Published in final edited form as:
Proteins. 2012 February ; 80(2): 671–675. doi:10.1002/prot.23236.

Solution NMR structure of Asl3597 from Nostoc sp. PCC7120, the first structure from protein domain family PF12095, reveals a novel fold

Erik A. Feldmann[1], **Theresa A. Ramelot**[1], **Yunhuang Yang**[1], **Rong Xiao**[2], **Thomas B. Acton**[2], **John K. Everett**[2], **Gaetano T. Montelione**[2,3], and **Michael A. Kennedy**[1,*]

[1] Department of Chemistry and Biochemistry, Miami University, Oxford, Ohio 45056 and Northeast Structural Genomics Consortium

[2] Center for Advanced Biotechnology and Medicine, Department of Molecular Biology and Biochemistry, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854 and Northeast Structural Genomics Consortium

[3] Department of Biochemistry, Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, Piscataway, New Jersey 08854

Abstract

The protein domain family PF12095 (DUF3571) is a functionally uncharacterized family of small proteins conserved from cyanobacteria to plants that are typically 85 to 95 amino acids in length in cyanobacteria. In this report, we describe the solution NMR structure of the 86-residue protein Asl3597 from *Nostoc* sp. PCC7120. The structure of Asl3597, which constitutes the first three-dimensional structure from protein family PF12095, has a unique α/β sandwich fold consisting of four anti-parallel β-strands opposite three continuous α-helices. Asl3597 may have a role in the assembly of the hydrophilic subcomplex of the cyanobacterial NAD(P)H complex as suggested by data for the orthologous Chlororespiratory reduction 7 protein from *Arabidopsis thaliana*.

Keywords

Asl3597; PF12095; DUF3571; structural genomics; 2KRX; NDH complex; chlororespiratory reduction 7; CRR7

INTRODUCTION

Here we present the solution NMR structure of the small (86-residue), acidic (pI 4.37) Asl3597 protein from *Nostoc* sp. PCC7120 (UniProtKB ID, Q8YR53_ANASP; NESG ID, NsR244; KEGG ID, ana:asl3597), a member of the (Pfam) PF12095 (DUF3571) protein domain family. Asl3597 was selected as a target by the Northeast Structural Genomics Consortium (NESG) as part of the National Institutes of Health Protein Structure

*Correspondence to: Michael A. Kennedy, Department of Chemistry and Biochemistry, Miami University, Oxford, Ohio 45056. kennedm4@muohio.edu. Phone: 1-513-529-8267 Fax: (513) 529-5715. Erik A. Feldmann feldmaea@muohio.edu Theresa A. Ramelot ramelota@muohio.edu Yunhuang Yang yangy9@muohio.edu Rong Xiao rxiao@cabm.rutgers.edu Thomas B. Acton acton@cabm.rutgers.edu John K. Everett everett.jk@gmail.com Gaetano T. Montelione guy@cabm.rutgers.edu.

Initiative-2, a massive international campaign to determine three-dimensional structures of broadly conserved protein domain families by NMR spectroscopy and X-ray crystallography. The PF12095 Pfam consists of proteins that are typically 85 to 95 amino acid residues in length found in heterocystous and non-heterocystous cyanobacteria, and upwards of 160 residues in plants due to the existence of a putative N-terminal 68-residue targeting sequence. The function of these proteins is still mostly unexplored with the exception of the 156-residue chlororespiratory reduction 7 (CRR7) protein from *Arabidopsis thaliana* (UniProtKB ID, Q9FL87_ARATH; KEGG ID, ath:AT5G39210, 33% identity). CRR7 is a soluble, chloroplast stromal protein required for assembly of the hydrophilic subcomplex of the chloroplast NAD(P)H dehydrogenase (NDH) complex together with CRR6.[1,2] The NDH complex in higher plants is located in the thylakoids where it is involved in chlororespiration and cyclic electron transport (CET) by interacting with photosystem-I (PSI) to form the NDH-PSI supercomplex as shown in *Arabidopsis*.[3,4] The NDH complex in cyanobacteria (NDH-1) is also responsible for CO_2 cellular uptake, in addition to respiration and PSI-mediated CET, and is believed to be the evolutionary ancestor of the chloroplast NDH.[5,6] The NDH-1 complex is composed of hydrophilic and membrane subcomplexes that together form an L-shape that is visible by electron microscopy.[7] However, little is still known about the molecular mechanism of CRR7 and Asl3597 and the function they play in assembly or stabilization of the NDH complex in plants and cyanobacteria. The solution NMR structure of Asl3597 reported here will hopefully assist in the future biochemical understanding of this family of proteins from Pfam PF12095.

MATERIALS AND METHODS

Samples of isotopically labeled [*U*-^{15}N,^{13}C] and *U*-^{15}N, 5% biosynthetically- directed ^{13}C (NC5) Asl3597 were generated for NMR through cloning, expression, and purification techniques using standard protocols of the NESG consortium.[8] Asl3597 behaved as a monomer under the conditions used for NMR data collection (0.8 mM Asl3597, 200 mM NaCl, 10 mM dithiothreitol, 5 mM $CaCl_2$, 0.02% NaN_3, and 20 mM MES pH 6.5 at 20°C) according to analytical gel filtration chromatography, in-line with static light scattering (Supplementary Fig. S1). An isotropic overall rotational correlation time (τ_c) of 12.1 ns was estimated from ^{15}N T_1 and $T_{1\rho}$ relaxation data (Supplementary Fig. S2). The longer than expected τ_c and shorter $T_{1\rho}$ measurements suggested that chemical or conformational exchange caused an increase the averaged transverse relaxation rate. The amide backbone $^1H^N$, $^1H^\alpha$, ^{15}N, $^{13}C^\alpha$ and side chain $^{13}C^\beta$ and $^1H^\beta$ resonances were assigned after manual peak picking of the ^{15}N-HSQC, HNCA, HNCO, HN(CO)CA, HNCACB, CBCA(CO)NH, and HBHA(CO)NH experiments, and data submission to the PINE server (NMRFAM)[9] for auto-assignment. Side chain resonance assignments were completed manually using the above experiments together with the H(C)CHTOCSY, H(C)CH-COSY, H(CC)(CO)NH-TOCSY, (H)CC(CO)NH-TOCSY, (H)CCH-TOCSY, two ^{13}C-edited NOESY-HSQC (τ_m = 70 ms) optimized for either aliphatic or aromatic carbons, and 4D ^{13}C-^{13}C-HMQC-NOESY-HMQC (τ_m = 70 ms) experiments collected on a [*U*-^{15}N,^{13}C] sample. Stereospecific assignments of isopropyl side chain methyl groups of Leu and Val residues were performed using a constant time ^{13}C-HSQC experiment collected on the NC5

sample.[10] Dihedral angle constraints were computed by Talos+[11] based on the chemical shift resonance assignments. CYANA version 2.1[12,13] was used to automatically assign NOEs. The solution NMR structure of Asl3597 was calculated using XPLOR-NIH-2.20,[14] followed by refinement in explicit water using CNS version 1.1.[15] Analysis of Asl3597 structures was performed using global structure quality scores and structural statistics computed from the PSVS software package[16] (Table 1). The final ensemble of 20 models sorted by the lowest energy, were deposited to the Protein Data Bank (PDB ID, 2KRX) and BioMagResDB (BMRB accession number, 16652).

RESULTS AND DISCUSSION

The solution NMR structure of Asl3597 is the first structure from Pfam domain PF12095, which remains a mostly uncharacterized family of proteins of unknown function. Asl3597 has a unique α/β sandwich fold consisting of four anti-parallel β-strands (β1, N11-E16; β2, Q19-T25; β3, L69-G72; β4, K75-A81) facing opposite a continuous three α-helix bundle (α1, T26-K39; α2, L47-K51; α3, L55-T65), and flanked by N- and C-terminal disordered tails with a secondary structure order of N-β1-β2-α1-α2-α3-β3-β4-C (Fig. 1A,B). As reported in the Materials and Methods section, the estimate of the overall rotational correlation time from the relaxation measurements was longer than expected for a monomer of Asl3597 in solution. Inspection of the ^{1}H-^{15}N HSQC spectrum of Asl3597 revealed several crosspeaks that were significantly weaker compared to the majority of crosspeaks. The linewidths of these crosspeaks were measured from an unapodized non-constant ^{1}H-^{15}N HSQC spectrum and used to generate a sausage representation of Asl3597 (Supplementary Fig S4). The amino acids with the broadest linewidths were clustered in and around the β-sheet face suggesting that these residues were undergoing chemical and/or conformational exchange. Therefore the longer than expected average correlation times (τ_c) and shorter than expected $T_{1\rho}$ relaxation times were likely due to an exchange contribution to the relaxation rates of these affected residues, and the contribution of these resonances to the computation of the overall average τ_c. Asl3597 has low sequence similarity with any other protein in the PDB (< 20%), and structural alignment using the Dali[17] server revealed that the closest structural similarity, which was still low, with small domains of RNA polymerase complex I and II from a variety of different source organisms (top result: RPO1N subunit of the archaeal 13-subunit DNA-directed RNA polymerase; PDB ID, 2WB1; Dali *Z*-score, 3.6; C^{α} RMSD, 4.6 Å) of which there also exists very low sequence identity (< 10%). Therefore it appears that Asl3597 exhibits an entirely novel protein fold.

Sequence analysis using a BLAST search provided by KEGG[18] yielded only proteins identified as either “hypothetical” or “uncharacterized” (*E*-values: 10^{1} - 10^{-35}) with the exception of the 156-residue CRR7 protein (*E*-value: 0.007, 33% identity) from *Arabidopsis thaliana*. Recent data indicated that CRR7, in conjunction with CRR6, is required for post-translational assembly of the hydrophilic subcomplex of the chloroplast NDH complex in *Arabidopsis*. CRR7 was found localized to the chloroplast stromal fraction, potentially facilitated by the N-terminal 68 residues that contain a predicted plastid targeting sequence.[2] CRR7 is believed to play a role in biogenesis of the hydrophilic subcomplex based on the study of *crr7* null mutants that were discovered to have impaired NDH electron transport activity.[1] The authors suggested that although CRR7 may be involved in the assembly of the

hydrophilic subcomplex, it most likely is not a subunit of NDH.[2] As expected from our BLAST search, the top results for Asl3597 were orthologous hypothetical proteins from other heterocyst forming and non-heterocystous cyanobacterial species all belonging to Pfam12095. Conservation of amino acid residues was assessed using BOXSHADE version 3.21 and the ConSurf server,[19] following generation of a multiple sequence alignment file using ClustalW.[20] Sequence alignment across cyanobacterial proteins and CRR7 from *Arabidopsis* showed that stretches of homology exist across the entirety of the protein sequence including charged, aromatic, and nonpolar amino acids (Fig. 2). ConSurf analysis revealed two conserved surface patches, one at the disordered C-terminal tail from R83 to K86, and one that extends across β1 (L15), β2 (E21), and β4 (Q78-Y80) (Fig. 1C). Analysis of the electrostatic surface potential[21] revealed a mostly negatively-charged β-sheet face (Fig. 1D). Unfortunately due to the novel fold and uncharacterized nature of Asl3597, it is currently difficult to ascertain any other functional information beyond that derived from the solution NMR structure provided here.

In summary, we have determined the solution NMR structure of Asl3597 from *Nostoc* sp. PCC7120, the first structure of the PF12095 (DUF3571) protein domain family, which remains functionally uncharacterized. Sequence analysis of Asl3597 led to identification of the ortholog CRR7 from *Arabidopsis thaliana* which has recently been suggested to play a role in the biogenesis of the hydrophilic subcomplex of the NDH complex in the chloroplast stroma. Additional biochemical relevance for this unique structural motif of Asl3597, in particular from a photosynthetic bacterium, remains to be determined.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

REFERENCES

1. Kamruzzaman Munshi M, Kobayashi Y, Shikanai T. Identification of a novel protein, CRR7, required for the stabilization of the chloroplast NAD(P)H dehydrogenase complex in Arabidopsis. Plant J. 2005; 44:1036–1044. [PubMed: 16359395]
2. Peng L, Cai W, Shikanai T. Chloroplast stromal proteins, CRR6 and CRR7, are required for assembly of the NAD(P)H dehydrogenase subcomplex A in Arabidopsis. Plant J. 2010; 63:203–211. [PubMed: 20444231]
3. Peng L, Shimizu H, Shikanai T. The chloroplast NAD(P)H dehydrogenase complex interacts with photosystem I in Arabidopsis. J Biol Chem. 2008; 283:34873–34879. [PubMed: 18854313]
4. Peng L, Shikanai T. Supercomplex formation with photosystem I is required for the stabilization of the chloroplast NADH dehydrogenase-like complex in Arabidopsis. Plant Physiol. 2011; 155:1629–1639. [PubMed: 21278308]
5. Ogawa T, Mi H. Cyanobacterial NADPH dehydrogenase complexes. Photosynth Res. 2007; 93:69–77. [PubMed: 17279442]
6. Friedrich T, Weiss H. Modular evolution of the respiratory NADH:ubiquinone oxidoreductase and the origin of its modules. J Theor Biol. 1997; 187:529–540. [PubMed: 9299297]
7. Arteni AA, Zhang P, Battchikova N, Ogawa T, Aro EM, Boekema EJ. Structural characterization of NDH-1 complexes of Thermosynechococcus elongatus by single particle electron microscopy. Biochim Biophys Acta. 2006; 1757:1469–1475. [PubMed: 16844076]
8. Acton T, Xiao R, Anderson S, Aramini J, Buchwald W, Ciccosanti C, Conover K, Everett J, Hamilton K, Huang Y, Janjua H, Kornhaber G, Lau J, Lee D, Liu G, Maglaqui M, Ma L, Mao L, Patel D, Rossi P, Sahdev S, Shastry R, GVT S, Tang Y, Tong S, Wang D, Wang H, Zhao L,

Montelione G. Preparation of protein samples for NMR structure, function, and small-molecule screening studies. Methods Enzymol. 2011; 493:21–60. [PubMed: 21371586]

9. Bahrami A, Assadi AH, Markley JL, Eghbalnia HR. Probabilistic interaction network of evidence algorithm and its application to complete labeling of peak lists from protein NMR spectroscopy. PLoS Comput Biol. 2009; 5:e1000307. [PubMed: 19282963]
10. Neri D, Szyperski T, Otting G, Senn H, Wüthrich K. Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional 13C labeling. Biochemistry. 1989; 28:7510–7516. [PubMed: 2692701]
11. Shen Y, Delaglio F, Cornilescu G, Bax A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J Biomol NMR. 2009; 44:213–223. [PubMed: 19548092]
12. Herrmann T, Güntert P, Wüthrich K. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. J Mol Biol. 2002; 319:209–227. [PubMed: 12051947]
13. Güntert P. Automated NMR structure calculation with CYANA. Methods Mol Biol. 2004; 278:353–378. [PubMed: 15318003]
14. Schwieters C, Kuszewski J, Clore G. Using Xplor-NIH for NMR molecular structure determination. Progress in Nuclear Magnetic Resonance Spectroscopy. 2006; 48:47–62.
15. Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr. 1998; 54:905–921. [PubMed: 9757107]
16. Battacharya A, Tejero R, Montelione G. Evaluating protein structures determined by structural genomics consotria. Proteins. 2007; 66:778–795. [PubMed: 17186527]
17. Holm L, Rosenström P. Dali server: conservation mapping in 3D. Nucleic Acids Res. 2010; 38:W545–549. [PubMed: 20457744]
18. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000; 28:27–30. [PubMed: 10592173]
19. Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res. 2010; 38:W529–533. [PubMed: 20478830]
20. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994; 22:4673–4680. [PubMed: 7984417]
21. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci U S A. 2001; 98:10037–10041. [PubMed: 11517324]
22. Huang YJ, Powers R, Montelione GT. Protein NMR recall, precision, and F-measure scores (RPF scores): structure quality assessment measures based on information retrieval statistics. J Am Chem Soc. 2005; 127:1665–1674. [PubMed: 15701001]

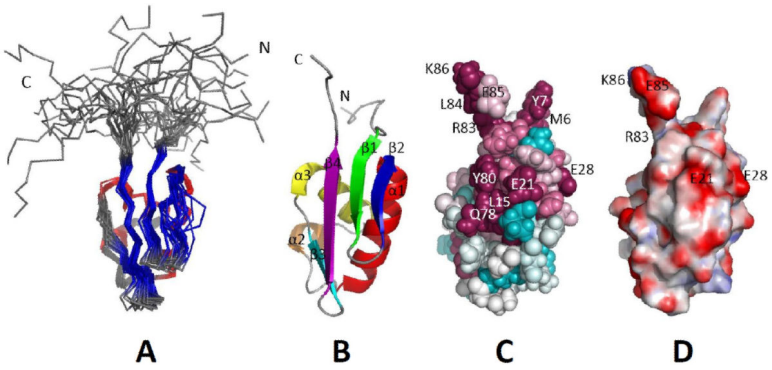

Figure 1.

(A) Superposition of the final ensemble of 20 conformers from the solution NMR structure of Asl3597 (PDB ID, 2KRX). (B) Cartoon representation of Asl3597 showing β-strand order 3412. Secondary structure elements are labeled; α-helices are shown in red, β-strands are shown in blue, and loops are shown in gray. (C) ConSurf19 image showing the conserved residues in Asl3597 from the top 10 BLAST results (KEGG18 database) aligned with ClustalW.20 The color scheme reflects the degree of residue conservation over these members from protein domain family PF12095, and is depicted as completely conserved (magenta), highly conserved (dark and light pink), average (white), and variable (cyan). Selected highly conserved residues are labeled. (D) ABPS21 image of the electrostatic surface potential of Asl3597. Positively charged are shown in blue, negatively charged shown in red and neutral shown in white (±20 kT/e). All structures were rendered using PyMOL molecular visualization software for residues 2-86.

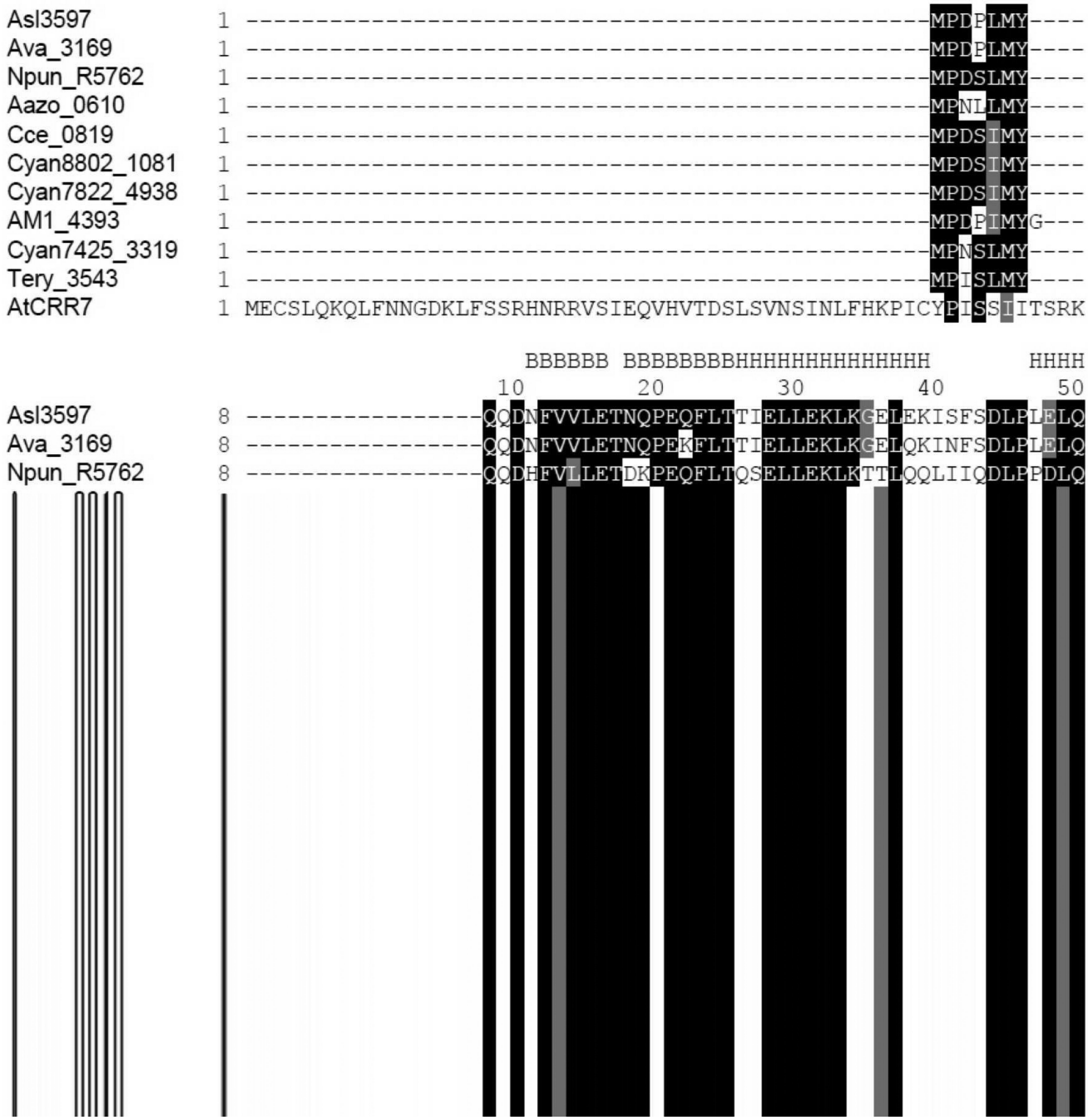

Figure 2.

Alignment of Asl3597 from Nostoc sp. PCC7120 with the top 10 homologous BLAST (KEGG) results using BOXSHADE v. 3.21. CRR7 from Arabidopsis thaliana is also included at the bottom with an accompanying solid black triangle indicating the predicted cleavage site of the plastid targeting signal. Secondary structure elements from Asl3597 are indicated above the amino acid residue numbers in the sequence.

Table 1

Summary of NMR and structural statistics for Asl3597 (PDB ID, 2KRX)

Completeness of resonance assignments[a]		
Backbone (%)		93.0
Side chain (%)		91.7
Aromatic (%)		87.1
Stereospecific methyl (%)		100.0
Conformationally-restricting constraints[b]		
Distance constraints		
Total		1717
Intraresidue ($i = j$)		350
Sequential [(i - 1) = 1]		463
Medium range [1 < ($i – j$) ≤ 5]		437
Long range [($i – j$) > 5]		467
Dihedral angle constraints		136
Hydrogen bond constraints		36
Number of constraints per residue		21.5
Number of long range constraints per residue		5.8
Residual constraint violations[b]		
Average number of distance violations per structure		
0.1 - 0.2 Å		13.65
0.2 – 0.5 Å		4.7
> 0.5 Å		0.1
RMSD from average coordinates (Å)[b,d]		
Backbone atoms		0.5
Heavy atoms		1.0
MolProbity Ramachandran statistics[b,d]		
Most favored regions (%)		92.3
Additionally allowed regions (%)		7.3
Generously allowed regions (%)		0.4
Disallowed regions (%)		0.0
Global quality scores (Raw/ Z-score)[b]		
Verify3D	0.2	-4.8
ProsaII	0.3	-1.4
Procheck (phi-psi)[d]	0.0	0.4
Procheck (all)[d]	0.0	-0.2
Molprobity clash	22.6	-2.4
RPF Scores[c]		
Recall / Precision	0.88	0.89

F-measure / DP-score	0.89	0.75

[a] Refers to chemical shifts for residues 3-83.

[b] Calculated for the ensemble of 20 structures using PSVS version 1.4.[16] Average distance violations were calculated using the sum over r^{-6}.

[c] RPF scores[22] calculated for the ensemble of 20 structures reflecting the goodness-of-fit to the NOESY data and resonance assignments.

[d] Ordered residue ranges: 11-20, 23-65, 69-83, with the sum of ϕ and ψ order parameters > 1.8.