M9 medium supplemented with 1 g 15NH4Cl and 2 g 13C-labeled glucose. After one hour incubation at 18 °C, cells were induced by 0.3 mM IPTG O/N at 18 °C. Cells were lysed in 12 ml ice-cold lysis buffer (5 mM imidazole, 50 mM Tris and 300 mM NaCl, pH ~7.5, supplemented with proteinase inhibitor cocktail, DNAse I and 500 times diluted β-mercaptoethanol) three times using French Press. Cell debris was pelleted at 7600 rpm (Beckman Coulter C0650 rotor, i.e. ~5950×g) for 1 h. Supernatant was purified using Ni–NTA affinity chromatography. His-tagged BAALC was eluted using buffer containing 250 mM imidazole, 50 mM Tris, 300 mM NaCl and 1:500 β-mercaptoethanol, pH ~7.5 and its concentration measured by NanoDrop (Thermo Scientific). His-tag removal was conducted by using at least 40 × weight excess of 3 mg/ml TEV (200 µl) for about an hour at room temperature (~21 °C) and subsequently the sample was two times dialyzed using a 10 kD cut-off membrane against 1 l, 20 mM Tris, 5 mM NaCl, 2 mM DTT, pH ~7.5 at ~4 °C for about 2 h each. Due to low ionic strength precipitation of TEV occurred that was pelleted at 7600 rpm (Beckman Coulter C0650 rotor, i.e. ~5950×g) for 30 min. The clear supernatant containing the 181-residue-long BAALC (pI = 5.48) was further purified on anion-exchange chromatography using DEAE Sepharose resin (GE Healthcare) with gradual increase of NaCl concentration (20 mM, 30 mM, 40 mM and 50 mM) before the final 500 mM NaCl elution. Eluted fractions (at 50 mM NaCl) were mixed and dialysed O/N at ~4 °C against 20 mM Tris, 100 mM NaCl and 2 mM DTT, pH ~7.3. Using a 3 kD cut-off membrane, all dialysed protein was concentrated (7600 rpm Beckman Coulter C0650 rotor) until less than 1 ml.

The protein concentration was ~1.55 mg/ml measured by NanoDrop that was further purified with size-exclusion chromatography on S75 10/300 GL Superdex column (GE Healthcare) using an ÄKTA Avant system. Three 0.5 ml fractions containing significantly purified BAALC were concentrated again (7600 rpm Beckman Coulter C0650 rotor) and a 3 kD cut-off cassette dialysed O/N at ~4 °C against 20 mM sodium phosphate, pH 6.5.

The shorter BAALC isofrom 6 was cloned from the isofrom 1 construct by inserting a stop codon after position 54 and replacing Val54 by a glycine. Expression and purification was done as described. Deviating from the procedure described above, after His-tag removal by using TEV an additional Ni–NTA affinity chromatography was applied and the flow through was collected and concentrated.

**NMR spectroscopy**

The NMR experiments for the 1H, 15N and 13C chemical shift assignments were acquired at 283 K in 20 mM sodium phosphate, pH 6.5 (90% H2O/10% D2O) on Bruker 600 (14.1 T) equipped with cryo-probe and 700 MHz (16.4 T) Avance III spectrometers. For sequential walk and backbone chemical shifts assignment, HNCO, HNCACB, HN(CA)CO, HN(CO)CA (Bax and Ikura 1991) and HNCA (Marion et al. 1989) spectra. The data was processed using TOPSPIN v.4.0.6 and analysed with CARA (Keller 2004) as well as CCPNmr Analysis (Vranken et al. 2005).

**Structure prediction**

The secondary structure elements of BAALC were examined by analysis of the chemical shift data with the program CSI v.3.0 (Hafsa et al. 2015) and the secondary structure propensity approach (Marsh et al. 2006). For the sequence-based prediction the IUPred2A server was used (Dosztányi 2017; Mézsáros et al. 2018).

**Extent of assignments and data deposition**

In contrast to the wild type, the BAALC protein used here for the NMR experiments exhibits one additional N-terminal amino acid (Gly0) arising from cloning purposes. This Gly0 is not considered in the following statistics.

The [1H,15N]-HSQC spectrum of BAALC (Fig. 2) allowed assignment of 99% of the backbone 1H and 15N resonances of the non-proline residues. Only the resonances of Gln167 of BAALC could not be determined. 100% of the 13Cα (180/180) and 99% of the 13Cγ (163/164) backbone chemical shifts were assigned and the procedure described above, after His-tag removal by using TEV an additional Ni–NTA affinity chromatography was applied and the flow through was collected and concentrated.

The NMR experiments for the 1H, 15N and 13C chemical shift assignments were acquired at 283 K in 20 mM sodium phosphate, pH 6.5 (90% H2O/10% D2O) on Bruker 600 (14.1 T) equipped with cryo-probe and 700 MHz (16.4 T) Avance III spectrometers. For sequential walk and backbone chemical shifts assignment, HNCO, HNCACB, HN(CA)CO, HN(CO)CA (Bax and Ikura 1991) and HNCA (Marion et al. 1989) spectra. The data was processed using TOPSPIN v.4.0.6 and analysed with CARA (Keller 2004) as well as CCPNmr Analysis (Vranken et al. 2005).

The secondary structure elements of BAALC were examined by analysis of the chemical shift data with the program CSI v.3.0 (Hafsa et al. 2015) and the secondary structure propensity approach (Marsh et al. 2006). For the sequence-based prediction the IUPred2A server was used (Dosztányi 2017; Mézsáros et al. 2018).

The [1H,15N]-HSQC spectrum of BAALC (Fig. 2) allowed assignment of 99% of the backbone 1H and 15N resonances of the non-proline residues. Only the resonances of Gln167 of BAALC could not be determined. 100% of the 13Cα (180/180) and 99% of the 13Cγ (except Pro36) backbone chemical shifts were assigned employing HNCO, HN(CA)CO, HNCA, HN(CO)CA and HNCACB spectra. 99% of assignments (164/166) for the β-carbon resonances were achieved (except: Met1, Arg21). In addition, we report 14 13Cγ resonances out of the 16 proline residues which allow to predict all the respective prolines possessing a trans conformation (~3 ppm shifts) (Schubert et al. 2002). Furthermore,
98% (191/194) of the Hα resonances were assigned. Only the Hα nuclei of Arg14, Gln167, Met168 could not be assigned.

Analysis of structural elements by the CSI web server (data not shown) resulted in an all-coil prediction. This supports the expectation based upon the appearance of the [1H, 15N]-HSQC spectrum (Fig. 2) which showed a reduced spectral dispersion of average chemical shifts implying flexibility typical for intrinsically disordered proteins. An IUPred2A analysis (Dosztányi 2017; Mészáros et al. 2018) also predicts that BAALC is predominantly disordered with very weak, short ordering tendency at residues 6–11, 18–25 and 76–85 (Fig. 3A). In order to reveal potential secondary structure propensity (SSP), which might not be detected by the other approaches, we analyzed the chemical shift data using the SSP method (Fig. 3B) (Marsh et al. 2006). By averaging the potential α-helical and β-sheet regions of the calculated SSP scores, an overall total of 6.3% α-structure and 3.3% β-structure, respectively, is estimated for BAALC. The large degree of disorder/flexibility is consistent with the findings of the other structure prediction tools and confirms the observation made from the 1H, 15N chemical shift dispersion.

A 1H, 15N chemical shift comparison of BAALC (isoform 1) with its shortest, neuroectodermal isoform 6 was performed (Fig. 4). The result indicates that only minor changes (less than 0.1 ppm) in the 1H and 15N chemical shifts occur. The only exception is residue 53, which can be explained by its penultimate position in isoform 6.

As described above, it is likely that also the human BAALC protein is anchored in the membrane. The same
applies to post-translational modifications (e.g. phosphorylation at some Ser, Thr and Tyr residues). It remains to be investigated whether the spatial proximity of the BAALC protein and its shorter isoforms to a membrane and/or additional modifications causes structural changes.

The $^1$H, $^{13}$C and $^{15}$N backbone chemical shifts of BAALC have been deposited in the BioMagResBank (BMRB) under the accession number 28084.

Reference

Bax A, Ikura M (1991) An efficient 3D NMR technique for correlating the proton and $^{15}$N backbone amide resonances with the alpha-atom of the preceding residue in uniformly $^{15}$N/$^{13}$C enriched proteins. J Biomol NMR 1:99–104. https://doi.org/10.1007/BF01874573

Baldus CD, Tanner SM, Kusewitt DF, Liyanarachchi S, Choi C, Caligiuri MA, Bloomfield CD, de la Chapelle A (2003a) BAALC, a novel marker of human hematopoietic progenitor cells. Exp Hematol 31:1051–1056. https://doi.org/10.1016/j.exphem.2003.08.004
Baldus CD, Tanner SM, Ruppert AS, Whitman SP, Archer KJ, Marucchi G, Caligiuri MA, Carroll AJ, Vardiman JW, Powell BL, Allen SL, Moore JO, Larson RA, Kolitz JE, de la Chapelle A, Bloomfield CD (2003b) BAALC expression predicts clinical outcome of de novo acute myeloid leukemia patients with normal cytogenetics: a Cancer and Leukemia Group B Study. Blood 102:1613–1618. https://doi.org/10.1182/blood-2003-02-0359

Bienz M, Ludwig M, Mueller BU, Oppeliger Leibundgut E, Ratschiller D, Selenhailer M, Fey MF, Pabst T (2005) Risk assessment in patients with acute myeloid leukemia and a normal karyotype. Clin Cancer Res 11:1416–1424. https://doi.org/10.1158/1078-0432.CCR-04-1552

Bracken C, Palmer AG 3rd, Cavanagh J (1997) (HN(COCA))NH and (HN(COCA))NH experiments for 1H–15N backbone assignments in 13C/15N-labeled proteins. J Biomol NMR 9:94–100. https://doi.org/10.1023/A:1018679819693

Clancy DM, Henry CM, Davidovich PB, Sullivan GP, Belotkerovskaya E, Martin SJ (2016) Production of biologically active IL-36 family cytokines through insertion of N-terminal caspase cleavage motifs. FEBS Open Bio 6:338–348. https://doi.org/10.1002/2211-5463.12044

Clubb RT, Thanabal V, Wagner G (1992) A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue 1H(N), 15N, and 13C chemical-shifts in 15N,13C-labeled proteins. J Magn Reson 97:213–217. https://doi.org/10.1016/0022-2364(92)90252-3

Dosztányi Z (2017) Prediction of protein disorder based on IUPred. Protein Sci 26:331–340. https://doi.org/10.1002/pro.3334

Eifeld AK, Schwind S, Patel R, Huang X, Santhanam R, Walker CJ, Markowitz J, Hoag KW, Jarvinen TM, Leffel B, Perrotti D, Carson III WE, Marcucci G, Bloomfield CD, de la Chapelle A (2014) Intronic miR-3151 within BAALC drives leukemogenesis by deregulating the TP53 pathway. Sci Signal 7:ra36. https://doi.org/10.1126/scisignal.2004762

Friedman AD (2015) C/EBP alpha in normal and malignant myelopoeisis. Int J Hematol 101:330–341. https://doi.org/10.1007/s12187-015-1764-6

Hafsa NE, Arndt D, Wishart DS (2015) CSI 3.0: a web server for identifying secondary and super-secondary structure in proteins using NMR chemical shifts. Nucleic Acids Res 43:W370–377. https://doi.org/10.1093/nar/gkv494

Hafsa NE, Arndt D, Wishart DS (2015) CSI 3.0: a web server for identifying secondary and super-secondary structure in proteins using NMR chemical shifts. Nucleic Acids Res 43:W370–377. https://doi.org/10.1093/nar/gkv494

Heuser M, Berg T, Kuchenbauer F, Lai CK, Park G, Fung S, Lin G, Hemsing AL, Hovland R, Tsykunova G, Reikvam H (2019) Trisomy 13C/15N-labeled proteins. J Biomol NMR 9:94–100. https://doi.org/10.1023/A:1018679819693

Marion D, Driscoll PC, Kay LE, Wingfield PT, Bax A, Gronenborn AM, Clore GM (1989) Overcoming the overlap problem in the assignment of 1H NMR spectra of larger proteins by use of three-dimensional heteronuclear 1H–15N Hartmann-Hahn-multiple quantum coherence and nuclear overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 beta. Biochemistry 28:6150–6156. https://doi.org/10.1021/bi00441a004

Mészáros B, Erdős G, Dosztányi Z (2018) IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. Nucleic Acids Res 46;W329–W337. https://doi.org/10.1093/nar/gky384

Montelione GT, Lyons BA, Emerson SD, Tashiro M (1992) An efficient triple resonance experiment using carbon-13 isotopic mixing for determining sequence-specific resonance assignments of isotopically-enriched proteins. J Am Chem Soc 114:10974–10975. https://doi.org/10.1021/ja00053a051

Mulder FA, Schipper D, Bott R, Boelens R (1999) Altered flexibility in the substrate-binding site of related native and engineered high-bacillic acid subtilisins. J Biol Chem 292:111–123. https://doi.org/10.1006/jbch.1999.3034

Schubert M, Labudde D, Oshkinat H, Schmiesser P (2002) A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on 13C chemical shift statistics. J Biomol NMR 24:149–154. https://doi.org/10.1023/A:1020997118364

Tanner SM, Austin JL, Leone G, Rush LJ, Plass C, Heinonen K, Mrozek K, Sill H, Knuttila S, Kolitz JE, Archer KJ, Caligiuri MA, Bloomfield CD, de la Chapelle A (2001) BAALC, the human member of a novel mammalian neuroectoderm gene lineage, is implicated in hematopoiesis and acute leukemia. Proc Natl Acad Sci USA 98:13901–13906. https://doi.org/10.1073/pnas.241525498

Tober J, Majenbourg MW, Speck NA (2016) Taking the leap: Runx1 in the formation of blood from endothelium. Curr Top Dev Biol 118:113–162. https://doi.org/10.1016/bs.ctdb.2016.01.008

Vanden W, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J, Laue ED (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins 59:687–696. https://doi.org/10.1002/prot.20449

Wang X, Tian QB, Okano A, Sakagami H, Moon IS, Kondo H, Endo S, Suzuki T (2005) BAALC 1–6–8 protein is targeted to post-synaptic lipid rafts by its N-terminal myristoylation and palmitylation, and interacts with alpha, but not beta, subunit of Ca/calmodulin-dependent protein kinase II. J Neurochem 92:647–659. https://doi.org/10.1111/j.1471-4159.2004.02902.x

Weber S, Alpermann T, Dicker F, Nadarajah N, Eder C, Fasan A, Kohlmann A, Meggendorfer M, Haferlach C, Kern W, Haferlach T, Schnittger S (2014) BAALC expression: a suitable marker for prognostic risk stratification and detection of residual disease in cytogenetically normal acute myeloid leukemia. Blood Cancer J 4:e173. https://doi.org/10.1038/bcj.2013.71

Wittekind M, Mueller L (1993) HNCACB: a high-sensitivity 3D NMR experiment to correlate amide proton and nitrogen resonances with chemical-shifts in 15N-13C-labeled proteins. J Magn Reson 89:496–514. https://doi.org/10.1016/0022-2364(90)90333-5

Keller RJ (2004) Computer aided resonance assignment tutorial. Cantina, Goldau

Kuhn H, Gokbuget N, Stroux A, Burmeister T, Neumann M, Heesch S, Haferlach T, Hoezler D, Hofmann WK, Thiel E, Baldus CD (2010) High BAALC expression predicts chemoresistance in adult B-precursor acute lymphoblastic leukemia. Blood 115:3737–3744. https://doi.org/10.1182/blood-2009-09-241943

Lin S, Mulloy JC, Goyama S (2017) RUNX1-ETO leukemia. Adv Exp Med Biol 962:151–173. https://doi.org/10.1007/978-1-03000-323-2_11

Marucchi G, Mrozek K, Bloomfield CD (2005) Molecular heterogeneity and prognostic biomarkers in adults with acute myeloid leukemia and normal cytogenetics. Curr Opin Hematol 12:68–75. https://doi.org/10.1097/01.moh.0000146082.29685.d1

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.