Journal of Biomolecular NMR, **11:** 97–102, 1998.
KLUWER/ESCOM

An efficient and cost-effective isotope labeling protocol for proteins expressed in *Escherichia coli*

Mengli Cai[a], Ying Huang[b], Kazuyasu Sakaguchi[c], G. Marius Clore[a,*], Angela M. Gronenborn[a,*] and Robert Craigie[b]
[a]*Laboratory of Chemical Physics and* [b]*Laboratory of Molecular Biology, Building 5, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892-0520, U.S.A.* [c]*Laboratory of Cell Biology, Building 37, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, U.S.A.*

Received 15 July 1997; Accepted 3 October 1997

Key words: heteronuclear NMR, isotope labeling, protein expression

Abstract

A cost-effective protocol for uniform ^{15}N and/or ^{13}C isotope labeling of bacterially expressed proteins is presented. Unlike most standard protocols, cells are initially grown in a medium containing nutrients at natural abundance and isotopically labeled nutrients are only supplied at the later stages of growth and during protein expression. This permits the accumulation of a large cell mass without the need to employ expensive isotopically labeled nutrients. The abrupt decrease in oxygen consumption that occurs upon complete exhaustion of essential nutrients is used to precisely time the switch between unlabeled and labeled nutrients. Application of the protocol is demonstrated for wild-type and a mutant of the N-terminal zinc-binding domain of HIV-1 integrase.

The size of macromolecular structures that can be solved by NMR has been dramatically increased over the last decade largely due to the advent of three- and four-dimensional heteronuclear spectroscopy (Clore and Gronenborn, 1991a,b; Bax and Grzesiek, 1993). Application of these techniques to proteins generally requires uniform ^{15}N and ^{13}C labeling. This is commonly achieved by expressing the protein of interest in a bacterial host, typically *E. coli*, using a growth medium containing ^{15}N and ^{13}C as the sole nitrogen and carbon sources, respectively. Modified minimal media containing $^{15}NH_4Cl$ or $(^{15}NH_4)_2SO_4$ as nitrogen sources (Muchmore et al., 1989) and ^{13}C-glucose, glycerol or acetate as carbon sources (Venters et al., 1991) have been employed. In addition, media containing isotopically enriched algal or microbial hydrolysates have been used (Reilly and Fairbrother, 1994). Although the costs for ^{15}N- and ^{13}C-labeled precursors have decreased substantially, for cases where protein expression is moderate, it may be desirable to achieve as high a yield of isotopically labeled protein per gram of labeled nutrient as possible. In this communication we present a simple strategy and protocol that achieves this goal. We demonstrate that by growing the bacterial cells initially in medium at natural isotopic abundance and providing isotopically labeled nutrients only after induction, the yield of labeled protein (>95%) per unit nutrient can be increased more than eightfold for ^{15}N labeling and threefold for ^{13}C,^{15}N labeling.

We developed and tested our protocol using as an example the N-terminal zinc binding domain of HIV-1 integrase (IN^{1-55}) and a mutant thereof, IN^{1-55} H12C. Both proteins were expressed in a T7 RNA polymerase-based system (Studier et al., 1990) employing pET-15b (Novagen) for cloning and *E. coli* Bl21(DE3) as the bacterial host. Two liter cultures were grown in a 10 l Bioflow IV fermentor (New Brunswick Scientific) controlled by AFS-BioCommand Bioprocessing Software (New Brunswick Scientific). A modified M9 medium (Table 1) was used as the basic growth medium which was either supplemented with $^{14}NH_4Cl$ or $^{15}NH_4Cl$

* To whom correspondence should be addressed.

Table 1. Recipe for 1 l of growth medium

1.	970 ml basic solution (anhydrous salts, autoclaved)	per 970 ml
	KH_2PO_4	13.0 g
	K_2HPO_4	10.0 g
	Na_2HPO_4	9.0 g
	K_2SO_4	2.4 g
	NH_4Cl	Varies (see text)
2.	10 ml trace element solution (sterile filtered)[a]	per 100 ml
	$FeSO_4$ ($7H_2O$)	0.60 g
	$CaCl_2$ ($2H_2O$)	0.60 g
	$MnCl_2$ ($4H_2O$)	0.12 g
	$CoCl_2$ ($6H_2O$)	0.08 g
	$ZnSO_4$($7H_2O$)	0.07 g
	$CuCl_2$ ($2H_2O$)	0.03 g
	H_3BO_3	2 mg
	$(NH_4)_6Mo_7O_{24}$ ($4H_2O$)	0.025 g
	EDTA	0.50 g
3.	10 ml 1 M $MgCl_2$ (sterile filtered)	
4.	6 ml thiamine (5 mg/ml, sterile filtered)[b]	
5.	1 ml ampicillin (100 mg/ml, sterile filtered)	
6.	D-Glucose (varies, see text for details)	
7.	10 µl 10% yeast extract (sterile filtered)[c]	

[a] The addition of the trace element solution ensures that the large number of metal ion containing enzymes in *E. coli* can function optimally.

[b] Thiamine (vitamin B_1) is provided since many commercial strains of *E. coli* are vitamin B_1 deficient.

[c] The addition of trace amounts of yeast extract seems to provide vitamins and cofactors that will allow for a reduced lag phase in the growth curve, although it is generally not essential.

(=98% ^{15}N; Cambridge Isotope Laboratories Inc.) and $^{12}C_6$-D-glucose or $^{13}C_6$-D-glucose (=98% ^{13}C; Cambridge Isotope Laboratories Inc.). The pH of the medium was maintained at 6.8 by titrating with 1 M NaOH and the oxygen level was kept between 15% and 18% (calibrated to 0% by nitrogen purging and 100% by oxygen purging). The dissolved oxygen concentration was maintained by adjusting the agitation rate while keeping the air flow constant (4–7 l/min). The oxygen level was monitored every 30 s and the agitation rate was increased by 2 rpm when the oxygen content fell below 15% or decreased by 2 rpm when it rose above 18%. During log-phase growth, oxygen demand and agitation rate are proportional to cell density. Exhaustion of any essential nutrient in the medium results in abrupt shutdown of cellular metabolism, indicated by a sharp drop in oxygen demand. This, in turn, causes a sudden rise in the dissolved oxygen level which is used to time the addition of further nutrients (i.e. the switch from unlabeled nutrients to isotopically labeled nutrients), thus avoiding dilution or starvation of cells. Although nutrient additions were carried out manually in the present protocol, it would be simple to automate this process.

Initial experiments were carried out to establish the relationship between generation of cell mass and amount of NH_4Cl and glucose consumed from our medium. Using the medium described in Table 1, a temperature of 37 °C and an oxygen level of 15–18%, we achieved an OD_{600} of ~5 for each g/l of NH_4Cl (in the presence of unlimiting amounts of glucose) and ~1.4 for each g/l of glucose (in the presence of excess NH_4Cl). These data provide the basis for the design of optimal growth media for either ^{15}N or ^{15}N,^{13}C labeling.

Figure 1 illustrates a typical growth for ^{15}N,^{13}C double labeling of the $IN^{1-55}H12C$ mutant. In this case the medium contained 2 g/l of ^{12}C-D-glucose and excess $^{15}NH_4Cl$ (2 g/l). Initially (time point a), with the minimum agitation rate set at 50 rpm and an air flow rate of 4 l/min, the oxygen supply was greater than its consumption, and thus the dissolved

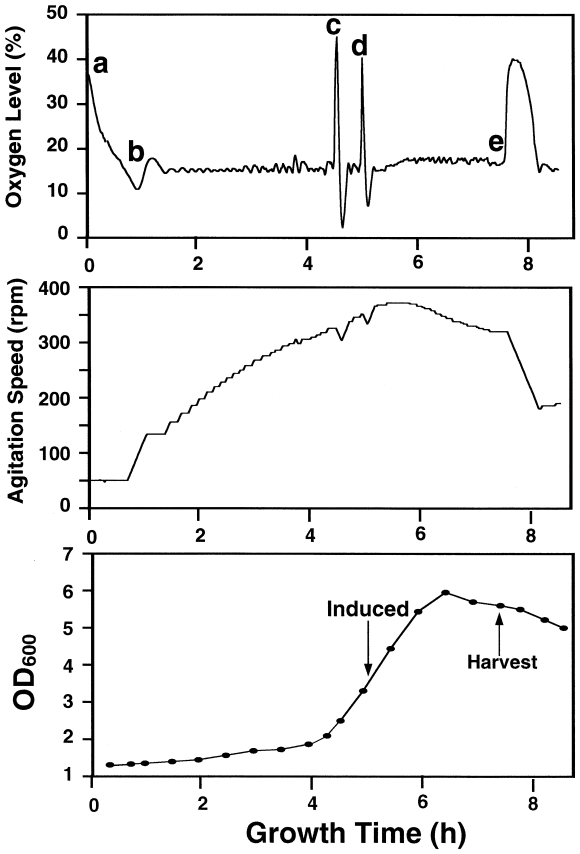

Figure 1. Oxygen level, agitation speed and OD_{600} as a function of time for $^{15}N,^{13}C$ labeling of the IN^{1-55}H12C mutant. The time points are as follows: (a) start of run; (b) start of cell growth; (c) depletion of essential nutrients and addition of labeled nutrient ($^{13}C_6$-glucose); (d) depletion of essential labeled nutrient, addition of labeled nutrient ($^{13}C_6$-glucose) and induction; and (e) cell harvest.

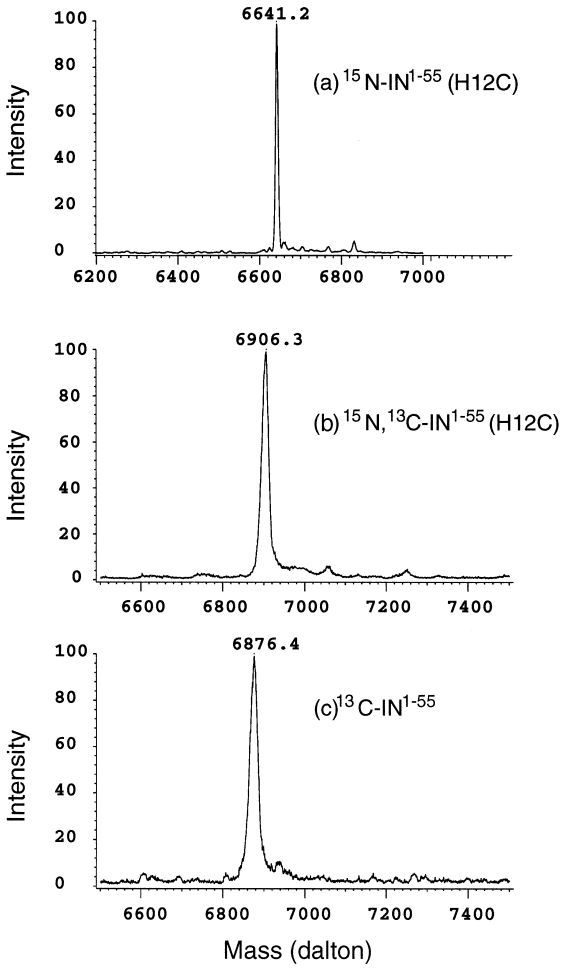

Figure 2. Mass spectra of isotopically enriched proteins: (a) ^{15}N-labeled IN^{1-55}H12C mutant; (b) $^{15}N,^{13}C$-labeled IN^{1-55}H12C mutant; and (c) wild-type ^{13}C-labeled IN^{1-55}.

oxygen level is above the set range. After about an hour, the oxygen demand increased and at time point b the dissolved oxygen level fell below the set range. Subsequently, the agitation speed progressively increased to maintain the dissolved oxygen level within the set range (middle panel). A sudden drop in oxygen demand and a consequent abrupt rise in the level of dissolved oxygen at time point c marked the exhaustion of an essential nutrient, in this case glucose. The addition of 0.5 g of $^{13}C_6$-D-glucose resulted in an immediate increase in oxygen demand and a concomitant drop in the dissolved oxygen to the set level. The addition of labeled nutrient at this point allows for any remaining small intracellular amounts of unlabeled nutrients to be consumed before protein production is initiated. Consumption of this additional amount of glucose was again indicated by an abrupt rise in dissolved oxygen. At this point (time point d), protein expression was induced by the addition of IPTG (at a final concentration of 1 mM) and a further 2.5 g of $^{13}C_6$-D-glucose. The $OD_{600}$was about 3.6 at the time of induction. After induction, the cellular metabolism slows down and the oxygen demand eventually falls as expected. Cells were harvested 2 h after induction (time point e), at which time all glucose was consumed. The harvesting point, as well as the final quantity of labeled glucose, needs to be determined empirically since it depends on the density of the culture at the time of induction, the length of induction for optimal yield of protein, and the response of the host cell to the expressed protein. For the example described above and illustrated in Figure 1, approximately 15 mg of labeled protein was purified from

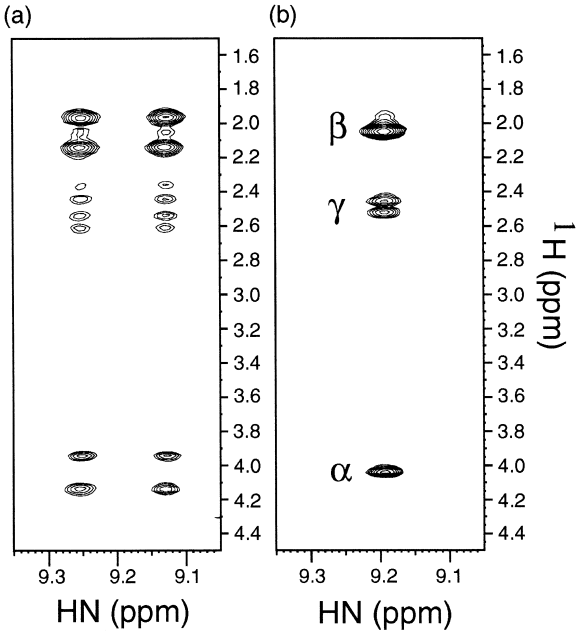

Figure 3. Selected region of the 2D ^{1}H-^{1}H NOE spectrum (120 ms mixing time) recorded at 30 °C on the ^{15}N,^{13}C-labeled IN$^{1-55}$H12C mutant (2 mM) in water without (a) and with (b) ^{15}N and ^{13}C decoupling. The spectra were recorded on a Bruker DMX750 spectrometer and decoupling in (b) was achieved using low-power (1.5 kHz rf) Waltz decoupling for ^{15}N during the acquisition period and a ^{13}C 180° hard pulse in the middle of the t_1 evolution period for ^{13}C decoupling.

1 l of culture, a yield that is 5 times greater than that achieved by growing in shake flasks. Comparison of the amount of protein per unit of labeled glucose yields a threefold increase given the fact that we used 2 g of labeled glucose/l for growth in flasks and a total of 3 g/l of labeled glucose in the fermentor.

The general procedure outlined above has also been modified for single ^{15}N and single ^{13}C labeling as illustrated in Table 2. As can be seen from the results, large amounts of ^{15}N-labeled protein can be obtained, in particular 38 mg of purified ^{15}N-labeled IN$^{1-55}$ H12C mutant per liter of culture. This is substantially more than the yield achieved by growing in shake flasks using labeled $^{15}NH_4Cl$ throughout (∼3 mg/l). This is equivalent to a ≥ eightfold increase in labeled protein per unit of labeled nutrient (i.e. ∼25 mg of protein per gram of $^{15}NH_4Cl$ with the above protocol versus 3 mg of protein per gram of $^{15}NH_4Cl$ in shake flasks). In general, ^{15}N labeling is not a particularly expensive procedure, but ^{13}C labeling is frequently regarded as such. As can be seen from the data presented in Table 2, a singly labeled ^{13}C sample was obtained with a final yield of purified protein of 5.0 mg per gram of $^{13}C_6$-D-glucose. The single ^{13}C labeling protocol was aimed at increasing the final isotopic enrichment of the purified protein by adding a small amount of labeled glucose twice instead of just once prior to induction and expression, resulting in only a modest increase in added glucose.

All proteins were purified by Ni^{2+} affinity chromatography followed by reverse-phase HPLC as described by Cai et al. (1997). The level of isotope enrichment was determined by electrospray ionization mass spectrometry using a Finnigan MAT SSQ 7000 spectrometer (Finnigan Mat, San Jose, CA) and NMR spectroscopy. Mass spectra of HPLC purified ^{15}N-, ^{15}N,^{13}C- and ^{13}C-labeled proteins are shown in Figure 2. The molecular mass for the IN$^{1-55}$H12C mutant at natural isotopic abundance was calculated to be 6566.94 Da and that for 100% ^{15}N-labeled and 100% ^{15}N,^{13}C-labeled protein 6644.94 and 6924.94 Da, respectively. The experimentally determined masses were 6641.2 and 6906.3 Da for the single and double labeled mutant, indicative of ^{15}N and ^{13}C enrichment levels of ∼95%. Similar results were obtained for other proteins grown with the equivalent protocol. If higher levels of enrichment are desired, two small additions of labeled nutrient can be made before induction and protein expression (Table 2) as exemplified by the results shown for the singly ^{13}C-labeled wild-type IN$^{1-55}$ protein (Figure 2c). For the wild-type domain the calculated mass for unlabeled and 100% ^{13}C-enriched protein is 6600.45 and 6883.23 Da, respectively. The experimentally determined mass of 6874.4 Da translates into an isotopic enrichment of 97.6%, which is within experimental error of the enrichment level for the starting nutrient.

In general, enrichment levels of ∼95% are sufficient for all of the heteronuclear NMR experiments currently employed for assignment and structure determination of proteins. Figure 3 shows a portion of a 2D ^{1}H-^{1}H NOE spectrum of the ^{15}N,^{13}C-labeled IN$^{1-55}$H12C mutant recorded in 90% H_2O/10% D_2O at pH 7.0 without (a) and with (b) ^{15}N and ^{13}C decoupling. NOE cross peaks from the α, β and γ side-chain protons of Glu^{13} to its backbone amide proton are depicted. In the spectrum recorded without decoupling (Figure 3a), each cross peak is split in the acquisition dimension by ∼97 Hz corresponding to the one-bond ^{1}H-^{15}N amide coupling, and in the indirect dimension by about 130–150 Hz resulting from the one-bond ^{1}H-^{13}C couplings. As can be appreciated from this spectrum, NOE cross peaks from unlabeled protein were

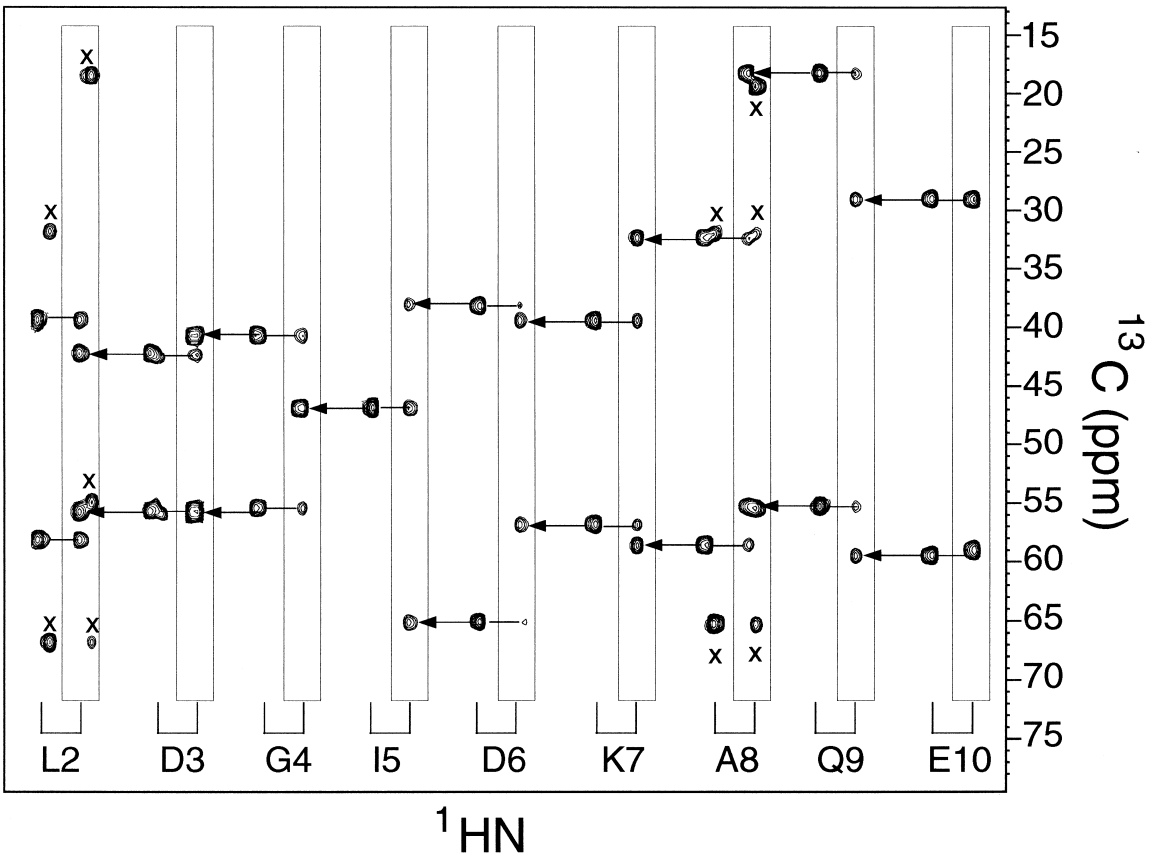

Figure 4. Strips from the 3D CBCANH (boxed) and CBCA(CO)NH triple-resonance experiments recorded at 30 °C on the ^{15}N,^{13}C-labeled IN$^{1-55}$H12C mutant showing sequential connectivities for residues 2–10 of the protein. Cross peaks marked with an x arise from other residues which have their maximum intensities in an adjacent ^{15}N plane in the 3D spectrum. The spectra were recorded on a Bruker DMX500 spectrometer with 16 scans per increment, and aquisition times of 6.26, 18.38 and 63.24 ms in the t_1, t_2 and t_3 dimensions, respectively.

Table 2. Protocols for different labeling schemes

Time point[a]	^{13}C,^{15}N	^{15}N	^{13}C
a	2 g $^{15}NH_4Cl$ 2 g $^{12}C_6$-D-glucose	12 g $^{12}C_6$-D-glucose 1 g $^{14}NH_4Cl$	2 g $^{14}NH_4Cl$ 0.5 g $^{12}C_6$-D-glucose
c	0.5 g $^{13}C_6$-D-glucose	0.25 g $^{15}NH_4Cl$	0.1 g $^{13}C_6$-D-glucose
c′			0.1 g $^{13}C_6$-D-glucose
d	Induction at OD_{600} ~3.6 2.5 g $^{13}C_6$-D-glucose	Induction at OD_{600} ~6.5 1.25 g $^{15}NH_4Cl$	Induction at OD_{600} ~1.0 0.7 g $^{13}C_6$-D-glucose
e	Harvest at OD_{600} ~6.0 15 mg purified protein	Harvest at OD_{600} ~10.5 38 mg purified protein	Harvest at OD_{600} ~1.6 4.5 mg purified protein

The basic medium is that described in Table 1 and only the variable amounts of nutrients per liter of medium are listed here.

[a] The time points a, c, d and e are indicated in Figure 1 and discussed in the text. Time point c′ (between time points c and d) corresponds to a second small addition of glucose prior to induction and expression (time point d) that was employed in the single ^{13}C labeling protocol (see text).

either extremely weak or not observed at all. Indeed, using our normal suite of 3D triple-resonance experiments complete assignments for the mutant integrase domain were achieved. Figure 4 presents examples of strips from the 3D CBCANH and CBCA(CO)NH experiments illustrating sequential connectivities along the polypeptide backbone for the first 10 residues of the ^{15}N,^{13}C-labeled IN$^{1-55}$H12C mutant. The excellent signal-to-noise ratio is readily appreciated from this figure.

In addition to substantially reducing the costs for the preparation of ^{15}N- and/or ^{13}C-labeled proteins, the labeling protocol presented in this paper also allows for efficient and cost-effective deuteration of proteins. A much higher cell mass per unit volume can be achieved in the fermentor, thereby limiting the amount of D_2O necessary for protein production. Furthermore, modifications of the above-described procedure can be easily implemented for different types of labeled nutrients.

Acknowledgements

We thank Ronglan Zheng for the plasmid expressing wild-type IN$^{1-55}$and the IN$^{1-55}$H12C mutant. This work was supported by the AIDS Targeted Anti-Viral Program of the Office of the Director of the National Institutes of Health (to G.M.C., R.C. and A.M.G.).

References

Bax, A. and Grzesiek, S. (1993) *Acc. Chem. Res.*, **26**, 131–138.
Cai, M., Zheng, R., Caffrey, M., Craigie, R., Clore, G.M. and Gronenborn, A.M. (1997) *Nat. Struct. Biol.*, **4**, 567–577.
Clore, G.M. and Gronenborn, A.M. (1991a) *Science*, **252**, 1390–1399.
Clore, G.M. and Gronenborn, A.M. (1991b) *Prog. Nucl. Magn. Reson. Spectrosc.*, **23**, 43–92.
Muchmore, D.C., McIntosh, L.P., Russel, C.B., Anderson, D.E. and Dahlquist, F.W. (1989) *Methods Enzymol.*, **177**, 44–73.
Reilly, D. and Fairbrother, W.J. (1994) *J. Biomol. NMR*, **4**, 459–462.
Studier, F.W., Rosenberg, A.H., Dunn, J.J., and Dubendorf, J.W. (1990) *Methods Enzymol.*, **185**, 60–89.
Venters, R.A., Calderone, T.L., Spicer, L.D. and Fierke, C.A. (1991) *Biochemistry*, **30**, 4491–1194.