Protein Engineering vol.4 no 1 pp.3–10, 1990

# Solution structure of the DNA-binding domain of the yeast transcriptional activator protein GCN4

**Vladimir Saudek, Annalisa Pastore[1,2], Maria A.Castiglione Morelli[3], Rainer Frank[2], Heinrich Gausepohl[2], Toby Gibson[2], Falk Weih[4] and Paul Roesch[5]**

Merrell Dow Research Institute, Strasbourg, France, [2]EMBL, Meyerhofstr. 1, Postfach 1022 09, D-6900 Heidelberg, FRG, [3]Universitá di Napoli, Napoli, Italy, [4]German Cancer Research Center, Heidelberg and [5]Max Planck Institut für Medizinische Forschung, Heidelberg, FRG

[1]To whom correspondence should be addressed

**The solution structure of an active synthetic peptide containing both the leucine zipper and the adjacent basic domain of the yeast transcription factor GCN4 (residues 220–280) was determined by NMR. The two domains show structurally distinct behaviours. In the absence of DNA, the basic domain is, although very flexible, structured and fluctuating around a helical conformation. The leucine zipper region forms a long, uninterrupted helix. From a suitable set of NMR distances the three-dimensional structure of the leucine zipper monomeric sub-domain was calculated by distance geometry algorithms. The structure of the symmetrical parallel dimer was obtained by model building using the NMR information. A smaller peptide with the sequence of the isolated basic region (residues 1–35 of the 61 residue peptide) was also synthesized. Circular dichroism studies showed 30–40% helicity. A flexible helix spans the region between residues 8 and 21. The comparison of our results with suggested models is discussed in detail.**

*Key words*: coiled-coil/DNA binding proteins/GCN4 protein/leucine zipper/NMR structure

## Introduction

A 60–70 residue DNA-binding domain occurs in some 25 known proteins defined as transcriptional regulators, some of which also show oncogenic potential (Landshutz *et al.*, 1988; Vinson *et al.*, 1989). To bind specifically to a recognition site in DNA, the domain must dimerize, and the best characterized binding sites of these proteins are palindromic (Hope and Struhl, 1987; Mueller *et al.*, 1989; Risse *et al.*, 1989).

The complete domain is characterized by two distinct regions. The C-terminal region, termed 'leucine zipper', contains a repeat of leucine every seven residues. It dimerizes and positions the contiguous region to interact specifically with DNA. Landshutz *et al.* (1988) proposed that dimerization occurred through interdigitating leucines in paired anti-parallel α-helices. This model was revised by O'Shea *et al.* (1989a), who demonstrated that a synthetic peptide encompassing the zipper of the GCN4 protein dimerized with the N-termini in close proximity. From a recent NMR study, this sub-domain was shown to be helical and inferred to be arranged as a parallel dimer (Oas *et al.*, 1990).

The N-terminal 'basic region' of the domain contains several positively charged amino acid residues. It shows strong helical propensity but lacks the hydrophobic residues which could allow hydrophobic packing. In the most recent proposed structure for the domain, Vinson *et al.* (1989) suggested a model christened the 'scissors grip'. According to this model, the basic regions each form two consecutive helices which track around the major groove of DNA, being held in place by the dimerized zipper. So far, no direct structural information about the basic region has been presented.

Any analysis of the role of the zipper proteins in the regulation of gene expression is complicated by the fact that in many cases they form heterodimers (Kouzarides and Ziff, 1989; Ransone *et al.*, 1989; Schuermann *et al.*, 1989; Turner and Tjian, 1989). Because GCN4 has been shown to form stable homodimers *via* its leucine zipper (Hinnebusch, 1985; Hope and Struhl, 1987; Sellers and Struhl, 1989), it appeared to be an excellent candidate for a high resolution structural analysis of the DNA-binding domain, which would be relevant to all members of the family.

In this paper we report a high resolution structure determination by NMR of a 61 residue synthetic peptide containing both the basic region and the leucine zipper of the yeast transcriptional activator GCN4 (BR-LZ peptide). We calculated the three-dimensional structure of the leucine zipper sub-domain of BR-LZ from NMR restraints and distance geometry calculations. A smaller fragment containing the basic region (BR peptide) with the amino acid sequence of the first N-terminal 35 residues was also studied. In this fragment, peak overlap was reduced and the results could be compared with the 'scissors grip' model.

## Materials and methods

The BR-LZ peptide amino acid sequence (residues 221–281 of the GCN4 protein) was:

```
        10        20        30        40        50        60
VPESSDPAALKRARNTEAARRSRARKLQRMKQLEDKVEELLSKNYHLENEVARLKKLVGER
```

The BR fragment includes the first 35 N-terminal residues.

*Peptide synthesis*

The peptides were synthesized on 1% cross-linked polystyrene resin by automated peptide synthesis under continuous flow conditions (Frank and Gausepohl, 1988). The raw products were purified to homogeneity by preparative HPLC and the identity of the peptide assessed by Edman degradation. To allow renaturation before use in experimentation, the BR-LZ peptide was dissolved in water and the sample placed in a beaker of water at 348 K (75°C) under nitrogen. The beaker was then left to cool down to room temperature. A similar treatment was also applied to the BR peptide.

*DNA-binding assays*

The oligonucleotides were synthesized on an Applied Biosystems Synthesizer (model 380B) using proprietary chemistry. We used Bluescribe vectors and confirmed the results by sequencing. The oligonucleotide sequences were:

GCN4$^{pal}$ 5′-AATTCCACCTAGCGGATGACTCATTTTTTTTCTTAGCGA-3′
GCN4$^{mut}$ 5′-AATTCCACCTAGCGGAAGGCTCATTTTTTTTCTTAGCGA-3′

Gel retardation experiments were carried out as described by Hope and Struhl (1987). Increasing amounts of the peptide were incubated with either 10 fmol of *Eco*RI−*Hin*dIII cleaved, end-labelled, palindromic binding site ($GCN4^{pal}$; Figure 1, lanes 1−4) or with the mutated version ($GCN4^{mut}$; Figure 1, lanes 5−8). Complex formation is strongly enhanced by increasing the peptide concentration. No complex formation could be observed under the same conditions with the mutated binding site ($GCN_4^{mut}$).

### *CD measurements*

The CD spectra were recorded on a JASCO J-500A spectrophotometer using a 1 mm pathlength cell.

### *NMR samples*

The NMR samples were dissolved in 90% $H_2O$/10% $D_2O$ or $D_2O$ containing 0.05 mol/l acetic acid. The concentration was varied between 2 and 8 mmol/l. Over this range the NMR spectra did not change significantly. The pH was adjusted with NaOH from 3.0 to 6.2, the temperature was varied between 280 and 320 K. Conditions for optimal resolution were 310 K and pH 3.2 for BR-LZ and 300 K pH 3.5 for BR. All NMR experiments were performed on a Bruker AM500 spectrometer. HOHAHA experiments were carried out using mixing times of between 10 and 120 ms. The ROESY experiments used radio-frequency field strengths of between 5 and 7 kHz with mixing times of 50−500 ms.

### *Distance geometry calculations on BR-LZ*

For some of the residues, Φ angles could be estimated from the coupling constants between HN and Hα resonances. Inter-hydrogen distances were obtained from the NOEs measured with mixing times in the range of 100−400 ms. Hydrogen bonds were inferred from the temperature dependence of the HN resonances and their stability towards exchange with $H_2O$ at pH values up to 6.2. The NOE cross-peaks were classified as strong, medium and weak, and upper bounds were assigned as 3.0, 3.5 and 4.0 Å for backbone-backbone NOEs and 3.0, 3.5 and 4.5 Å for backbone−side chain and side chain−side chain NOEs. Hydrogen bonds were not explicitly included in the list of restraints for the distance geometry calculations. A standard protocol was used for DISMAN, where the initial structures were either generated completely at random or, in the final calculations, within specified regions of the conformational space (α-helical region ±30°). Five hundred cycles of steepest descent minimization of the target function were then performed. The

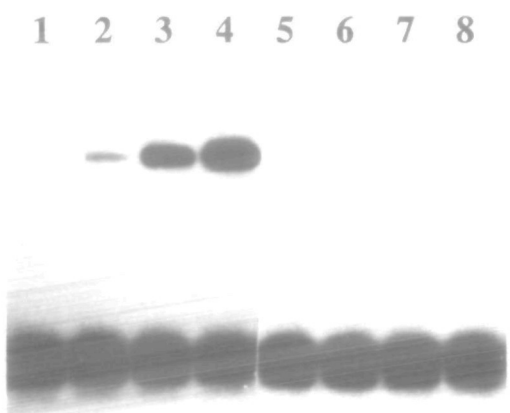


**Fig. 1.** Gel retardation assays were used to demonstrate DNA-binding properties of the BR-LZ peptide. **Lanes 1−4** show the binding to a DNA fragment containing the GCN4 binding site at increasing concentrations of the peptide. The specificity was proved by using a double-point mutated DNA sequence. **Lanes 5−8** show that no complex formation is observed with this fragment

**Table I.** Chemical shift assignment of the H resonances of the zipper peptide

| | NH | αH | βH | γH | δH | εH | Others |
|---|---|---|---|---|---|---|---|
| 1 Val t | n a | 4 20 | 2 33 | 1 13 1 03 | | | |
| 1 Val c | n a | 4 11 | 2 52 | 0.97 | | | |
| 2 Pro t | | 4 52 | 2 33 2.00 | 2 08 1 98 | 3 77 3 63 | | |
| 2 Pro c | | 4 42 | 2 47 2 08 | 2 18 | 3 41 3 45 | | |
| 3 Glu | 8 47 | 4.35 | 2 15 2 03 | 2 53 | | | |
| 4 Ser | 8 29 | 4 45 | 3 85 3 83 | | | | |
| 5 Ser | 8 26 | 4 48 | 3.93 3 85 | | | | |
| 6 Asp t | 8 25 | 4 93 | 2 92 2 77 | | | | |
| 6 Asp c | n a | 4 72 | 3 07 2.63 | | | | |
| 7 Pro t | | 4 33 | 2 37 2 03 | 2 08 | 3 87 | | |
| 7 Pro c | | n a | 2 43 1 82 | 2 22 | 3 58 3 68 | | |
| 8 Ala | 8 15 | 4 20 | 1 42 | | | | |
| 9 Ala | 7 83 | 4 18 | 1 43 | | | | |
| 10 Leu | 7 77 | 4.23 | 1 71 | 1 63 | 0 90 0 85 | | |
| 11 Lys | 7 82 | 4 10 | 1 82 1 65 | 1.41 | 1 48 | 2 99 | 7 50 |
| 12 Arg | 7 84 | 4 25 | 1 85 1 62 | 1 63 | 3 22 | | 7 18 |
| 13 Ala | 8 06 | 4 17 | 1 46 | | | | |
| 14 Arg | 8 18 | 4 25 | 1 83 1 62 | 1 64 | 3 21 | | 7 17 |
| 15 Asn | 8 34 | 4 75 | 2 92 2.88 | | | | 7 42 6 79 |
| 16 Thr | 8 13 | 4 37 | 4 25 | 1 27 | | | |
| 17 Glu | 8.29 | 4 22 | 2 11 | 2 50 | | | |
| 18 Ala | 8 17 | 4 17 | 1 41 | | | | |
| 19 Ala | 8 02 | 4 18 | 1 46 | | | | |
| 20 Arg | 8 07 | 4 22 | 1 80 1 78 | 1 65 | 3 22 | | 7 30 |
| 21 Arg | 8 18 | 4 25 | 1 83 1 74 | 1 64 | 3 22 | | 7 30 |
| 22 Ser | 8 14 | 4 35 | 3.97 3 89 | | | | |
| 23 Arg | 8 18 | 4 25 | 1 83 1 74 | 1 64 | 3 21 | | 7 30 |
| 24 Ala | 8 08 | 4 18 | 1 42 | | | | |
| 25 Arg | 8 29 | 4 45 | 1 84 1 78 | 1 66 | 3 21 | | 7 30 |
| 26 Lys | 8 24 | 4 17 | 1 81 1.77 | 1 46 | 1 66 | 3 00 | 7 40 |
| 27 Leu | 8 11 | 4 40 | 1 71 1.60 | 1 23 | 0 90 0 85 | | |
| 28 Gln | 7 85 | 4 43 | 2.52 | 2 17 2 98 | | | 7 42 6 78 |
| 29 Arg | 7 98 | 4 23 | 1 75 1 80 | 1.68 | 3.24 | 7 30 | 6 67 |
| 30 Met | 8.37 | 4.05 | 2 35 2.36 | 2 58 2 53 | 2 07 | | |
| 31 Lys | n a | 4 07 | 1 98 | 1 55 | 1 69 | 3 03 | |
| 32 Gln | 7 98 | 4 11 | 2 15 2 28 | 2 54 2 42 | | | 7 42 6 78 |
| 33 Leu | 8 22 | 4.18 | 1 82 1 77 | 1 40 | | | |
| 34 Glu | 8 42 | 3 96 | 2 28 2 12 | 2 64 2 43 | | | |
| 35 Asp | 8 46 | 4 45 | 3 03 2 88 | | | | |
| 36 Lys | 8 05 | 4 18 | 2 07 | 1 57 | 1 59 | 2 95 | |
| 37 Val | 8.37 | 3 43 | 2.30 | 1 05 0.88 | | | |
| 38 Glu | 7 84 | 4 04 | 2 30 2 25 | 2 67 2 60 | | | |
| 39 Glu | 8 25 | 4 17 | 2 32 2 24 | 2 68 | | | |
| 40 Leu | 8 68 | 4 03 | 2 12 1 89 | 1 27 | 0 93 0 83 | | |
| 41 Leu | 8 88 | 4 03 | 1 93 1 77 | 1 49 | 0.91 0 83 | | |
| 42 Ser | 7 72 | 4 35 | 4 09 | | | | |
| 43 Lys | 8 28 | 4 20 | 1 85 1 95 | 1.70 | 1 47 | 3 00 | |
| 44 Asn | 8.79 | n a | 2 72 3 24 | n a | | | 7 49 6 30 |
| 45 Tyr | 8.32 | 4 37 | 3.18 3 28 | | | | 7 08 6 83 |
| 46 His | 8 03 | 4.38 | 3 47 | | | | 7 38 8 67 |
| 47 Leu | 8 73 | 4 06 | 2 18 1.91 | 1 85 | 0 99 0 89 | | |
| 48 Glu | 8.82 | 3 97 | 2.27 2.10 | 2 50 | | | |
| 49 Asn | 7 78 | 4 45 | 2.63 2 82 | 6 16 7 28 | | | |
| 50 Glu | 8.03 | 4.17 | 2 20 | 2.45 | | | |
| 51 Val | 8.59 | 3 41 | 2.15 | 1 01 0.84 | | | |
| 52 Ala | 7.75 | 4.02 | 1 50 | | | | |
| 53 Arg | 7.80 | 4 03 | 1.95 | 1 77 1 55 | 3 25 3 19 | 7 33 | 6.67 |
| 54 Leu | 8.43 | 4.04 | 1 80 | 1 32 | 1.01 0.83 | | |
| 55 Lys | 8.79 | 3.84 | 1 82 1.76 | 162 | 1.60 | 3 91 2 83 | |
| 56 Lys | 7.32 | 4.12 | 1 98 1 94 | 1 60 | 1 81 | 3 00 | |
| 57 Leu | 7.61 | 4 21 | 1.98 1 82 | 1.63 | 0 94 0.90 | | |
| 58 Val | 7.88 | 3.94 | 2 18 | 1 02 0 94 | | | |
| 59 Gly | 7.85 | 4 10 3 92 | | | | | |
| 60 Glu | 8 31 | 4 22 | 2 15 2.00 | 2 57 | | | |
| 61 Arg | 8 07 | 4 22 | 1 92 1 77 | 1.68 | 3 22 | 7 19 | 6 67 |

n a not assigned
Values refer to DSS

range of the distance and the van der Waals constraints included in the target function were changed in a step of one residue up to level five. Steps of 10 were used afterwards, since no distances involving residues more than four residues apart were observed.

For the model of the dimeric structure, the $\phi$ and $\psi$ values were taken from Parry and Suzuki (1969) and McLachlan (1978) whereas the $\chi$ angles were chosen to be consistent to McGregor *et al.* (1987) and with the NMR data. The interface leucine side chains are in the $\chi_1 = -60$, $\chi_2 = 180$ conformation. Molecular graphics was carried out using a program of Lesk and Hardman (1982) and the INSIGHT (Dayringer *et al.*, 1986) graphics programs running on a PS390 system. The list of distance restrains used in the structure determination is available from the authors.

## Results

### *DNA-binding properties of the BR-LZ peptide*

The BR-LZ peptide was tested for its ability to bind sequence specifically to DNA using gel retardation and competition assays (Hope and Struhl, 1987). The peptide binds to a DNA fragment containing the palindromic GCN4 binding site of the *Saccharomyces cerevisiae* HIS3 promoter, increasing amounts of complex being observed at higher peptide concentrations (Figure 1, lanes 1–4). To show that the DNA binding is site-specific, a similar DNA fragment with a double point mutation was used as a negative control. When the same experiment was carried out with the mutated binding site, no complex formation could be observed under the same conditions, even at high peptide concentration (Figure 1, lanes 5–8).

The interaction of the peptide with DNA was further characterized by methylation interference analysis (Siebenlist and Gilbert, 1980), and orthophenanthroline copper footprinting (Kuwabara and Sigman, 1987), showing that the BR-LZ specifically interacts with the GCN4 cognate binding site (data not shown).

### *CD measurements*

Preliminary CD measurements were carried out on both peptides. The spectra on BR-LZ confirmed previous studies on the leucine zipper region alone demonstrating a high percentage of helical conformation (O'Shea *et al.*, 1989b). Various concentrations ($10^{-4}$–$10^{-6}$ M), temperatures (283–318 K) and pH conditions (3–9) were checked. The helicity decreased with dilution and temperature and increased with pH.

Similar studies on the BR peptide showed 40–25% of helicity at the highest concentration ($10^{-4}$ M) and temperature 278–300 K.

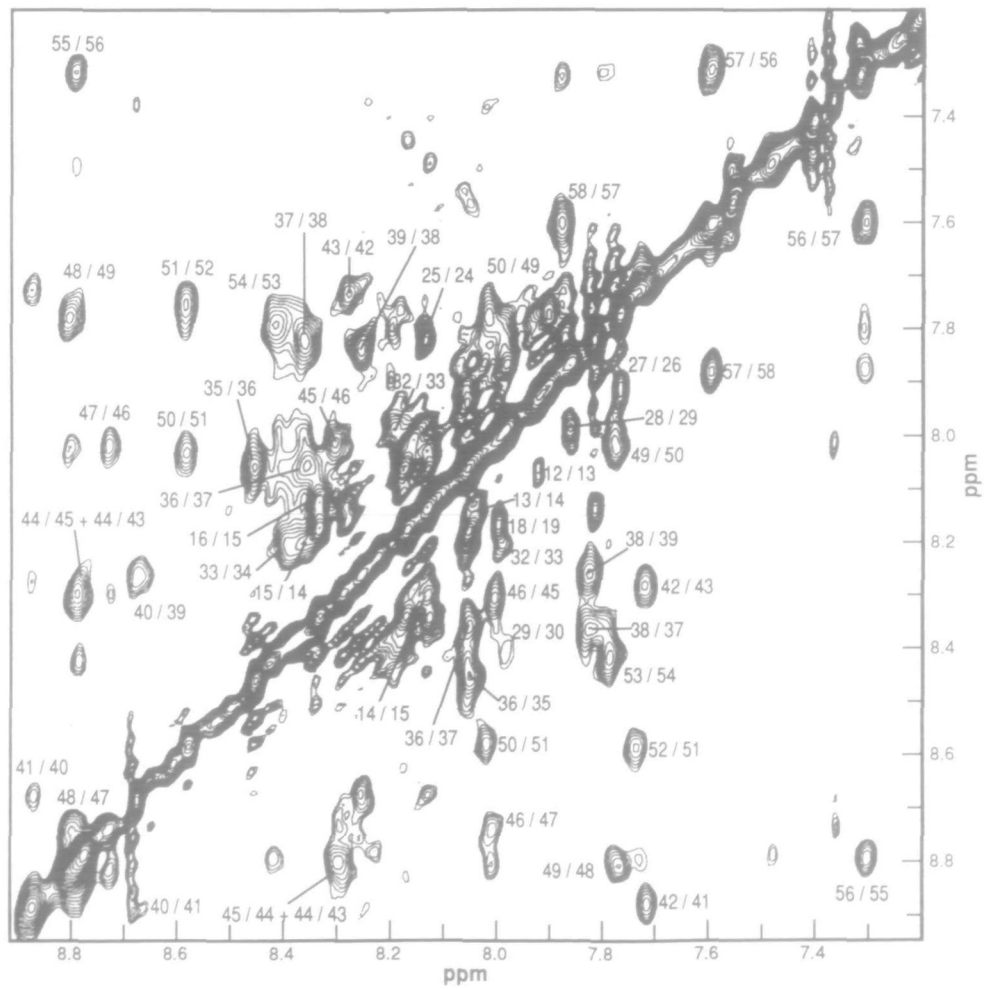


Fig. 2. Diagonal region of a NOESY spectrum of the BR-LZ peptide in $H_2O$ at 310 K and mixing time 200 ms showing the sequential cross-peaks between the amide hydrogens typical of a helical conformation. The numbers indicate the residues involved.

*Monitoring renaturation by one-dimensional NMR on BR-LZ*

Preliminary one-dimensional spectra of BR-LZ showed proton resonances much broader than expected for peptides of this size, even though dimeric. The temperature was then increased to 343 K until sharper resonances and a small chemical shift dispersion was observed. Under these conditions the protein should unfold completely. The temperature was then decreased to 310 K. Although the chemical shift spreading was now similar to the original spectrum, much sharper lines were observed. This effect was especially clear on isolated resonances such as the aromatic protons of Y45 and H46. Other resonances, such as the methyl group of the methionine, continued to show line width broadening. No variation of the line width was observed on dilution.

All subsequent experimentation utilized annealed samples (see Materials and methods).

*Proton NMR assignment of BR-LZ*

The interpretation of the spectra was complicated by the presence of several amino acid types and by the lack of dispersion of the chemical shifts in some of the relevant regions. Nonetheless, almost complete sequence-specific assignment of the resonances was achieved using the procedures described by Wuethrich (1986), as well as the main chain directed method (Englander and Wand, 1987). Inter- and intra-residue NOE connectivity patterns characteristic of secondary structure units were analysed first. The NOESY spectrum in $H_2O$ revealed a large number of strong $d_{HN-HN}$ connectivities that were used as starting points for the sequential assignment. These connectivities, together with sequential $d_{H\alpha-HN}$ and $d_{H\beta-HN}$, led to recognition of stretches of adjacent residues. To locate these short regions in the sequence, unique residues (e.g. Met, Tyr and His) were recognized easily and used as starting points to trace the sequential assignment.

The spin systems of the individual amino acid residues were then identified in $D_2O$ solution using NOESY (Kumar *et al.*, 1981), DQF COSY (Nagayama *et al.*, 1980) and HOHAHA (Branschweiler and Ernst, 1983; Bax and Davis, 1985) experiments at different mixing times. Nearly all the C$\beta$ protons could be observed as well as most of the C$\gamma$ protons. Although the spin system of Arg20, Arg21 and Arg61 could be observed, the attribution of these three residues in the sequence was prevented by the overlap of the signals. The typical NOE pattern and ROESY exchange cross peaks (Bothner-By *et al.*, 1984) showed that both prolines were present in equilibrium between *cis* and *trans* conformations.

The full list of assignments is given in Table I. As an example, the amide proton resonance region of a NOESY experiment of BR-LZ with the main chain connectivities indicated is shown in Figure 2.

*Secondary structure assignment of BR-LZ*

Direct evidence for the presence of helical secondary structure was given by the presence of the $d_{HN-HN}(i,i+1)$, $d_{H\alpha-HN}(i,i+3)$ and $d_{H\beta-HN}(i,i+3)$ connectivities. Interruptions in observed connectivities exist where peaks were occluded by overlap. No H$\alpha$−HN($i,i+4$) distances could be observed. This is not unusual, however, because these peaks are weak and even small distortions from the $\alpha$-helical conformation can prevent this observation.

According to the information inferred from these NMR parameters most of the peptide is helical. However, there are three clearly distinct structural regions. The first seven residues in the sequence are disordered. Residues 8−31 (the basic region) show only a small number of inter-residue interactions, indicating high flexibility. All the observable NMR parameters are consistent with a helical conformation. HN−HN connectivities could be followed in ROESY and NOESY experiments with longer

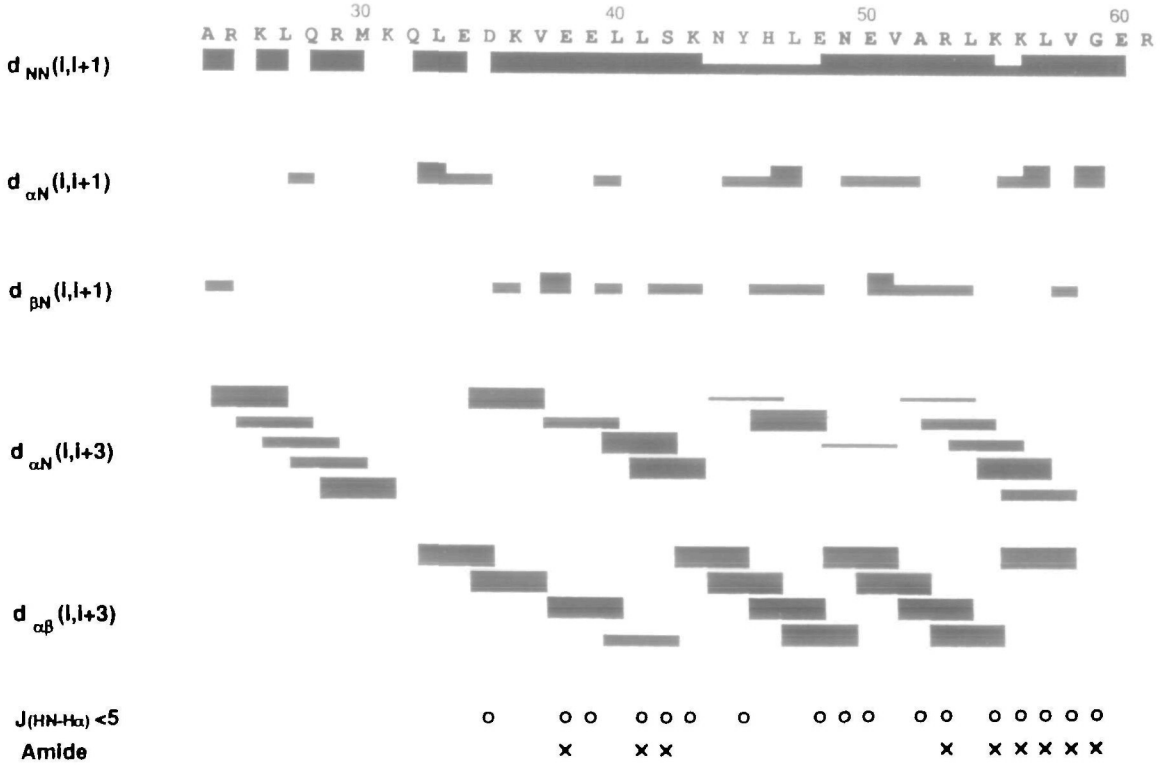


Fig. 3. Primary sequence of the leucine zipper sub-domain with the summary of the experimental data used for the sequential assignment of the main chain resonances. Sequential connectivities $d_{HN-H\alpha}$, $d_{HN-H\beta}$ and $d_{HN-HN}$ are classified into three categories according to their intensities and indicated by the height of the bars. Crosses indicate the position of the slowly exchanging HN hydrogens.

mixing times in the range 8−20. The region between residues 32 and 59 (the leucine zipper sub-domain) forms a long, uninterrupted helix in agreement with previous results on the isolated sub-domain (Oas *et al.*, 1990).

The structures of the two sub-domains are reflected in the spectra quite differently. The difference is most probably caused by the very different mobilities of the two sub-domains. Residues in rigid regions of a protein cross relax with many neighbours and relax quickly, giving rise to many cross relaxation effects but very poorly resolved *J*-coupling effects. The converse is true for mobile regions (Jardetzky and Roberts, 1981).

The NOESY spectra of LZ-BR contain mainly the signals from the residues of the leucine zipper region. For instance, there is no trace at all of the leucines from the leucine zipper sub-domain in the HOHAHA spectra, whereas they dominate the NOESY spectra. A similar behaviour is also shown by the side chains of the leucine zipper arginines, lysines and by the $\epsilon$ methyl of M30, groups usually not readily observed in the NOESY spectra of proteins (Wuethrich, 1986, p.143). Most of the side chains of residues in the basic region are clearly observable only in the *J*-correlated spectroscopy (HOHAHA, DQF-COSY).

Experiments involving hydrogen exchange elucidated the nature of the interaction between monomers. Some amide hydrogens were preferentially stable against exchange in deuterated solvent, i.e. V37, L40, L41, A52, L54, K55, K56, L57, V58 and G59 (Figure 3). Because these residues are predominantly on the hydrophobic side of the helix, their inertia against exchange cannot be explained solely by increased stability due to the hydrogen bonding pattern, typical of helices. This part of the structure must therefore be the contact surface of the dimer, protected from access to the solvent.

The next question was the arrangement of the two monomers in the dimer. This point has been already discussed by Oas *et al.* (1990), with the conclusion that the lack of perceived distances between residues more than four units apart must imply the presence of a parallel symmetrical arrangement. Our results extend this conclusion to the complete DNA-binding domain. The observation of an unbroken helix over the whole leucine zipper sub-domain and the lack of head−tail distances definitely exclude any possibility other than a parallel symmetrical arrangement of the helices. Of the two models presented by Oas *et al.* for the parallel dimer, interdigitation of the leucine side chains can also be excluded *a priori* because the side chains would break well-known packing rules (Chothia *et al.*, 1981).

*Distance geometry calculations and model building*

Calculations were performed using the variable target function program DISMAN (Braun and Go, 1985). Not enough NOEs could be observed to allow calculations on the basic region. We restricted the calculations to the well-defined region only, which includes the leucine zipper and seven adjoining residues (24−30) from the contiguous region. A total of 305 upper distance restraints could be extracted in this region. A survey of the short-range connectivities $d_{HN-H\alpha}$, $d_{HN-H\beta}$ and $d_{HN-HN}$ observed is shown in Figure 3. Seventy-seven distances were between protons belonging to contiguous residues, 58 between protons belonging to residues more than one residue apart.

Several preliminary runs were performed to check the internal consistency of the data. First, no restrictions were imposed on the dihedral angles of backbone atoms to test whether the distance restraints were enough to fold the peptide. Having established this, we then used the $J_{HN-H\alpha}$ information to add additional restraints on the $\phi$ angles. The next problem was how to distinguish between intra- and inter-molecular effects. In prin-

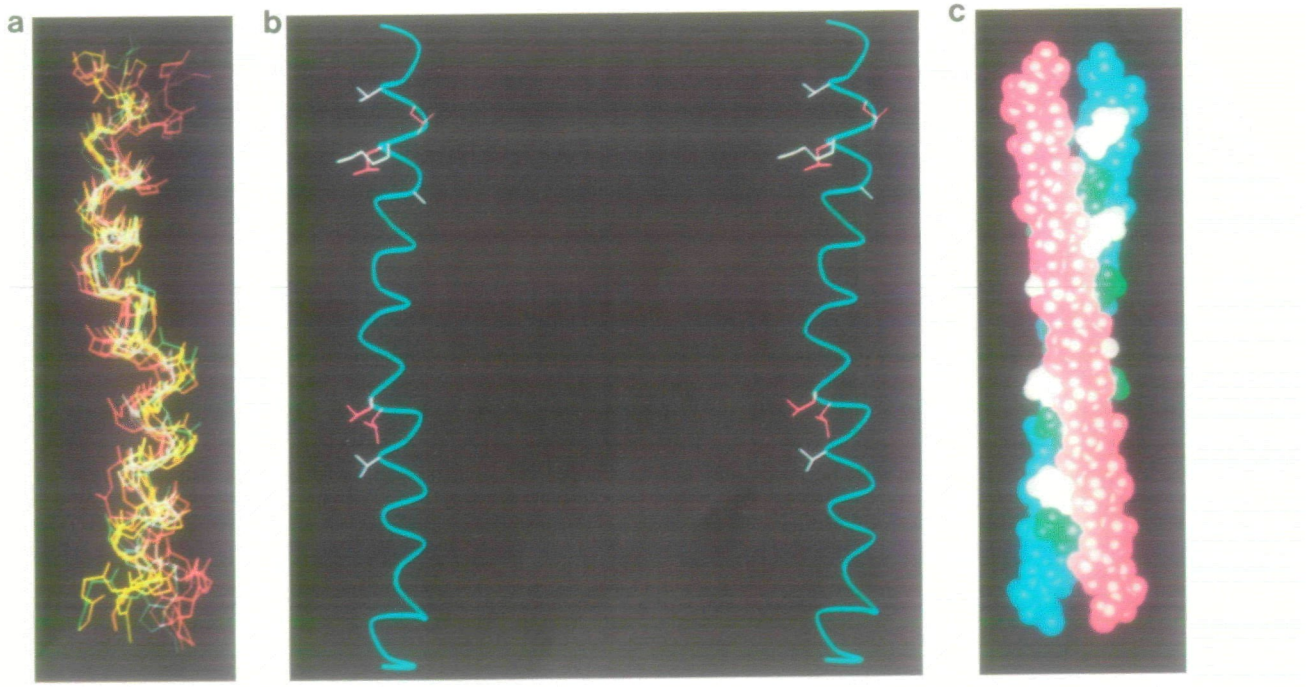


**Fig. 4.** (a) Superposition of the eight structures with the lowest number of violations of the leucine zipper domain of the BR-LZ peptide derived from distance geometry calculations. The structures are superposed for minimum RMS deviation between backbone N, $C_\alpha$, C and O atoms. (b) A stereo representation of the leucine zipper domain showing the side chains of residues with amide hydrogens preferentially stable against exchange in deuterated solvent (see text). The backbone is shown as a cyan ribbon, leucines are magenta and the other side chains white. Two stable regions are separated by a less stable region containing N44 in the interface. (c) The structure of the dimer modelled according to the observed NMR restraints in space-filled representation, showing only the backbone and the interface side chains. One monomer is represented in magenta for the backbone and green side chains, the other with cyan backbone and white side chains.

ciple, because of the intrinsic symmetry of the parallel arrangement of the dimer, each distance involving residues in the interface could be a sum of the two effects (Breg *et al.*, 1989). Experimental solutions require asymmetric isotopic labelling (Weiss, 1990), which might be possible in the future. From previous experience on another dimeric protein, ROP (W.Eberle, A Pastore, C.Sander and P.Roesch, submitted), we observed that inter-molecular distances are usually rather weakly discriminated and often observable only with very long mixing times. In addition, since the main interaction is determined by side chain packing, inter-molecular effects are observed mainly between side chains.

From these assumptions, we repeated two sets of calculations. In the first we included only distances from NOESY experiments with the shortest mixing time. In the second we excluded distances involving side chains. The results were then compared and the backbone conformation shown to be the same in both cases.

Twenty final structures using the complete set of restraints were calculated. Figure 4(a) shows a representive set of the eight final structures with the lowest number of violations clearly folded in helical conformation. For these structures, the maximum violation of the upper bound distance restraints ranged between 0.45 and 0.68 Å, whereas the sum of violations was 6.1−9.1 Å. The final average root mean square (RMS) deviation between structure pairs for the backbone atoms was 2.87 Å.

We note a slight distortion of the helix axis accommodating the hydrophobic interface. Ten additional structures were also calculated including distances $d_{H\alpha-HN}$ $(i,i+3)$ and $d_{H\beta-HN}$ $(i,i+3)$ where they could be unobservable because of other overlapping cross peaks. These distances were all included as medium restraints (3.5 Å). The result of this calculation shows a more regular helix.

The extension of these calculations to the dimer structure is precluded by the intrinsic symmetry of the molecule. We used instead a modelling approach to investigate the compatibility of our experimental observations with the coiled-coil model for paired α-helices. Starting from standard values for $\phi$ and $\psi$ angles (Parry and Suzuki, 1969; McLachlan, 1978), we built a model of an idealized coiled-coil helix. In our sequence, position 2 of the coiled-coil heptad repeat would correspond to A23 so that the residues in the interface would correspond to M30, L33, V37, L40, N44, L47, V51, L54 and V57, in agreement with the experimental data. The conformation of the side chains in the interface was consistent with the statistically allowed values (McGregor *et al.*, 1987). When more than one rotamer was possible (leucines, methionine), the conformation was chosen

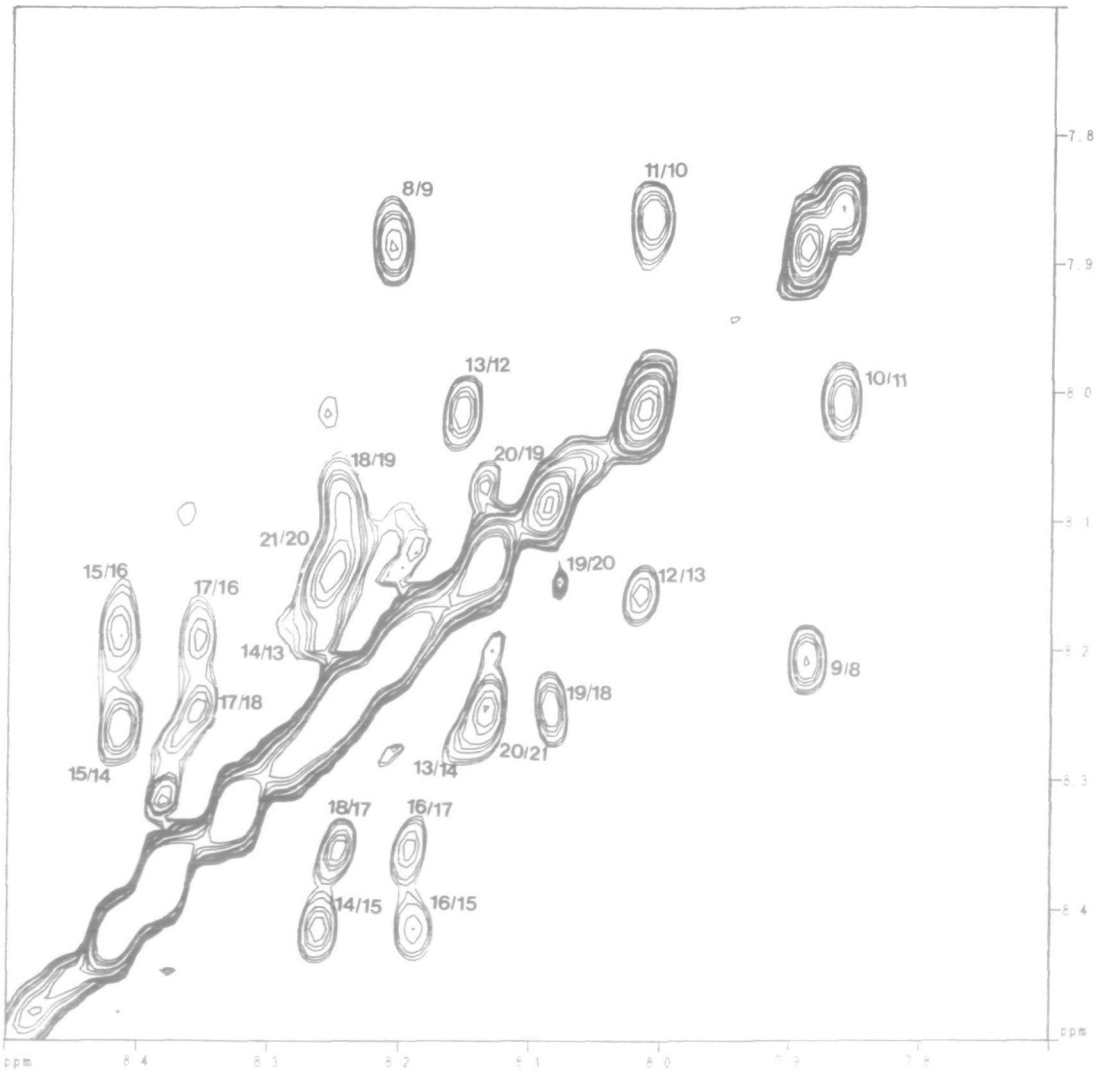


**Fig. 5.** Diagonal region of a NOESY spectrum of the BR fragment in $H_2O$ at 300 K and mixing time 300 ms showing the sequential cross-peaks between the amide hydrogens typical of a helical conformation The numbers indicate the residues involved.

which matched the NMR data. Two copies of the monomer were then docked together with graphic support (Dayringer *et al.*, 1986) as a symmetrical parallel dimer.

A space-filled picture of the dimer model is shown in Figure 4(c).

*NMR studies on the BR peptide*

Peak overlap and higher flexibility prevented a detailed observation of the structure of the basic region in BR-LZ. We thought of using the BR peptide, a fragment from the complete synthesis, to study this region more specifically. However, we were expecting the sub-domain, already flexible in BR-LZ, to be completely disordered when isolated. To the contrary, the NMR spectrum of BR is very similar to the corresponding resonances in BR-LZ. Relatively longer mixing times in NOESY experiments were necessary (between 300 and 800 ms), but very clear HN−HN connectivities could be observed between residues 8 and 21 (see Figure 5). In this range some of the other cross-peaks typical of helices could also be observed. The structural independence of the basic region from the leucine zipper sub-domain confirms previous studies on domain swapping (Sellers and Struhl, 1989). Further studies on this peptide are now in progress and a more detailed comparison between the 61 residue peptide and this fragment will be reported elsewhere (V.Saudek, H.Pasley, T.Gibson, H.Gausepohl, R.Frank and A.Pastore, in preparation).

## Discussion

*Structure elucidation of synthetic peptides*

The 61 residue BR-LZ peptide is the largest synthetic peptide yet reported to yield a high quality two-dimensional NMR dataset. This work shows that relatively large synthetic peptides corresponding to known protein domains are suitable for structural studies using NMR techniques. It is much quicker to synthesize a peptide of 50 or 60 residues in milligram quantities than to prepare sufficient quantities by expression *in vivo*. The synthesis and the purification were completed in 2 weeks.

One of the concerns with synthetic peptides is whether they can fold into the native structure. Temperature-induced denaturation and renaturation of the GCN4 peptide were followed by one-dimensional NMR. Furthermore, a functional assay demonstrated the structural integrity of the peptide. Correct folding does in fact appear to be realizable with synthetic peptides, since in another example the structure of a functional 99 residue synthetic retroviral protease has been determined by X-ray crystallography (Wlodawer *et al.*, 1989). Thus, no fundamental barriers appear to exist to the structural elucidation, piece by piece, of the numerous multi-domain proteins of eukaryotes (of which the transcription factors are only a subset) by *in vitro* synthesis and analysis of each domain separately.

*Structure of the GCN4 DNA-binding domain in solution and comparison with models*

In solution, the BR-LZ peptide is essentially a single continuous helix, dimerized through the leucine zipper sub-domain. The basic region shows none of the side chain packing interactions necessary to stabilize a helix and fluctuates about the helical conformation. The amino acid sequence of this region lacks in fact sufficient hydrophobic residues to form an interface with its counterpart or any other part of the peptide. In the absence of packing interactions, helices are unstable (Chothia *et al.*, 1981).

In the 'scissors grip' model, the basic region is suggested to form two consecutive helices, broken at the conserved N15, which track around the major groove of the target site (Vinson *et al.*, 1989). However, N15 is definitely helical in the BR peptide (see Figure 5), though this information was hidden by peak overlap in BR-LZ. The connection between the leucine zipper and the basic region (residues 23−32) is also clearly helical in BR-LZ with continuous $i$-$i$+3 connectivity. Breaks in the helicity upon binding DNA cannot be excluded but are not supported by the solution data.

Other reasons for the conservation of N15 can be envisaged. In particular, the basic region must at some point come close to the sugar-phosphate backbone whence the amide group, with its partial positive charge, makes asparagine the ideal small side chain to interact with a phosphate. Despite the curvature of the major groove, a single helix in the basic region would be quite adequate to interact specifically with 3 bp as required for the GCN4 half-site. Indeed, a high resolution structure of a bacterial repressor/DNA complex shows that the recognition helix can interact with up to 5 bp (Aggarwal *et al.*, 1988).

## Acknowledgements

We gratefully acknowledge Prof. A.M.Lesk for valuable discussions, advice and constant encouragement. Also we would like to thank Dr P.Sassone-Corsi and Dr S.J.Busch for helpful discussions and Dr M.Saviano for assistance in the CD measurements.

## References

Aggarwal,A K , Rodgers,D.W., Drottar,M , Ptashne,M. and Harrison,S.C. (1988) *Science*, **242**, 899−907.
Bax,A. and Davis,D.G. (1985) *J. Magn. Resonance*, **65**, 355−360.
Bothner-By,A A , Stephens,R.L., Lee,J., Warren,C.D. and Jeanloz,R.W. (1984) *J. Am. Chem. Soc.*, **106**, 811−813
Branschweiler,L. and Ernst,R R. (1983) *J. Magn. Resonance*, **53**, 521−528.
Braun,W. and Go,N. (1985) *J Mol. Biol.*, **186**, 611−626.
Breg,J.N , Boelens,R., George,A.V.E. and Kaptein,R. (1989) *Biochemistry*, **28**, 9826−9833.
Chothia,C , Levitt,M. and Richardson,D. (1981) *J. Mol. Biol.*, **145**, 215−250.
Dayringer,H E., Tramontano,A., Sprang,S.R. and Fletterick,R.J. (1986) *J. Mol. Graph* , **4**, 82−87.
Englander,S.W. and Wand,A.J. (1987) *Biochemistry*, **26**, 5953−5958.
Frank,R. and Gausepohl,H. (1988) In Tschesche,H (ed.), *Modern Methods in Protein Chemistry*. de Gruyter, Berlin, Vol 3, pp. 41−60.
Jardetzky,O. and Roberts,G.C.K. (1981) *NMR in Molecular Biology*. Academic Press, New York.
Kouzarides,T. and Ziff,E. (1989) *Nature*, **340**, 568−571.
Kumar,A , Wagner,G., Ernst,R.R. and Wuethrich,K. (1981) *J. Am. Chem. Soc.*, **103**, 3654−3658.
Kuwabara,M.D. and Sigman,D S. (1987) *Biochemistry*, **26**, 7234−7238.
Hinnebusch,A.G. (1984) *Proc. Natl Acad. Sci. USA*, **81**, 6442−6446.
Hope,I.A. and Struhl,K. (1987) *EMBO J.*, **9**, 2781−2784.
Landschutz,W.H., Johnson,P F. and McKnight,S.L. (1988) *Science*, **240**, 1759−1764.
Lesk,A.M. and Hardman,K.D (1982) *Science*, **216**, 539−540.
McGregor,M J., Islam,S.A. and Sternberg,M.J. (1987) *J. Mol. Biol.*, **198**, 295−310
McLachlan,A.D. (1978) *J. Mol. Biol.*, **12**, 493−506.
Mueller,U., Roberts,M.P., Engel,E.A., Doerfler,W. and Shenk,T. (1989) *Genes Dev* , **3**, 1991−2002
Nagayama,K , Kumar,A., Wuethrich,K. and Ernst,R.R. (1980) *J. Magn. Resonance*, **40**, 321−334.
Oas,T.G., McIntosh,L.P., O'Shea,E.K., Dahlquist,F.W and Kim,P.S. (1990) *Biochemistry*, **29**, 2891−2894.
O'Shea,E.K., Rutkowski,R. and Kim,P.S. (1989a) *Science*, **243**, 538−542.
O'Shea,E.K., Rutkowski,R., Stafford,W.F. and Kim,P.S. (1989b) *Science*, **245**, 646−648.
Parry,D.A.D. and Suzuki,E. (1969) *Biopolymers*, **7**, 189−206.
Ransone,L.J., Visvader,J., Sassone-Corsi,P. and Verma,I.M. (1989) *Genes Dev.*, **3**, 770−781.
Risse,G., Jooss,K., Neuberg,M., Brueller,H.-J. and Mueller,R. (1989) *EMBO J.*, **8**, 3825−3832.

Schuermann,M., Neuberg,M , Hunter,B , Jenuwein,T., Ryseck,R.P., Bravo,R and Mueller,R. (1989) *Cell*, **56**, 507–516
Sellers,J.W. and Struhl,K (1989) *Nature*, **341**, 74–76.
Siebenlist,U. and Gilbert,W. (1980) *Proc Natl Acad Sci. USA*, **77**, 122–126.
Turner,R. and Tjian,R. (1989) *Science*, **243**, 1689–1694
Vinson,C.R., Sigler,P.B. and McKnight,S.L (1989) *Science*, **246**, 911–916
Weiss,M.A (1990) *J. Magn. Resonance*, **86**, 626–632
Wlodawer,A., Miller,M., Jaskolski,M , Sathyanarayana,B K., Baldwin,E , Weber,I.T., Selk,L.M , Clawson,L., Schneider,J and Kent,S.B H (1989) *Science*, **245**, 616–621
Wuethrich,K. (1986) *NMR of Proteins and Nucleic Acids*. Wiley, New York

*Received on May 5, 1990, accepted on July 17, 1990*