Derivation of structural restraints using a thiol-reactive chelator

Alex Dvoretsky[1], Vadim Gaponenko[1], Paul R. Rosevear*

Department of Molecular Genetics, Biochemistry, and Microbiology, University of Cincinnati, College of Medicine, 231 Albert Sabin Way, Cincinnati, OH 45267, USA

Received 28 June 2002; revised 1 August 2002; accepted 5 August 2002

First published online 29 August 2002

Edited by Thomas L. James

Abstract Recognition and identification of protein folds is a prerequisite for high-throughput structural genomics. Here we demonstrate a simple protocol for covalent attachment of a short and more rigid metal-chelating tag, thiol-reactive EDTA, by chemical modification of the single cysteine residue in barnase(H102C). Conjugation of the metal-chelating tag provides the advantage of allowing a greater range of paramagnetic metal substitutions. Substitution of Yb^{3+}, Mn^{2+}, and Co^{2+} permitted measurement of metal–amide proton distances, dipolar shifts, and residual dipolar couplings. Paramagnetic-derived restraints are advantageous in the NMR structure elucidation of large protein complexes and are shown sufficient for validation of homology-based fold predictions.

Key words: NMR; Paramagnetic restraint; Residual dipolar coupling; Dipolar shift; Barnase; Protein fold

1. Introduction

A comprehensive description of cellular function will require detailed protein structures. New nuclear magnetic resonance (NMR) methodologies have facilitated protein structure determination. However, data collection, analysis time, as well as a scarcity of long-range distance and orientational restraints continue to impede structure characterization by NMR. This has increased interest in simple approaches for evaluating protein fold predictions and for collection of long-range distance and angular restraints.

New methods that expedite protein fold recognition and structure determination independent of short-range distance restraints are required. Towards this goal, considerable effort has been focused on utilizing residual dipolar couplings (RDCs) and, to a lesser extent, dipolar chemical shifts (DCSs) to facilitate structure elucidation and to obtain the relative orientation of protein domains [1–3]. RDCs can be measured in proteins oriented in the magnetic field such that dipolar interactions do not average to zero. RDCs were initially measured in paramagnetic proteins [4] and later in diamagnetic proteins that partially align in the presence of bicelles and filamentous viruses [5,6]. DCSs have also been exploited in conjunction with short-range restraints to assist in structure elucidation and refinement of metal-binding proteins [7–10]. Recently, we have demonstrated that non-metal-binding proteins can be partially aligned in the magnetic field by fusing a zinc finger tag to the C-terminus of a protein and substitution of the bound zinc with cobalt [11].

Incorporation of metal-binding tags at unique sites in the target protein by chemical modification has the potential to yield a wider range of paramagnetic metal substitutions and protein attachment sites. Here we demonstrate a simple protocol for covalent attachment of a metal-binding tag using site-directed labeling. This approach has been applied to barnase, a ribonuclease secreted by *Bacillus amiloliquifacience*. Covalent attachment of thiol-reactive EDTA (*S*-(2-pyridylthio)cysteaminyl-ethylenediaminetetraacetic acid) to monocysteine derivatives of barnase permitted site-specific incorporation of Yb^{3+}, Mn^{2+}, Co^{2+}, and Zn^{2+}. Varying the tag-bound metal permitted the measurement of RDCs, DCSs, and metal–nuclear distances advantageous in NMR structure determination of larger proteins and protein complexes. In addition, the paramagnetic-derived restraints are useful for validation of homology-based fold predictions.

2. Materials and methods

2.1. Monocysteine barnase proteins

Barnase(H102C) and barnase(H102A) were ^{15}N labeled and purified as previously described [12]. The conjugation of barnase(H102C) with thiol-reactive EDTA was performed as described previously [13]. The extent of sulfhydryl labeling was assessed spectrophotometrically using 5,5′-dithio-bis(2-nitrobenzoic acid) and found to be greater than 95%. Ytterbium, cobalt, and manganese-loaded samples were prepared by titration with ultra-pure $MnCl_2$, $CoCl_2$, and $YbCl_3$. Excess metals were removed by gel filtration and the proteins concentrated to 0.5 mM in a buffer containing 20 mM HEPES, pH 6.9, and 10% ^2H_2O for NMR analysis.

2.2. NMR spectroscopy

NMR spectra were collected on Varian Inova 500, 600, and 800 spectrometers. Assignments for barnase(H102C-EDTA–Co^{2+}) were confirmed by NOESY-HSQC experiments at 500 and 800 MHz using a mixing time of 75 ms. T_1 inversion recovery ^1H–^{15}N HSQC spectra were collected with delay times of 10, 200, 400, 600, 800, 1000, 1400, 1800, 2200, and 3000 ms for barnase(H102C-EDTA–Mn^{2+}) and barn-

*Corresponding author. Fax: (1)-513-558 8474.
E-mail address: paul.rosevear@uc.edu (P.R. Rosevear).

[1] Equal scientific contribution from both authors.

Abbreviations: thiol-reactive EDTA, *S*-(2-pyridylthio)cysteaminyl ethylenediamine tetraacetic acid; barnase(H102C-EDTA), barnase with histidine 102 mutated to cysteine and modified with *S*-(2-pyridylthio)cysteaminyl-ethylenediaminetetraacetic acid; barnase(H102A), barnase with histidine 102 mutated to alanine; α/β TROSY, α/β transverse relaxation optimized spectroscopy; RDC, residual dipolar coupling; DCS, dipolar chemical shift; NMR, nuclear magnetic resonance; rms, root mean square

0014-5793/02/$22.00
PII: S0014-5793(02)03297-0

ase(H102A). Spectra were processed using the Felix software (Accelrys) with resolution enhancement as previously described [11].

2.3. Distance calculations

Paramagnetic contributions to the ^1H_N relaxation rates, $1/T_{1p}$, were calculated from the difference between the longitudinal relaxation rates of barnase(H102C-EDTA–Mn^{2+}) and barnase(H102A) [14]. The similarity in ^1H_N correlation times, measured from the frequency dependence of the paramagnetic effects on T_{1p} [11], justified using an average uniform correlation time of 2.9 ns. Distances were calculated using the Solomon–Bloembergen equation [15] as previously described [11].

2.4. Measurement of RDCs

Dipolar couplings were collected using a generalized α/β transverse relaxation optimized spectroscopy (TROSY) experiment [16]. RDCs were obtained from the differences in $^1J_{HN}$ couplings observed in the apo- or Zn^{2+}-loaded (unoriented) and the Co^{2+}-loaded (partially oriented) proteins. The magnitude and the orientation of the alignment tensor were calculated as described previously [11].

2.5. Measurement of DCSs

DCSs were measured as ^1H_N chemical shift differences between metal-loaded Yb^{3+} or Co^{2+}, and apo-samples of barnase(H102C-EDTA). Chemical shifts were measured from high-resolution 500 and 600 MHz ^1H–^{15}N HSQC spectra using ^1H and ^{15}N spectral widths of 8 and 2.2 kHz, respectively, and 512 increments in the indirect dimension.

3. Results

Introduction of paramagnetic probes into non-metal-binding proteins can be achieved through alkylation of unique cysteine residues with thiol-reactive EDTA (Fig. 1). As wild-type barnase lacks cysteine residues, a unique cysteine was introduced at position 102 using site-directed mutagenesis. The purpose of this study is to evaluate the use of a site-directed metal-binding tag to obtain long-range distance and orientational restraints useful in protein fold verification and in structure determinations of larger proteins. No significant amide proton chemical shift differences were observed between barnase(H102C-EDTA) and barnase(H102C), suggesting that covalent attachment of the EDTA tag did not significantly perturb protein structure.

3.1. Metal–^1H_N distances

As expected, the addition of Mn^{2+} to barnase(H102C-EDTA) resulted in selective line-broadening of ^1H–^{15}N correlations in the HSQC spectra. Paramagnetic relaxation enhancement methodology was used to estimate ^1H_N–metal distances. To quantitate the paramagnetic effect of Mn^{2+} on the ^1H_N nuclei, proton longitudinal relaxation rates, $1/T_1$, for amide protons of barnase(H102C-EDTA–Mn^{2+}) were measured from a series of inversion recovery experiments collected at 500, 600, and 800 MHz. Measured $1/T_1$ values in barnase(H102A) were used to estimate the diamagnetic contributions to the measured relaxation rates. Longitudinal relaxation rates measured after addition of 2 mM $MnCl_2$ to barnase(H102A) demonstrated that effects from the non-specific binding of Mn^{2+} were negligible. Individual paramagnetic contributions to the measured amide proton relaxation rates, $1/T_{1p}$, were determined by subtraction of $1/T_1$ values for barnase(H102A) from $1/T_1$ values for barnase(H102C-EDTA–Mn^{2+}). Together with the estimated correlation time, 2.9 ns, ^1H_N–metal distances were calculated using the Solomon–Bloembergen equation [15]. Metal–^1H_N distances could be measured for 66 of the 70 resolvable amide backbone protons. Measured distances ranged from 18 to 46 Å with an average error of ~11%. The correlation between fitted and experimentally obtained Mn^{2+}–^1H_N distances for barnase(H102C-EDTA–Mn^{2+}) is shown in Fig. 2. The average difference in ^1H_N–metal distance obtained from paramagnetic relaxation enhancement and those calculated from amide proton DCSs was determined to be 3.2 Å.

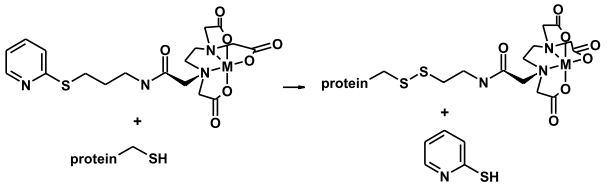

Fig. 1. Reaction of thiol-reactive EDTA with –SH groups of a protein to produce the side-chain EDTA derivative.

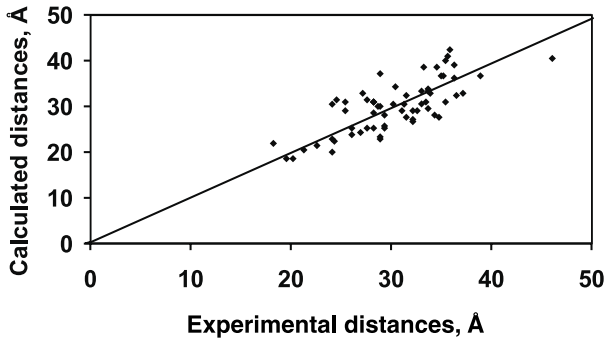

Fig. 2. Correlation between calculated (based on the crystal structure of barnase) and experimental Mn^{2+}–^1H_N distances for barnase(H102C-EDTA–Mn^{2+}). The correlation coefficient is 77%. Experimental distances were obtained from paramagnetic effects of bound Mn^{2+} on ^1H_N longitudinal relaxation rates collected at 500, 600, and 800 MHz.

3.2. Dipolar chemical shifts

Barnase(H102C-EDTA) with Yb^{3+} or Co^{2+} bound was prepared by titration with the appropriate metal salt. To validate the accuracy of DCSs measured using a paramagnetic metal-binding tag, measured values were used to orient the paramagnetic susceptibility tensor relative to the crystal structure of barnase [11,17], allowing a direct comparison of measured DCSs and those calculated based on the crystal structure. The correlation between the measured and predicted ^1H_N DCSs in barnase(H102C-EDTA–Co^{2+}) and (H102C-EDTA–Yb^{3+}) are shown in Fig. 3.

3.3. Residual dipolar couplings

RDCs for ^1H–^{15}N bond vectors have been shown to be an excellent tool for protein structure refinement [1,18,19]. In particular, low-resolution protein folds determined from paramagnetic relaxation enhancement and DCSs can be refined using RDCs. Binding of Co^{2+} to barnase(H102C-EDTA) induced sufficient magnetic susceptibility anisotropy to measure RDCs using a generalized α/β TROSY experiment [16]. RDCs ranged from −4 to 4 Hz and were measured by subtraction of one-bond ^1H–^{15}N couplings for barnase(H102A) from barnase(H102C-EDTA–Co^{2+}). The fitted axial component of the magnetic susceptibility was negative as predicted by the measured RDCs. The calculated alignment tensor was found to be axial ($|\Delta\chi_{ax}| = 1.49 \times 10^{-31}$ m^3) and in agreement with both

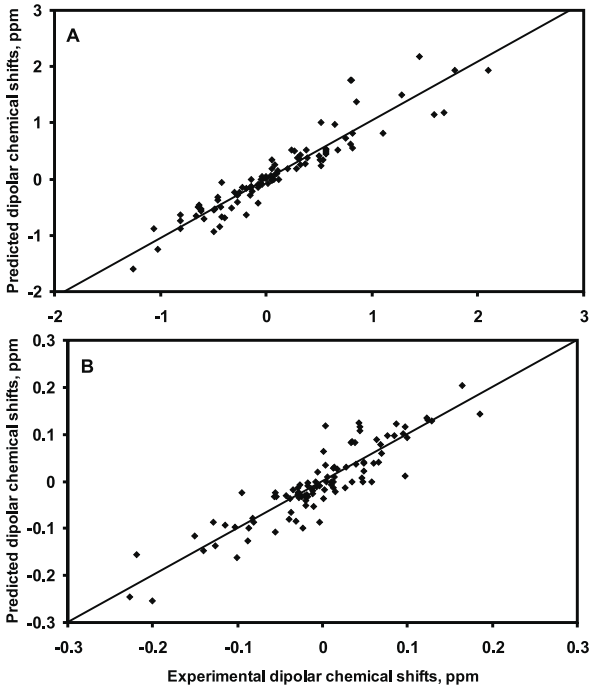

Fig. 3. Experimentally determined versus predicted amide proton DCSs for barnase(H102C-EDTA–Co^{2+}) (A) and barnase(H102C-EDTA–Yb^{3+}) (B). Correlation coefficients of 95% and 90%, respectively, were determined for the Co^{2+} and Yb^{3+}-tagged protein. Predicted DCSs were calculated by a tensor optimization procedure based on the crystal structure of wild-type barnase [17].

the paramagnetic susceptibility tensor determined from DCSs and the symmetry of the magnetic susceptibility deduced from a 35 GHz Q-band EPR spectrum at 2 K (data not shown). The magnitude and orientation of the alignment tensor obtained from the RDCs was indistinguishable by the *F*-test from the paramagnetic susceptibility tensor. This suggests that the alignment of the protein in the external magnetic field is dominated by the effect of bound Co^{2+}.

4. Discussion

The presence of a paramagnetic center can yield a wealth of long-range structural information in naturally occurring metal-binding proteins [7,20,21]. This approach can be expanded to non-metal-binding proteins using covalently attached metal-binding tags at unique positions within the target protein. We have utilized a broadly applicable method for precise introduction of a paramagnetic center into a target protein through the use of site-directed cysteine mutants reacted with a thiol-specific EDTA tag. The ability to incorporate a range of paramagnetic metals into the covalently attached metal-binding tag provides a variety of magnetically induced orientations useful in resolving degeneracy in amide proton bond vectors. Furthermore, introduction of Mn^{2+} allowed the direct measurement of long-range metal–nuclear distances using paramagnetic relaxation enhancement methodology. A linear relationship between distances calculated from the crystal structure of wild-type barnase and experimental Mn^{2+}-^1H_N distances for barnase(H102C-EDTA–Mn^{2+}) was observed (Fig. 2). Measured Mn^{2+}–^1H_N distances ranged from 18 to 46 Å with an average error of 3.2 Å.

Introduction of a paramagnetic center with significant anisotropy induces DCSs that can be related to the distance and angle between the principal axis of the paramagnetic susceptibility tensor and the metal–nucleus vector. Measured DCSs for both Yb^{3+} and Co^{2+}-loaded barnase(H102C-EDTA) could be accurately correlated with those predicted from the crystal structure of barnase (Fig. 3). Similarly, RDCs measured in Yb^{3+} and Co^{2+}-loaded barnase(H102C-EDTA) could be correlated with those calculated using the crystal structure of barnase (data not shown). These results demonstrate that incorporation of a paramagnetic reference frame into a non-metal-binding protein, through covalent attachment of a metal-binding tag, provides important long-range distance and orientational restraints useful in protein structure determination.

The thiol-specific EDTA tag offers numerous advantages over a terminal metal-binding tag to provide a site for incorporation of a paramagnetic probe. In our previous work, a zinc finger from the nucleic acid-binding domain of Rauscher murine leukemia virus was attached at the C-terminus of barnase [11]. While this permitted measurement of RDCs, DCSs, and Mn^{+2}–^1H_N distances, flexibility in the linker between barnase and the fused zinc finger severely limited the accuracy of the measurements [11]. Flexibility in the linker also impacts the magnitude of the RDCs and DCSs through increased motional averaging. A disulfide linked EDTA tag was chosen in order to provide a more rigid linkage between the protein and the metal ligand [22,23]. This increased rigidity is evidenced by the improved precision of measured paramagnetic restraints in this system (Figs. 2 and 3).

Unlike the zinc finger tag, the attachment of thiol-reactive EDTA is not restricted to only the N- and C-termini. The EDTA tag permits introduction of metal-binding sites at multiple points along the protein sequence. The ability of EDTA to chelate a wide variety of cations increases the range of paramagnetic metal ions available for incorporation. It has been demonstrated that attachment of chemical tags at solvent-exposed sites induces only minor structural perturbations to the backbone fold [24]. Thus, the thiol-specific EDTA tag provides a more robust method for specific incorporation of metal ions into a protein of interest.

Paramagnetic-derived structural restraints can also be used to validate and improve fold prediction based on threading-alignment methodologies. It has previously been shown that secondary structure information and a limited number of NOEs can significantly improve the threading quality in both fold recognition and threading alignment [25]. Main-chain RDCs can be used to discriminate between similar and dissimilar folds from homology-based modeling predictions [19]. To illustrate the sensitivity in using DCSs and metal–nuclear distances to validate protein fold predictions, two folds having differing sequence homology to barnase were fit to the experimental Mn^{2+}–^1H_N distances and DCSs. Both folds contained a central β-sheet, at least one α-helix, and belong to the same family as barnase. Binase (PDB# 1BUJ) [26], having 84% identity with barnase, was fit to the Mn^{2+}–^1H_N distances experimentally measured for barnase(H102C-EDTA–Mn^{2+}), yielding a unique position for Mn^{2+} relative to the binase structure with a correlation coefficient of 75%. This position for the Mn^{2+} site was used to fit the measured DCSs onto the binase structure with a correlation coefficient of 89%, indicating binase represents a good candidate fold. This result is consistent with a 1.2 Å root mean square (rms)

deviation between the structure of binase and the crystal structure of barnase. In addition, the structural similarity between barnase and binase can be quantitated using the Z-score [27,28]. The Z-score is a measure of structural similarity, derived utilizing a distance matrix approach, with a higher score indicating a greater degree of structural similarity [28]. The Z-score derived from comparing barnase and binase is 18.9.

In contrast, RNase M (PDB# 1BUJ) [29], which has only a 22% identity with barnase, yielded a poor fit to the Mn^{2+}–^1H_N distances, with a correlation coefficient of 33%. Likewise, a poor correlation was obtained when the position for Mn^{2+} relative to the RNase M structure was used to fit the DCSs: a correlation coefficient of 46%. This poor fit is in agreement with the 3.2 Å rms deviation between the structures of RNase M and barnase. The Z-score [27,28] comparing the two structures is 3.7. This low score suggests only slight structural similarity between barnase and RNase M (proteins with Z-score < 2.0 are considered to be structurally dissimilar).

In this query, Mn^{2+}–^1H_N distances in combination with amide proton DCSs appear sufficient to discriminate between homologous and non-homologous folds. Thus, easily obtainable structural restraints using a thiol-reactive chelator should prove useful when experimentally evaluating protein threading or homology-based modeling predictions.

Acknowledgements: This work was supported by Grants AR 44324 and GM 62153.

References

[1] Clore, G.M. and Gronenborn, A.M. (1998) Proc. Natl. Acad. Sci. USA 95, 5891–5898.
[2] Bewley, C.A. (2001) J. Am. Chem. Soc. 123, 1014–1015.
[3] Tjandra, N., Marquardt, J. and Clore, G.M. (2000) J. Magn. Reson. Imaging 142, 393–396.
[4] Tolman, J.R., Flanagan, J.M., Kennedy, M.A. and Prestegard, J.H. (1995) Proc. Natl. Acad. Sci. USA 92, 9279–9283.
[5] Tjandra, N. and Bax, A. (1997) Science 278, 1111–1114.
[6] Hansen, M.R., Mueller, L. and Pardi, A. (1998) Nat. Struct. Biol. 5, 1065–1074.
[7] Gochin, M. and Roder, H. (1995) Protein Sci. 4, 296–305.
[8] Biekofsky, R.R., Muskett, F.W., Schmidt, J.M., Martin, S.R., Browne, J.P., Bayley, P.M. and Feeney, J. (1999) FEBS Lett. 460, 519–526.
[9] Banci, L., Bertini, I., Bren, K.L., Cremonini, M.A., Gray, H.B., Luchinat, C. and Turano, P. (1996) J. Biol. Inorg. Chem. 1, 117–126.
[10] Arnesano, F., Banci, L., Bertini, I. and Felli, I.C. (1998) Biochemistry 37, 173–184.
[11] Gaponenko, V., Dvoretsky, A., Walsby, C., Hoffman, B.M. and Rosevear, P.R. (2000) Biochemistry 39, 15217–15224.
[12] Gaponenko, V., Howarth, J.W., Columbus, L., Gasmi-Seabrook, G., Yuan, J., Hubbell, W.L. and Rosevear, P.R. (2000) Protein Sci. 9, 302–309.
[13] Ebright, Y.W., Chen, Y., Pendergrast, P.S. and Ebright, R.H. (1992) Biochemistry 31, 10664–10670.
[14] Mildvan, A.S., Granot, J., Smith, G.M. and Liebman, M.N. (1980) in: Structure and Function of Metalloproteins (Darnell, A.W. and Wilkins, R.G., Eds.), Vol. 2, pp. 211–235, Elsevier, Amsterdam.
[15] Solomon, I. (1955) Phys. Rev. 99, 559–565.
[16] Andersson, P., Annila, A. and Otting, G. (1998) J. Magn. Reson. Imaging 133, 364–367.
[17] Mauguen, Y., Hartley, R.W., Dodson, E.J., Dodson, G.G., Bricogne, G., Chothia, C. and Jack, A. (1982) Nature 297, 162–164.
[18] Chou, J.J., Li, S., Klee, C.B. and Bax, A. (2001) Nat. Struct. Biol. 8, 990–997.
[19] Aitio, H., Annila, A., Heikkinen, S., Thulin, E., Drakenberg, T. and Kilpelainen, I. (1999) Protein Sci. 8, 2580–2588.
[20] Hus, J.C., Marion, D. and Blackledge, M. (2000) J. Mol. Biol. 298, 927–936.
[21] Contreras, M.À., Ubach, J., Millet, Ò. and Pons, M. (1999) J. Am. Chem. Soc. 121, 8947–8948.
[22] Columbus, L., Kalai, T., Jeko, J., Hideg, K. and Hubbell, W.L. (2001) Biochemistry 40, 3828–3846.
[23] Langen, R., Oh, K.J., Cascio, D. and Hubbell, W.L. (2000) Biochemistry 39, 8396–8405.
[24] McHaourab, H.S., Lietzow, M.A., Hideg, K. and Hubbell, W.L. (1996) Biochemistry 35, 7692–7704.
[25] Xu, Y., Xu, D., Crawford, O.H. and Einstein, J.R. (2000) J. Comput. Biol. 7, 449–467.
[26] Reibarkh, M., Nolde, D.E., Vasilieva, L.I., Bocharov, E.V., Shulga, A.A., Kirpichnikov, M.P. and Arseniev, A.S. (1998) FEBS Lett. 431, 250–254.
[27] Holm, L. and Sander, C. (1993) J. Mol. Biol. 233, 123–138.
[28] Holm, L. and Sander, C. (1996) Science 273, 595–602.
[29] Nonaka, T., Nakamura, K.T., Uesugi, S., Ikehara, M., Irie, M. and Mitsui, Y. (1993) Biochemistry 32, 11825–11837.