Journal of Biomolecular NMR, **18:** 229–238, 2000.
KLUWER/ESCOM

An approach for high-throughput structure determination of proteins by NMR spectroscopy

Ales Medek, Edward T. Olejniczak, Robert P. Meadows & Stephen W. Fesik*
Pharmaceutical Discovery Division, Abbott Laboratories, Abbott Park, IL 60064, U.S.A.

Received 23 June 2000; Accepted 10 August 2000

Key words: cryoprobe, high-throughput, protein, structure

Abstract

An approach is described for rapidly determining protein structures by NMR that utilizes proteins containing ^{13}C-methyl labeled Val, Leu, and Ile (δ1) and protonated Phe and Tyr in a deuterated background. Using this strategy, the key NOEs that define the hydrophobic core and overall fold of the protein are easily obtained. NMR data are acquired using cryogenic probe technology which markedly reduces the spectrometer time needed for data acquisition. The approach is demonstrated by determining the overall fold of the antiapoptotic protein, Bcl-xL, from data collected in only 4 days. Refinement of the Bcl-xL structure to a backbone rmsd of 0.95 Å was accomplished with data collected in an additional 3 days. A distance analysis of 180 different proteins and structure calculations using simulated data suggests that our method will allow the global folds of a wide variety of proteins to be determined.

Introduction

As a result of massive, high-throughput sequencing, several microbial genomes have been sequenced and soon the sequences of all human genes will be known. This information is having a dramatic impact on biological research. By comparing the new amino acid sequences to those of proteins with known function, the function of the new protein can often be inferred. However, there are still many cases in which the function of a newly discovered protein cannot be assigned from its sequence alone. In these cases, the three-dimensional structure of the protein may be useful for characterizing its function by comparing the structure to proteins with known biological activities (Muchmore et al., 1996; Zarembinski et al., 1998; Burley et al., 1999). Protein structures are also important for identifying active sites and guiding site-directed mutagenesis to test functional predictions. In addition, by determining the structures of novel proteins, new protein folds will be discovered that will improve homology modeling and provide more data for the development of structure prediction algorithms. Indeed, three-dimensional structures of proteins that have been determined thus far have been extremely useful in many ways.

Although the number of new protein structures that have been deposited in the Protein Data Bank has been steadily increasing, the number of experimentally determined structures is still relatively small compared to the number of protein sequences. This disparity is principally due to the difficulties in the overall process of determining protein structures by X-ray crystallography or NMR spectroscopy. However, both approaches are rapidly improving. Through the use of a synchrotron and MAD phasing techniques (Hendrickson et al., 1990; Ogata, 1998), the time required for obtaining X-ray crystal structures has been markedly reduced. Providing that suitable crystals can be obtained, it is now possible to determine the crystal structure of a new protein in less than a day. Recent advances in NMR hardware and new labeling techniques have also opened up the possibility of dramatically decreasing the time necessary to obtain protein structures by NMR. This will be especially important for those proteins that do not crystallize.

*To whom correspondence should be addressed. E-mail: stephen.fesik@abbott.com

Table 1. NMR experiments and parameters used in the collection of the NMR data

	Sample	dim	NS	Time (h:min)	Number of increments (t1×t2×t3)	Dwell (μs) (t1×t2×t3)	Field (MHz)	Carrier frequency (ppm)
^{15}N HSQC	a	2D	2	0:09	128(N)×512(H)	720×120	500	116 (N); 4.74 (H)
HNCO	a	3D	2	5:54	48(CO)×48(N)×512(H)	600×720×120	500	175 (CO); 116 (N); 4.74 (H)
HN(CA)CO	a	3D	2	5:34	48(CO)×48(N)×512(H)	600×720×120	500	175 (CO); 116 (N); 4.74 (H)
ct-HNCA	a	3D	2	8:56	72(CA)×48(N)×512(H)	260×720×120	500	57.2(CA); 116 (N); 4.74 (H)
HN(CO)CA	a	3D	2	4:16	36(CA)×48(N)×512(H)	260×720×120	500	57.2 (CA); 116 (N); 4.74 (H)
HNCACB	a	3D	2	6:44	56(CB)×48(N)×512(H)	132×720×120	500	44 (CB); 116 (N); 4.74 (H)
HN(CO)CACB	a	3D	2	6:04	48(CB)×48(N)×512(H)	132×720×120	500	44 (CB); 116 (N); 4.74 (H)
(H)C(CO)NH-TOCSY	a	3D	2	3:47	32(C)×48 (N)×512(H)	331.4×720×120	500	22.2 (C); 116 (N); 4.74 (H)
H(CCO)NH-TOCSY	a	3D	2	3:38	32(H)×48 (N)×512(H)	720×720×120	500	0.25 (H); 116 (N); 4.74 (H)
TOCSY (aliphatic)	b	2D	16	1:14	128(H)×1024(H)	250×100	800	1.74 (H); 4.74 (H)
TOCSY (aromatic)	b	2D	16	1:14	128(H)×1024(H)	400×100	800	7.38 (H); 4.74 (H)
TOCSY (aromatic)	c	2D	16	1:14	128(H)×1024(H)	400×100	800	7.38 (H); 4.74 (H)
^{15}N HSQC-NOESY	b	3D	4	10:09	64 (H)× 32(N)×512(H)	100×400×100	800	4.74 (H); 116 (N); 4.74 (H)
NOESY	b	2D	16	4:08	320(H)×1024(H)	100×100	800	4.74 (H); 4.74 (H)
^{15}N filtered-NOESY	b	2D	8	2:04	320(H)×1024(H)	100×100	800	4.74 (H); 4.74 (H)
^{15}N filtered-NOESY	c	2D	8	2:05	320(H)×1024(H)	100×100	800	4.74 (H); 4.74 (H)
^{13}C HMQC-NOESY ($\tau_m = 300$ ms)	b	3D	8	27:15	48(C)×48(H)×512(H)	220×400×100	800	22.2 (C); 0.54 (H); 4.74 (H)
Total time (d:h:min)				3:22:25				
HNCO ^{1}HN-^{15}N (TROSY, $\kappa = 0$)	a	3D	4	4:24*	24(CO)×36(N)×512(H)	600×720×120	500	175 (CO); 116 (N); 4.74 (H)
HNCO ^{1}HN-^{15}N (TROSY, $\kappa = 2$)	a	3D	4	4:28*	24(CO)×36(N)×512(H)	600×720×120	500	175 (CO); 116 (N); 4.74 (H)
HNCO^{13}C′-^{13}C$^{\alpha}$ (TROSY)	a	3D	4	11:53*	72(CO)×32(N)×512(H)	600×720×120	500	175 (CO); 116 (N); 4.74 (H)
HNCO ^{15}N-^{13}C′,^{1}HN-^{13}C′ (TROSY, $\kappa = 3$)	a	3D	4	8:06*	24(CO)×64(N)×512(H)	600×720×120	500	175 (CO); 116 (N); 4.74 (H)
HNCO^{1}HN-^{15}N	a	3D	2	1:57*	24(CO)×32(N)×512(H)	600×720×120	500	175 (CO); 116 (N); 4.74 (H)
HNCO ^{13}C′-^{13}C$^{\alpha}$	a	3D	2	5:56*	72(CO)×32(N)×512(H)	600×720×120	500	175 (CO); 116 (N); 4.74 (H)
Total time (d:h:min)				6:23:53				

(a) ^{15}N-, ^{13}C-, ^{2}H-labeled sample of Bcl-xL with the methyl groups of Val, Leu, and Ile (δ1) protonated.
(b) ^{15}N-, ^{2}H-labeled Bcl-xL with ^{13}C-,^{1}H-methyl labeled Val, Leu, Ile, and protonated Phe and Tyr.
(c) ^{15}N-, ^{2}H-labeled Bcl-xL with ^{13}C-,^{1}H-methyl labeled Val, Leu, Ile, and protonated Tyr.
*The actual experimental times were twice as long since reference and phage experiments had to be acquired. This is reflected in the total time.

One approach for decreasing the amount of time needed to determine NMR structures of proteins is to reduce the number of restraints used in the structure calculations. If only the restraints that are easy to obtain and important for protein folding are used, then the data collection and analysis time could be reduced. Using ^{15}N-, ^{13}C-, ^{2}H-labeled proteins with Val, Leu, Ile (δ1) methyl protonation and modern multidimensional NMR techniques, Kay and co-workers have shown that the backbone and methyl groups can be readily assigned (Yamazaki et al., 1994; Shan et al., 1996). Furthermore, NOEs involving the limited set of protons in these deuterated proteins can be rapidly interpreted in terms of a set of methyl-methyl, methyl-NH, and NH-NH distance restraints (Gardner et al., 1997). Using only these NOE-derived restraints and ϕ, ψ angular restraints based on the secondary structure predicted from the backbone chemical shifts, the global folds of proteins can be obtained as previously illustrated for the C-terminal SH2 domain of bovine phospholipase Cδ1 (Gardner et al., 1997). However, based on structure calculations using simulated restraints, Gardner and co-workers concluded that the quality of the protein structures determined using their approach will be highly dependent on the secondary structure and topology. In particular, it was shown that accurate structures will be especially difficult to obtain for highly helical proteins (Gardner et al., 1997).

Here we describe an approach for rapidly determining the structures of proteins by NMR spectroscopy. The method is demonstrated using the all α-helical, antiapoptotic protein, Bcl-xL (Muchmore et al., 1996). Our strategy is an extension of the method developed by Gardner et al. (1997) in which we include additional, easy-to-obtain restraints to the structure calculations. These new distance restraints are derived from aromatic-methyl, aromatic-aromatic, and aromatic-NH NOEs obtained using ^{15}N-, ^{2}H-labeled proteins containing ^{13}C-labeled and protonated Val, Leu, and Ile (δ1) methyl groups and protonated Phe and Tyr. By including these additional NOE-derived distance restraints, a substantial improvement in the precision and accuracy of protein structures can be obtained as previously demonstrated by Clore and co-workers for the barrier-to-autointegration factor BAF and cyanovirin-N (Clore et al., 1999). In addition, like previously described approaches that utilize deuterated proteins (Torchia et al., 1988; LeMaster, 1989; Grzesiek et al., 1995; Venters et al., 1995; Gardner et al., 1997), this approach will be applicable for determining the global folds of larger proteins. To reduce the time required for data collection, cryogenic NMR probe technology was used for the acquisition of the NMR data in which the preamplifier and radio frequency coils of the probe are cooled to low temperatures (Styles et al., 1984; Black et al., 1993). Using our strategy for protein structure determination and cryogenic probe technology, protein structures can be determined by NMR in a high-throughput mode which may allow NMR to contribute to the structural genomics initiative (Terwilliger et al., 1998; Burley et al., 1999).

Materials and methods

Sample preparation

Bcl-xL used in the NMR studies lacks the putative C-terminal transmembrane region and residues 45–84 which constitute a flexible loop in the full length protein. The protein was cloned and expressed with a C-terminal six histidine tag as previously described (Sattler et al., 1997) using the T7 system and purified by affinity chromatography on a nickel-IDA column (Invitrogen). The amino acid precursors [3-^{13}C]α-ketobutyrate and [3,3′-^{13}C]α-ketoisovalerate were prepared as previously reported (Hajduk et al., 2000) by single or double alkylation of the N,N-dimethylhydrazone of pyruvate t-butyl ester using ^{13}C-labeled methyl iodide followed by hydrolysis to the α-keto ester and removal of the t-butyl group. To prepare the ^{15}N-, ^{13}C-, ^{2}H-labeled sample of Bcl-xL in which the methyl groups of Val, Leu, and Ile (δ1) are protonated (Goto et al., 1999), bacterial cells (BL21-DE3) over-expressing Bcl-xL were grown at 37 °C on a miminal medium containing [U-^{13}C]α-ketobutyrate (50 mg/l), [U-^{13}C]α-ketoisovalerate (100 mg/l), [U-^{13}C, ^{2}H] glucose, (3 g/l), $^{15}NH_4Cl$ (1 g/l), and $^{2}H_2O$ obtained from Cambridge Isotopes. The amino acid precursors were added 30 min prior to induction with 1 mM IPTG. ^{15}N-, ^{2}H-labeled Bcl-xL containing ^{13}C-methyl labeled Val, Leu, Ile, and protonated Phe and Tyr was isolated from bacterial cells grown on a medium containing [3-^{13}C]-α-ketobutyrate (50 mg/l), [3,3′-^{13}C]-α-ketoisovalerate (100 mg/l), [U-^{13}C, ^{2}H] glucose (3 g/l), $^{15}NH_4Cl$ (1 g/l), $^{2}H_2O$, and ^{15}N-labeled Phe and Tyr. A sample was also prepared using the same medium except for the lack of ^{15}N-labeled Phe. In both of these samples, the β-proton of the Val residues and γ-protons of Leu and Ile are partially

protonated, since fully protonated α-ketoisovalerate and α-ketobutyrate were used in the growth medium. [U-^{13}C, ^{15}N, ^{2}H (75%)] Bcl-xL was prepared from cells grown on a minimal medium containing [U-^{13}C] glucose (3 g/l), $^{15}NH_4Cl$ (1 g/l), and 75% $^{2}H_2O$.

NMR spectroscopy

All NMR spectra were collected at 30 °C on either a Bruker DMX500 NMR spectrometer equipped with a cryoprobe^{TM} or a Bruker DRX800 NMR spectrometer. The parameters and spectrometers used to acquire the data for the individual NMR experiments are given in Table 1. The backbone (Yamazaki et al., 1994) and side chain TOCSY (Montelione et al., 1992; Logan et al., 1992) experiments were performed using the ^{15}N-, ^{13}C-, ^{2}H-labeled sample of Bcl-xL in which the methyl groups of Val, Leu, and Ile (δ1) are protonated; whereas the NOESY experiments utilized ^{15}N-, ^{2}H-labeled Bcl-xL containing ^{13}C-labeled and protonated Val, Leu, and Ile (δ1) methyl groups and ^{15}N-labeled Phe and Tyr that was fully protonated. A mixing time of 300 ms was chosen to yield the most NOEs. This relatively long mixing time is appropriate because of the reduction of spin diffusion due to the deuteration of the protein. Residual dipolar couplings were collected using the ^{15}N-, ^{13}C-, ^{2}H-labeled sample of Bcl-xL that contained the protonated Val, Leu, and Ile (δ1) methyl groups. Modified versions of TROSY-based HNCO pulse sequences were employed to obtain ^{1}HN-^{15}N, ^{13}C′-^{13}Cα, ^{15}N-^{13}C′, and ^{1}HN-^{13}C′ residual dipolar couplings (Yang et al., 1999). Data were collected in the absence (reference spectra) and presence of 20 mg/ml Pf1 phage (Hansen et al., 1998; Clore et al., 1998) on a Bruker DMX500 equipped with a cryoprobe. The ^{1}HN-^{15}N and ^{13}C′-^{13}Cα residual dipolar couplings were measured using two sets of HNCO experiments to resolve peaks in the amide proton dimension that could not be resolved in the nitrogen dimension in the TROSY-based experiment. The final values of ^{1}HN-^{15}N and ^{13}C′-^{13}Cα couplings used in the structure calculations were obtained as arithmetical averages from the TROSY and conventional HNCO experiments.

Structure calculations

Three-dimensional structures of Bcl-xL were generated with a distance geometry/simulated annealing protocol (Nilges et al., 1988) using the CNX program (MSI Inc., San Diego, CA). Distance restraints derived from the ^{13}C-HSQC-NOESY spectra acquired using a mixing time of 300 ms were categorized in 4 bins (4.0, 5.0, 6.0, and 7.0 Å) based on the cross-peak intensities. An additional 0.5 Å was added for NOEs involving the methyl groups. For the distance restraints, a square-well potential using a force constant of 50.0 kcal mol^{-1} Å^{-2} was employed. The torsional restraints derived from TALOS were used with a harmonic potential and a force constant of 200 kcal mol^{-1} rad^{-2}. The N-H, N-C′, Cα-C′, and H^N-C′ residual dipolar couplings were included in the structure refinement using a protocol that has been previously described (Tjandra et al., 1997).

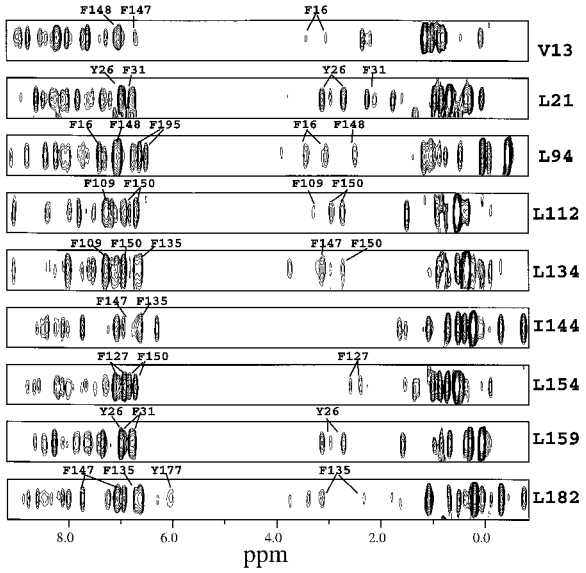

Figure 1. Selected regions of ^{1}H, ^{1}H planes from the ^{13}C-HMQC-NOESY spectrum extracted at the ^{13}C-methyl frequencies of the residues shown at the right. The long-range NOEs observed in the spectrum between aromatic and methyl groups are labeled.

Results and discussion

Strategy

In order to rapidly determine the global fold of a protein, an approach was required that utilized structural restraints derived from NOEs that could be obtained with high sensitivity, were easy to assign with a minimum number of experiments, and involved protons in the hydrophobic core of the protein. The method should be applicable to any protein, regardless of the secondary structure and topology and ideally to relatively large proteins as well. The approach of

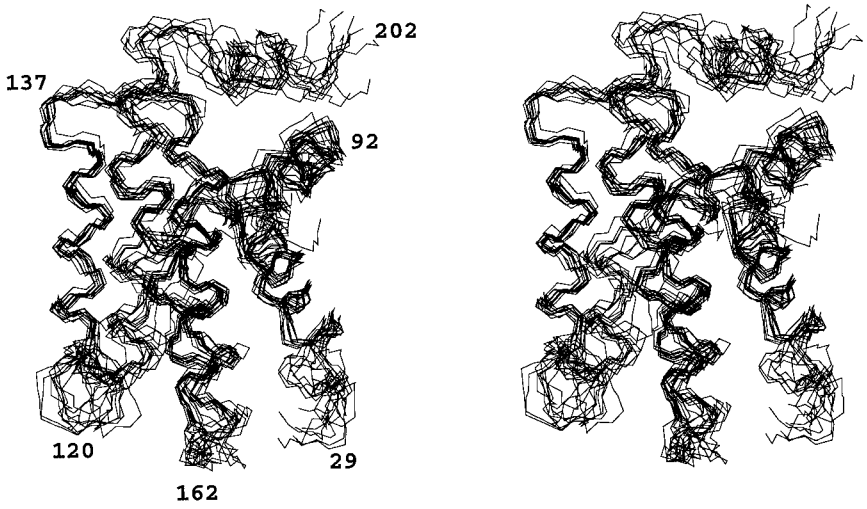

Figure 2. Stereoview of the ensemble of Bcl-xL structures using restraints derived from all of the NMR data.

Gardner et al. (1997) for obtaining global folds of highly deuterated, methyl-protonated proteins by multidimensional NMR satisfies most of these criteria. However, based on simulations, the approach does not appear to be generally applicable to all classes of proteins (Gardner et al., 1997). To extend this method to all proteins, additional structural restraints are needed involving the amino acids located in the interior of the protein. In addition to Val, Leu, and Ile, aromatic amino acids such as Phe are often found in the interior of proteins (Janin et al., 1988). To obtain NOEs involving aromatic amino acids, reverse labeling strategies have been previously employed (Vuister et al., 1994; Kuboniwa et al., 1995; Aghazadeh et al., 1998). In our case, protein samples were prepared that contained only protonated Phe, Tyr, and the ^{13}C-labeled methyl groups of Val, Leu, and Ile in an otherwise deuterated protein (except for the exchangeable protons). Our labeling strategy is very similar to that previously described by Aghazadeh et al. (1998) in which unlabeled Phe and Tyr were present in uniformly ^{13}C-,^{15}N-,^{2}H-labeled protein containing Val, Leu, and Ile (δ1) with protonated methyl groups. However, in our approach only the methyl groups of Val, Leu, and Ile (δ1) are ^{13}C-labeled. Thus, ^{13}C-^{13}C couplings are absent, resulting in narrow signals without the need for constant time experiments, the sensitivity is improved, and spectral windows can be reduced which decreases the data acquisition time. Another advantage of this labeling scheme is that it is less expensive compared to uniform ^{13}C-labeling due to the low cost of ^{13}C-labeled methyl iodide used in the synthesis of the amino acid precursors (Hajduk et al., 2000). Even though the use of deuteration is not necessary for small proteins, a significant improvement in sensitivity is observed for larger proteins, the spectra are simplified which facilitates the data analysis, and spin diffusion is reduced (LeMaster, 1989; Arrowsmith et al., 1990; Reisman et al., 1991; Brodin et al., 1989; Kelly et al., 1999).

A second, critical part of our approach for high-throughput protein structure determination is the marked reduction in data acquisition times due to the use of recently developed cryogenic probe technology. A threefold improvement in sensitivity can be achieved compared to a conventional probe which is equivalent to a ninefold reduction in acquisition time (Hajduk et al., 1999). This allowed the total number of NMR experiments needed for protein structure determination to be collected in 4–7 days depending on the desired resolution of the final structures.

Assignments

The H^N, ^{15}N, Cα, C$'$ signals of the backbone and Cβ resonances were assigned from six triple resonance experiments that correlate the amides to the Cα, C$'$, and Cβ signals of the same and the adjacent amino acid residue. These experiments were recorded in a total time of about 1.5 days (Table 1). Despite the short amount of time for data acquisition, nearly all of the expected signals appeared in the spectra, allowing the backbone assignments to be rapidly obtained.

The ^{1}H and ^{13}C resonances of the methyl groups were readily assigned from two 3D experiments that

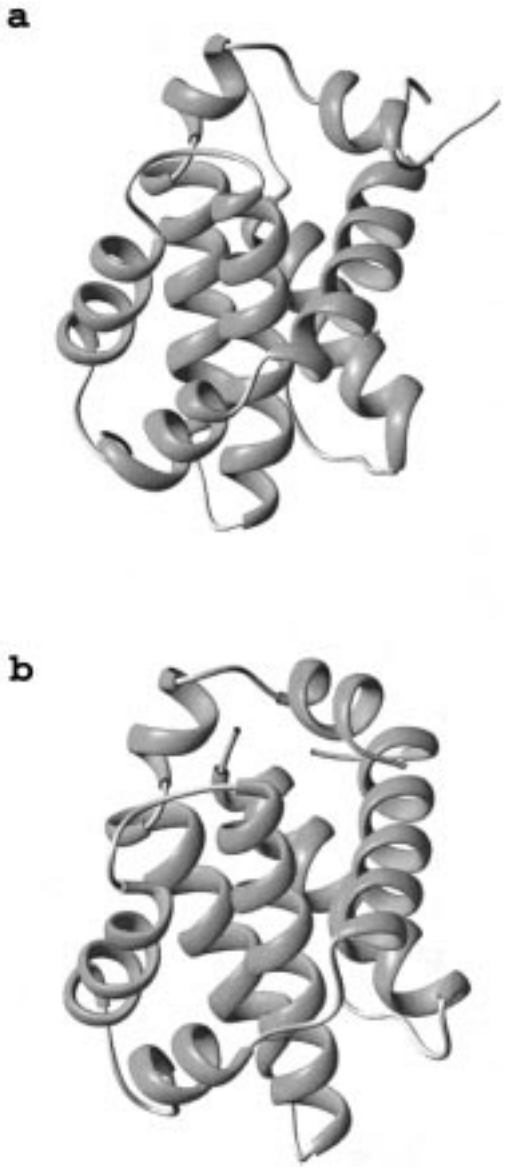

Figure 3. Ribbon (Carson et al., 1987) representation of (a) the NMR structure of Bcl-xL and (b) the X-ray structure of Bcl-xL (Muchmore et al., 1996).

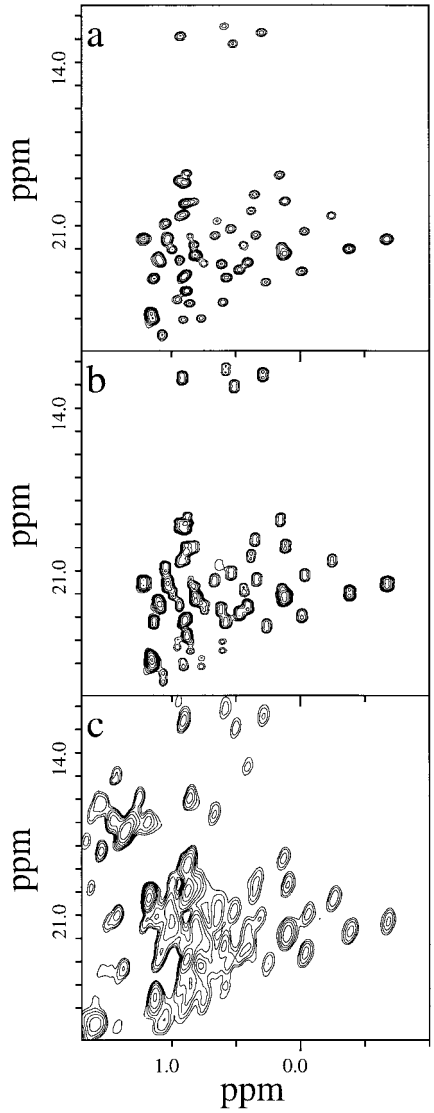

Figure 4. ^{1}H, ^{13}C HSQC spectra of (a) [U-^{15}N-, ^{2}H] Bcl-xL with ^{13}C-methyl labeled and protonated Leu, Val, and Ile (δ1), (b) [U-^{13}C-, ^{15}N-, ^{2}H] Bcl-xL in which the methyl groups of Leu, Val, and Ile (δ1) are protonated, and (c) [U-^{13}C-, ^{15}N-, ^{2}H (75%)] Bcl-xL.

correlate the methyl group signals to the ^{1}H and ^{15}N amide signals of the adjacent (i+1) residue. The protons and carbons of the methyl groups were identified from the correlations observed in the 3D ^{13}C-resolved NOESY experiment.

The proton assignments of Tyr and Phe were made by correlating the signals of the through-bond coupled protons in 2D homonuclear TOCSY spectra. By comparing TOCSY spectra of a protein sample containing Phe and Tyr to one containing Tyr, the amino acid spin systems of Phe and Tyr could be distinguished. The sequence specific assignments of Phe and Tyr was accomplished by correlating the amide signals to the β-protons and the β-protons to the aromatic signals in NOESY spectra (Table 1). The β-proton of the Val residues and the γ-protons of the Ile and Leu were also assigned from TOCSY and NOESY spectra.

Structural restraints

From the backbone chemical shifts, the secondary structure was obtained using the TALOS program (Cornilescu et al., 1999). On the basis of the secondary structure, ϕ, ψ angular restraints and distance restraints involving NH and CO groups of the α-helices were derived for those residues whose secondary structure could be unambiguously defined. Distance restraints were obtained from NH-NH, NH-CH_3, CH_3-CH_3, aromatic-CH_3, aromatic-NH, and aromatic-aromatic NOEs. In addition, a few distance restraints were obtained from NOEs involving the β-proton of the Val residues and γ-protons of the Ile and Leu. Several checks of the NOE assignments are performed during the structure refinement. Initially, a large uniform upper bound is placed on all NOEs. Any NOE that is consistently violated is carefully scrutinized by re-inspection of the NMR data. Furthermore, consistency checks are performed against the known secondary structure. NOEs that are not corroborated by restraints to the same or nearby residues are included only in later stages of the refinement (Meadows et al., 1994). As illustrated in Figure 1, several long-range NOEs between the methyl groups and aromatic residues were observed in the 3D ^{13}C-resolved NOESY spectra. A total of 970 NOEs were observed in the NOESY experiment. Many of these were long-range NOEs involving Phe and Tyr that are important for protein folding. To further improve the quality of the NMR structures, restraints derived from 224 residual dipolar couplings were included in the structure calculations.

Structure calculations

The initial structure calculations of Bcl-xL included a total of 1404 NMR-derived restraints. For a set of 10 structures, the rmsd about the mean coordinate positions for residues 5–30 and 89–200 of Bcl-xL (numbering of the full length protein) was 1.25 $\pm$ 0.12 Å for the backbone. As a measure of the accuracy, the NMR-derived structure was compared to the corresponding portion of the X-ray structure of Bcl-xL that contained the large loop between α1 and α2 (Muchmore et al., 1996). The rmsd to the X-ray crystal structure of Bcl-xL was 2.61 $\pm$ 0.12 Å. The total time to acquire the NMR data to obtain these structures was only 4 days. To obtain improved structures, additional restraints derived from residual dipolar couplings were added to the structure calculations. Using these additional restraints, the rmsd of the backbone decreased to 0.95 $\pm$ 0.11 Å (Figure 2). As shown in Figure 3, the NMR structure is similar to the X-ray structure of Bcl-xL with a backbone rmsd of 2.12 $\pm$ 0.09 Å (Table 2).

Comparison to other methods

An advantage of our approach is the use of selectively ^{13}C-labeled Val, Leu, and Ile (δ1) methyl groups. Due to the lack of the ^{13}C-^{13}C coupling, the signals are narrower (Figure 4). To obtain this resolution for the methyl groups using uniformly ^{13}C-labeled proteins (Gardner et al., 1997), constant time experiments are required which are less sensitive compared to conventional experiments.

Another advantage of our strategy compared to that of Gardner et al. (1997) is the use of additional NOEs involving Phe and Tyr. From an analysis of 180 NMR-derived protein structures, a significant number of NOEs involving Phe and Tyr would be expected (Figure 5). An average of 42% more NOEs could be included in the structure determinations, and as demonstrated for Bcl-xL, these additional restraints should produce more precise and more accurate NMR structures of these proteins. To test whether the additional NOE-derived restraints would generally improve the quality of structures, we generated NMR structures for three different classes of proteins whose structures were previously determined in our laboratory. The structures were determined using only a subset of the actual NMR restraints (Table 3). Significant improvements in both the accuracy and precision were observed by including the additional NOE-derived restraints involving Phe and Tyr in the structure calculations of the Dbl homology domain of Trio (Liu et al., 1998) and stromelysin (Table 3). Furthermore, while the additional NOEs did not improve the precision, the structures of the PTB domain of IRS-1 (Zhou et al., 1996) calculated with the extra restraints were more accurate (Table 3).

An alternative labeling strategy that has been utilized for obtaining protein structures by NMR is the use of fractionally deuterated ^{13}C-, ^{15}N-labeled proteins (Kushlan and LeMaster, 1993; Grzesiek et al., 1993). Using this protocol, the NMR samples are less expensive to prepare, *all* of the side chain signals are observed and may be assigned, and additional NOE-derived distance restraints involving the protons of the side chains can be included in the structure calculations. However, using fractionally deuterated proteins, the signals are less resolved (Figure 4), and

Table 2. Precision and accuracy of the NMR structures of Bcl-xL

NMR data[a]	# NOEs	Precision[b] (Å)	Accuracy[c] (Å)
NOEs (τ_m = 300 ms)	970	1.25 ± 0.12	2.61 ± 0.12
All NOEs + residual dipolar couplings	970 (+ 224 RDC)	0.95 ± 0.11	2.12 ± 0.09
NOEs - F, Y NOEs	557	1.79 ± 0.41	3.87 ± 0.42

[a]NMR data also includes ϕ, ψ restraints from TALOS.
[b]The rmsd for the backbone of the ensemble of structures compared to the mean.
[c]The rmsd for the backbone atoms of the ensemble compared to the X-ray structure of Bcl-xL (Muchmore et al., 1996).

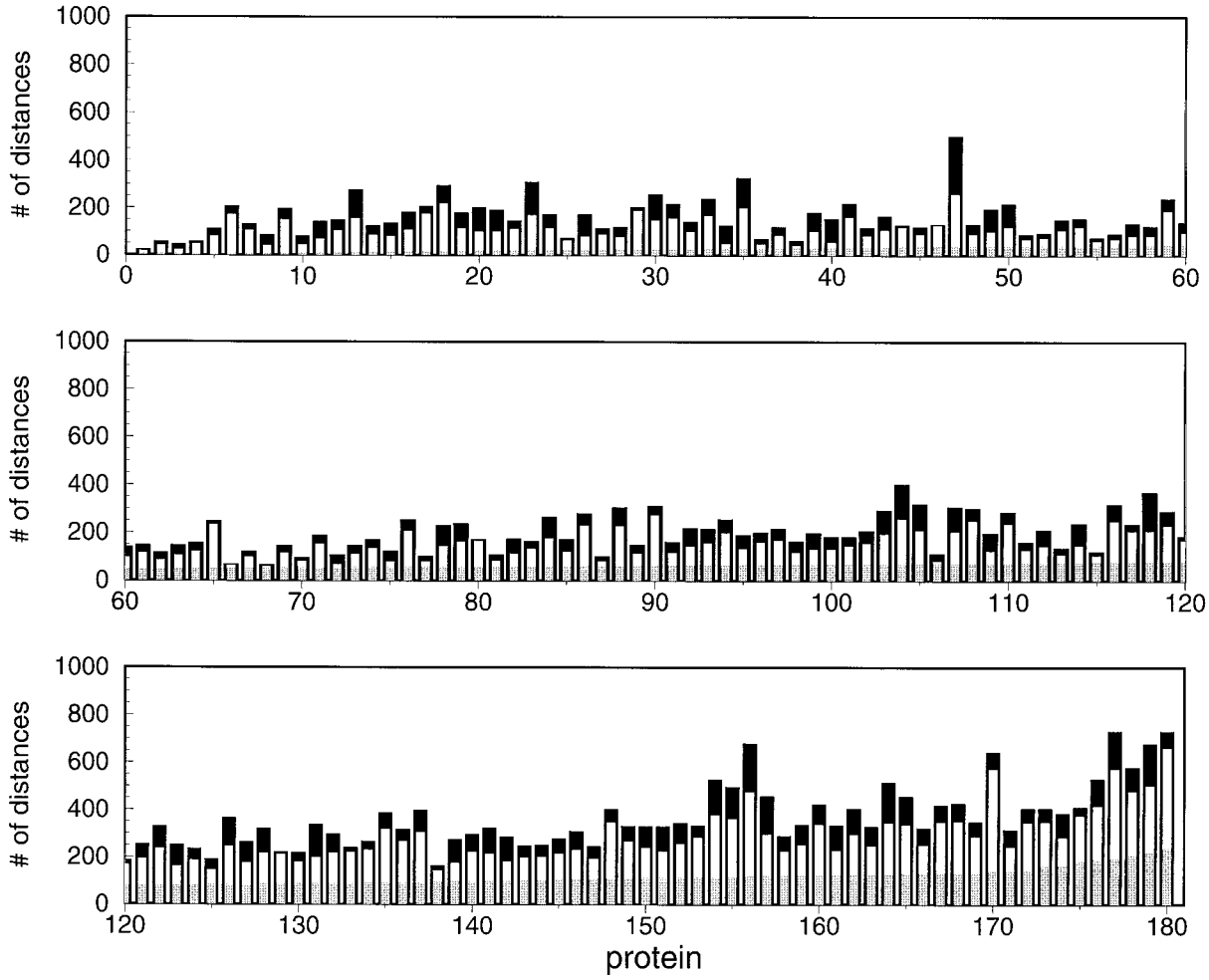

Figure 5. The number of proton-proton distances < 6 Å in 180 proteins. The bottom gray bars represent NH-NH distances, the middle white bars correspond to the number of protons close to Leu, Val, and Ile (δ1) methyl groups, and the top black bars represent short distances involving the aromatic protons of Tyr and Phe.

Table 3. Structure calculations of proteins of different classes using a subset of NOEs

Protein	Class	Restraints[a]	# NOEs	Precision (Å)	Accuracy (Å)
IRS-PTB	β	+F, Y NOEs	339	1.13 ± 0.09	2.36 ± 0.07
		−F, Y NOEs	261	1.13 ± 0.21	3.32 ± 0.24
Trio DH	α	+F, Y NOEs	416	1.76 ± 0.25	3.15 ± 0.43
		−F, Y NOEs	341	2.55 ± 0.50	5.83 ± 0.75
Stromelysin	α and β	+F, Y NOEs	357	1.76 ± 0.16	3.43 ± 0.30
		−F, Y NOEs	244	2.55 ± 0.57	4.45 ± 0.76

[a]The NMR-derived restraints used in the structure calculations included φ, ψ angular restraints based on the secondary structure and NOEs involving amide protons and the methyl groups of Val, Leu, and Ile (δ1).

the sensitivity of all the NMR experiments is markedly reduced. For example, the sensitivity of the constant time HNCA and HNCACB experiments is reduced by at least a factor of two which would require more than a fourfold increase in data acquisition time to obtain the same signal to noise. The difference in sensitivity was even more pronounced for the 3D HN(CA)CO experiment which was four times less sensitive when acquired on fractionally deuterated Bcl-xL compared with the perdeuterated protein.

Conclusions

An approach is described for the high-throughput structure determination of proteins by NMR spectroscopy that involves the collection of easy-to-assign and easy-to-obtain structural restraints using a cryoprobe for data acquisition. For Bcl-xL, the total time for data collection was only 4 days for obtaining the global fold of the protein and less than one week for obtaining a more resolved structure of Bcl-xL. Because the protein is deuterated, the spectra are simplified with a limited set of defined NOEs to analyze. This simplifies the manual interpretation of the NMR data which can be time-consuming for unknown proteins and should facilitate automated data analysis. Another advantage of using deuterated proteins is the narrow line widths of the NMR signals which should allow this approach to be applied to relatively large proteins. Indeed, based on simulations, our approach should be applicable to a wide variety of proteins from different structural classes.

References

Aghazadeh, B., Zhu, K., Kubiseski, T.J., Liu, G.A., Pawson, T., Zheng, Y. and Rosen, M.K. (1998) *Nat. Struct. Biol.*, **5**, 1098–1107.

Arrowsmith, C.H., Pachter, R., Altman, R.B., Iyer, S.B. and Jardetzky, O. (1990) *Biochemistry*, **29**, 6332–6341.

Black, R.D., Early, T.A., Roemer, P.B., Mueller, O.M., Mogro-Campero, A., Turner, L.G. and Johnson, G.A. (1993) *Science*, **259**, 793–795.

Brodin, P., Drakenberg, T., Thulin, E., Forsén, S. and Grundström, T. (1989) *Protein Eng.*, **2**, 353–358.

Burley, S.K., Almo, S.C., Bonanno, J.B., Capel, M., Chance, M.R., Gaasterland, T., Lin, D., Šali, A., Studier, F.W. and Swaminathan, S. (1999) *Nat. Genet.*, **23**, 151–157.

Carson, M.J. (1987) *J. Mol. Graphics*, **5**, 103–106.

Clore, G.M., Starich, M.R. and Gronenborn, A.M. (1998) *J. Am. Chem. Soc.*, **120**, 10571–10572.

Clore, G.M., Starich, M.R., Bewley, C.A., Cai, M. and Kuszewski, J. (1999) *J. Am. Chem. Soc.*, **121**, 6513–6514.

Cornilescu, G., Delaglio, F. and Bax, A. (1999) *J. Biomol. NMR*, **13**, 289–302.

Gardner, K.H., Rosen, M.K. and Kay, L.E. (1997) *Biochemistry*, **36**, 1389–1401.

Goto, N.K., Gardner, K.H., Mueller, G.A., Willis, R.C. and Kay, L.E. (1999) *J. Biomol. NMR*, **13**, 369–374.

Grzesiek, S., Anglister, J., Ren, H. and Bax, A. (1993) *J. Am. Chem. Soc.*, **115**, 4369–4370.

Grzesiek, S., Wingfield, P., Stahl, S., Kaufman, J.D. and Bax, A. (1995) *J. Am. Chem. Soc.*, **117**, 9594–9595.

Hajduk, P.J., Gerfin, T., Boehlen, J.-M., Häberli, M., Marek, D. and Fesik, S.W. (1999) *J. Med. Chem.*, **42**, 2315–2317.

Hajduk, P.J., Augeri, D.J., Mack, J., Mendoza, R., Yang, J., Betz, S. and Fesik, S.W. (2000) *J. Am. Chem. Soc.*, **122**, 7898–7904.

Hansen, M.R., Mueller, L. and Pardi, A. (1998) *Nat. Struct. Biol.*, **5**, 1065–1074.

Hendrickson, W.A., Horton, J.R. and LeMaster, D.M. (1990) *EMBO J.*, **9**, 1665–1672.

Janin, J., Miller, S. and Chothia, C. (1988) *J. Mol. Biol.*, **204**, 155–164.

Kelly, M.J.S., Krieger, C., Ball, L.J., Yu, Y., Richter, G., Schmieder, P., Bacher, A. and Oschkinat, H. (1999) *J. Biomol. NMR*, **14**, 79–83.

Kuboniwa, H., Tjandra, N., Grzesiek, S., Ren, H., Klee, C.B. and Bax, A. (1995) *Nat. Struct. Biol.*, **2**, 768–776.

Kushlan, D.M. and LeMaster, D.M. (1993) *J. Biomol. NMR*, **3**, 701–708.

LeMaster, D.M. (1989) *Methods Enzymol.*, **177**, 23–43.
Liu, X., Wang, H., Eberstadt, M., Schnuchel, A., Olejniczak, E.T., Meadows, R.P., Schkeryantz, J.M., Janowick, D.A., Harlan, J.E., Harris, E.A.S., Staunton, D.E. and Fesik, S.W. (1998) *Cell*, **95**, 269–277.
Logan, T.M., Olejniczak, E.T., Xu, R.X. and Fesik, S.W. (1992) *FEBS Lett.*, **314**, 413–418.
Meadows, R.P., Olejniczak, E.T. and Fesik, S.W. (1994) *J. Biomol. NMR*, **4**, 79–96.
Montelione, G.T., Lyons, B.A., Emerson, S.D. and Tashiro, M. (1992) *J. Am. Chem. Soc.*, **114**, 10974–10975.
Muchmore, S.W., Sattler, M., Liang, H., Meadows, R.P., Harlan, J.E., Yoon, H.S., Nettesheim, D., Chang, B.S., Thompson, C.B., Wong, S.-L., Ng, S.-C. and Fesik, S.W. (1996) *Nature*, **381**, 335–341.
Nilges, M., Gronenborn, A.M., Brünger, A.T. and Clore, G.M. (1988) *Protein Eng.*, **2**, 27–38.
Ogata, C.M. (1998) *Nat. Struct. Biol.*, **5** (Suppl S), 638–640.
Reisman, J., Jariel-Encontre, I., Hsu, V.L., Parello, J., Geiduschek, E.P. and Kearns, D.R. (1991) *J. Am. Chem. Soc.*, **113**, 2787–2789.
Sattler, M., Liang, H., Nettesheim, D., Meadows, R.P., Harlan, J.E., Eberstadt, M., Yoon, H.S., Shuker, S.B., Chang, B.S., Minn, A.J., Thompson, C.B. and Fesik, S.W. (1997) *Science*, **275**, 983–986.
Shan, X., Gardner, K.H., Muhandiram, D.R., Rao, N.S., Arrowsmith, C.H. and Kay, L.E. (1996) *J. Am. Chem. Soc.*, **118**, 6570–6579.
Styles, P., Soffe, N.F., Scott, C.A., Cragg, D.A., Row, F., White, D.J. and White, P.C.J. (1984) *J. Magn. Reson.*, **60**, 397–404.
Terwilliger, T.C., Waldo, G., Peat, T.S., Newman, J.M., Chu, K. and Berendzen, J. (1998) *Protein Sci.*, **7**, 1851–1856.
Tjandra, N., Omichinski, J.G., Gronenborn, A.M., Clore, G.M. and Bax, A. (1997) *Nat. Struct. Biol.*, **4**, 732–738.
Torchia, D.A., Sparks, S.W. and Bax, A. (1988) *J. Am. Chem. Soc.*, **110**, 2320–2321.
Venters, R.A., Metzler, W.J., Spicer, L.D., Mueller, L. and Farmer II, B.T. (1995) *J. Am. Chem. Soc.*, **117**, 9592–9593.
Vuister, G.W., Kim, S.J., Wu, C. and Bax, A. (1994) *J. Am. Chem. Soc.*, **116**, 9206–9210.
Yamazaki, T., Lee, W., Arrowsmith, C.H., Muhandiram, D.R. and Kay, L.E. (1994) *J. Am. Chem. Soc.*, **116**, 11655–11666.
Yang, D., Venters, R.A., Mueller, G.A., Choy, W.Y. and Kay, L.E. (1999) *J. Biomol. NMR*, **14**, 333–343.
Zarembinski, T.I., Hung, L.-W., Mueller-Dieckmann, H.-J., Kim, K.-K., Yokota, H., Kim, R. and Kim, S.-H. (1998) *Proc. Natl. Acad. Sci. USA*, **95**, 15189–15193.
Zhou, M.M., Huang, B., Olejniczak, E.T., Meadows, R.P., Shuker, S.B., Miyazaki, M., Trüb, T., Shoelson, S.E. and Fesik, S.W. (1996) *Nat. Struct. Biol.*, **3**, 388–393.