

## Solution structure of the DNA binding octapeptide repeat of the *K10* gene product

Masashi Suzuki<sup>1</sup>, David Neuhaus, Mark Gerstein<sup>2</sup> and Saburo Aimoto<sup>3</sup>

MRC Laboratory of Molecular Biology, Hills Road, Cambridge, CB2 2QH, UK and <sup>3</sup>Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka, 565 Japan

<sup>2</sup>Present address: Beckman Laboratory for Structural Biology, Department of Cell Biology, Stanford Medical School, Stanford, CA 94305-5400, USA.

<sup>1</sup>To whom correspondence should be addressed

A putative transcription factor, the *Drosophila K10* gene product, contains eight repeats of the octapeptide sequence SPNQQQHP or close variants. The solution structure of the *K10* repeat was studied by NMR using a peptide composed of two SPNQQQHP units (referred to here as HP2). To overcome problems caused by degeneracy of backbone amide signals of Gln residues, a series of synthetic peptides containing an <sup>15</sup>N-labelled main chain amide at different positions in HP2 were synthesized. In aqueous trifluoroethanol solution, HP2 folds into two structural units; the SPNQ part of each unit folds into a turn structure, while the C-terminal part shows some helical characteristics but is less structured. The N-terminal turn is likely to provide a core that produces a more stable helical structure upon binding to DNA and probably 'caps' the segmented helical unit at its N-terminus. This model is supported by a DNA footprinting study which shows that one SPNQQQHP unit spans four base pairs upon binding to A/T-rich sequences of DNA.

**Key words:** CD/DNA–protein interaction/footprinting/NMR

### Introduction

The *Drosophila K10* gene is one of several maternally acting genes which establish the dorso-ventral polarity of an embryo (Prost *et al.*, 1988). Its gene product is a nuclear protein and a putative transcription factor (Prost *et al.*, 1988), but does not contain any well-established DNA binding sequence motifs. Its DNA binding mode has not been characterized previously.

The *K10* gene product has eight octapeptide repeats (Figure 1a, see also Suzuki, 1991, 1992), the first of which begins with the Ser-Pro-Lys-Lys sequence. Repeats of similar sequences, 'SPKK' motifs, are found in sea urchin histones, H1 and H2B (Suzuki, 1989). It is clear that the SPKK motif binds to DNA and the structure and DNA binding mode of the SPKK motif have been studied by physicochemical and biochemical methods (Churchill and Suzuki, 1989; Suzuki, 1989; Suzuki *et al.*, 1993). Thus, the *K10* repeat is also likely to be a DNA binding domain. However, the most frequently observed unit in *K10* is Ser-Pro-Asn-Gln-Gln-Gln-His-Pro. Thus, a novel peptide structure for DNA binding is expected.

### Materials and methods

#### Peptides

The parent peptide, HP2 [H-(Ser-Pro-Asn-Gln-Gln-Gln-His-Pro)<sub>2</sub>-NH<sub>2</sub>], was synthesized by a solid phase method and

purified by HPLC. Peptides in which the main chain NH of the residue marked with an asterisk (\*) was labelled with <sup>15</sup>N, were synthesized similarly.

HP2[Q5]: H-(Ser-Pro-Asn-Gln-Gln\*-Gln-His-Pro)-(Ser-Pro-Asn-Gln-Gln-Gln-His-Pro)-NH<sub>2</sub>

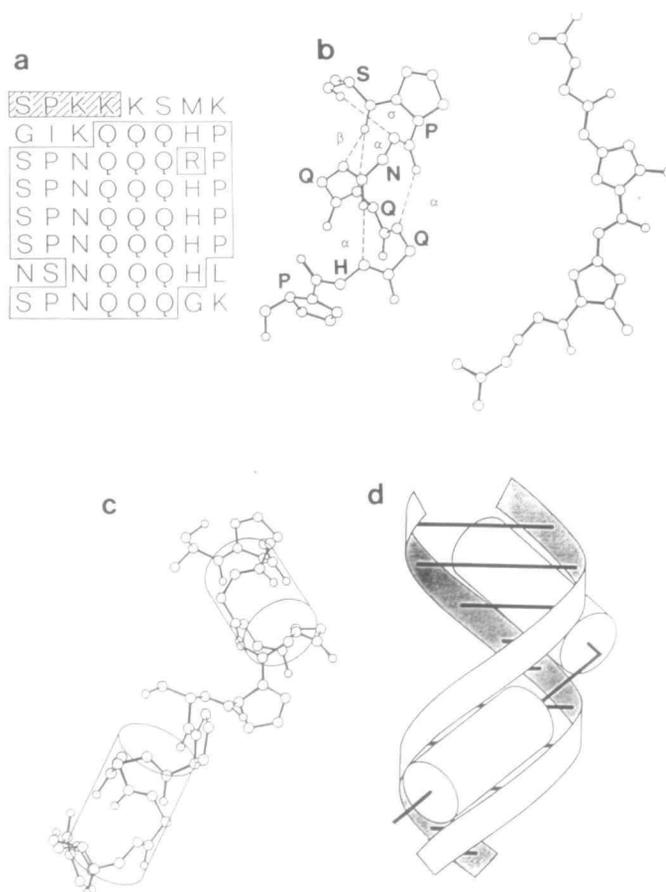
HP2[Q6]: H-(Ser-Pro-Asn-Gln-Gln-Gln\*-His-Pro)-(Ser-Pro-Asn-Gln-Gln-Gln-His-Pro)-NH<sub>2</sub>

HP2[Q13]: H-(Ser-Pro-Asn-Gln-Gln-Gln-His-Pro)-(Ser-Pro-Asn-Gln-Gln\*-Gln-His-Pro)-NH<sub>2</sub>

HP2[Q14]: H-(Ser-Pro-Asn-Gln-Gln-Gln-His-Pro)-(Ser-Pro-Asn-Gln-Gln-Gln\*-His-Pro)-NH<sub>2</sub>

The following peptides were also synthesized.

HP1: H-Ser-Pro-Asn-Gln-Gln-Gln-His-Pro-NH<sub>2</sub>



**Fig. 1.** Sequence and proposed structure of the *K10* octapeptide repeat. (a) The sequence of *K10* is that determined by Prost *et al.* (1988). The sequence contains eight tandem repeats of an octapeptide. The models proposed for (b) one unit SPNQQQHP and (c) two units (SPNQQQHP)<sub>2</sub>. The dihedral angles used are (–57°, 130°)–(–57°, –36°)–(–63°, –47°)–(–57°, –47°)–(–57°, –47°)–(57°, –47°)–(–113°, 105°)–(–88°, 120°) for the SPNQQQHP sequence. Side chains of Asn, Gln and His are omitted. In panel b the model is compared with the structure of netropsin (Kopka *et al.*, 1985), showing hydrogen bonds with dashed lines. (d) A schematic drawing of DNA binding by two SPNQQQHP units.

PH2: H-Pro-(Ser-Pro-Asn-Gln-Gln-Gln-His-Pro)-Ser-Pro-Asn-Gln-Gln-Gln-His-NH<sub>2</sub>

HP4: H-(Ser-Pro-Asn-Gln-Gln-Gln-His-Pro)<sub>4</sub>-NH<sub>2</sub>

QN1: H-Gln-Gln-Gln-His-Pro-Ser-Pro-Asn-OH

S4: H-Gly-(Ser-Pro-Arg-Arg)-(Ser-Pro-Arg-Lys)-(Ser-Pro-Lys-Lys)-(Ser-Pro-Arg-Lys)-NH<sub>2</sub>

#### NMR measurements

The HP2 peptide was dissolved in 75% 1,1-[<sup>2</sup>H<sub>2</sub>]2,2,2-trifluoroethanol (TFE)–25% H<sub>2</sub>O to yield a 10 mM solution. The isotope-labelled peptides were dissolved in 90% TFE–10% H<sub>2</sub>O. Three sets of 2-D [<sup>1</sup>H]NMR spectra were measured: COSY, TOCSY and NOESY spectra of HP2 at 5°C in 75% TFE; COSY, TOCSY, NOESY and ROESY spectra of HP2[Q6] at 5°C in 90% TFE; and COSY, TOCSY, NOESY and ROESY spectra of HP2[Q14] at 5°C in 90% TFE. Spectra were recorded at 500 MHz using a Bruker AMX500 spectrometer and processed according to the method described in Suzuki *et al.* (1993). For NOESY experiments in the mixing time was 200 ms and for ROESY 100 ms. For 2-D experiments with labelled peptides, <sup>15</sup>N-decoupling in F<sub>1</sub> was achieved using a 180° <sup>15</sup>N-pulse in the middle of t<sub>1</sub> and in F<sub>2</sub> using composite pulse decoupling (GARP) during t<sub>2</sub>. Spectra were referenced relative to internal ~0.1 mM sodium 2,3-[<sup>2</sup>H<sub>4</sub>]3-trimethylsilylpropionate.

#### CD measurements

CD spectra were recorded at 20°C using a circular dichrograph (Jobin Yvon CD VI) equipped with a computer. A 1 mm cuvette was used for measurements.

#### DNA footprinting

A 160 bp DNA fragment, TyrT (Drew and Travers, 1984), was prepared from Δ98 plasmid by *Ava*I–*Eco*RI digestion and labelled at its 3-terminus with <sup>32</sup>P[α-dCTP] by reverse transcriptase. Copper(II)-*o*-phenanthroline footprinting experiments were carried out essentially according to Spassky and Sigman (1985). After electrophoresis, gel densitometry was carried out according to Smith and Thomas (1900).

#### Model building study

Models were built using the program CHARMM (version 2.1, Brooks *et al.*, 1983) and XPLOR (version 2.2, Brunger *et al.*, 1987), displayed using INSIGHT (Dayringer *et al.*, 1986) and FRODO (Jones, 1982) and plotted with INSIGHT (Dayringer *et al.*, 1986) and ARTPLOT (Lesk and Hardman, 1982).

## Results

### General features of the NMR spectra

Sequence-specific resonance assignments for the peptide HP2 in both 90 and 75% aqueous TFE solution were made following the standard methodology (Wüthrich, 1986); three portions of the NOESY spectrum of HP2[Q14] are shown in Figure 2. The chemical shifts of the various isotope-labelled peptides are essentially identical (Figure 3a and b). The spectra recorded in 75% TFE were more heavily overlapped than those recorded in 90% TFE, particularly amongst the resonances of the six Gln residues (Figure 3b). Although most effort was therefore concentrated on analysing spectra from the 90% TFE solution, many of the corresponding signals could be identified in spectra from the 75% TFE solution and in those cases that could be compared, the NOE connectivities observed were similar in the two solvent compositions (Figure 3c). All NOE cross-peaks had the same sign as the diagonal, i.e. the molecules are within the negative NOE regime in both solvent compositions.

At least one set of minor resonances was also observed, but the intensity of these minor signals was only a few per cent of that of the major set. We conclude that the minor signals probably arise from different conformations of the same peptide rather than from an impurity, because the intensity of the minor signals varies according to the solvent composition, being slightly higher in the 75% TFE solution. The most likely origin of this conformational equilibrium is *trans*–*cis* isomerism of X–Pro amide linkages (see for example, Suzuki *et al.*, 1993).

Although only partial assignments were made for the minor set of signals (data not shown), it was clear from these that the most likely sites for such isomerism were the two His–Pro linkages. Thus, the major conformer showed NOE cross-peaks between His C<sub>α</sub>H and Pro C<sub>β</sub>H for both His–Pro linkages (see cross-peaks 35A+B and 15A+B in Figure 2), as expected for a *trans* arrangement (Wüthrich, 1986), whereas the minor set of signals included NOE cross-peaks linking His C<sub>α</sub>H and Pro C<sub>α</sub>H signals (see cross-peaks 305 and 306 in Figure 2), as expected for a *cis* amide linkage. Presumably, the minor signals result from two species, in each of which one of the two His–Pro linkages was *cis*, but this could not be detected directly from the spectra. The two Ser–Pro links of the major conformation are *trans* (see cross-peaks 4 and 104 in Figure 2), while the arrangement of the corresponding links in the minor forms remains undetermined. In what follows, the structure of the major component only is discussed.

### Structure of the SPNQQQHP unit

The two Ser-Pro-Asn-Gln-Gln-Gln-His-Pro units in HP2 (hereafter referred to as units I and II) show very similar patterns of NOE connectivities, suggesting that their structures are essentially identical. In each unit, the N-terminal part (Ser-Pro-Asn-Gln-Gln) is structured, but the C-terminal part (Gln-His-Pro) is less so, with non-sequential NOE connectivities only being observed within the N-terminal part (Figure 4d).

The NOE connectivities observed for the N-terminal tetrapeptide of each unit (Ser-Pro-Asn-Gln) show that this region folds into a turn structure. Previous work has shown that the Ser-Pro-X-X sequence has a strong tendency to form a turn structure, stabilized by a β-turn type hydrogen bond linking the CO of Ser(*i*) to the main chain of X(*i*+3) and/or by a σ-turn-type hydrogen bond linking O<sub>γ</sub> of Ser(*i*) to the main chain NH of X(*i*+2) (Figure 4a; see also Suzuki and Yagi, 1991). The NOE connectivities expected for the two turn types have been described in detail in Suzuki *et al.* (1993) and are also summarized in Figure 4a of this paper. The two turn types are compatible only if the β-turn is of type I (or its variant type III), but not when it is of type II, since then the main chain NH of residue X(*i*+2) is oriented away from the side chain of Ser(*i*). Based on the observation of a large number of these characteristic NOE connectivities in each Ser-Pro-Asn-Gln unit of HP2 (see Figure 4b), we conclude that each such unit folds into a combination of a type I β-turn and a σ-turn, that is, they each form a βσ-turn.

The remaining C-terminal part of each Ser-Pro-Asn-Gln-Gln-His-Pro unit is less structured. However, there are some indications that the C-terminal part forms a 'nascent helix'. The term, nascent helix, has been coined by Dyson *et al.* (1988) to refer to a peptide undergoing a dynamical interconversion between extended and structured states, in which local turn structures form transiently but do not possess the long-range characteristic of a fully formed helix. Such a nascent helix typically shows stronger sequential *d*<sub>NN</sub> connectivities than expected for an extended structure, but does not show the medium

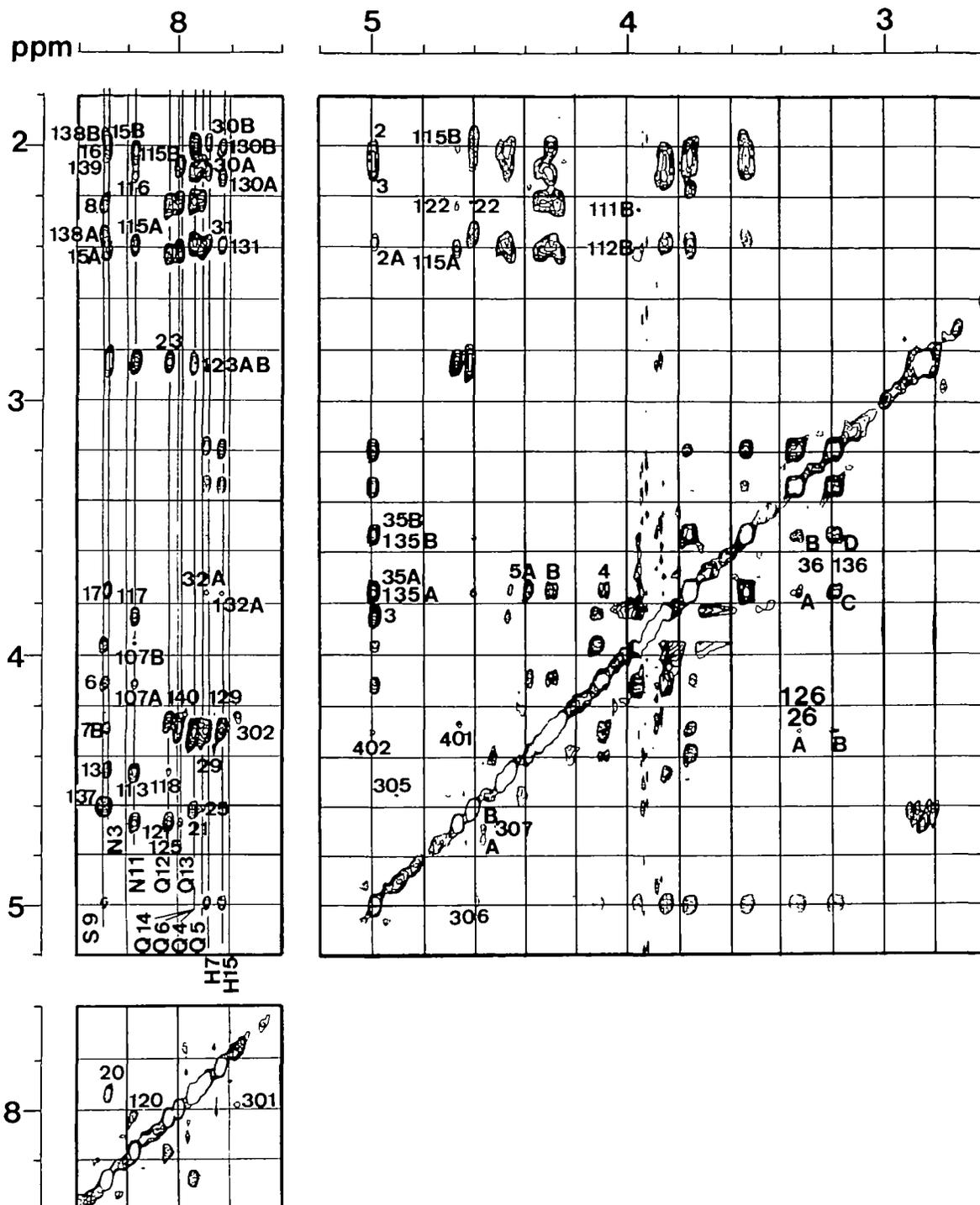


Fig. 2. Three portions of a NOESY spectrum of HP2. Regions of a NOESY spectrum of HP2[Q14] measured in 90% TFE–10% water at 5°C are shown. Numbers <100 and those between 100 and 200 are used respectively for NOE connectivities of the first and second units, respectively (see Figure 3c). Connectivities 301–306 are those identified for the minor conformation; 301 Gln6 NH–His7 NH, 302 His7 NH–Gln6 C $\alpha$ H, 305 His7 C $\alpha$ H–Pro8 C $\alpha$ H, 306 His15 C $\alpha$ H–Pro16 C $\alpha$ H, 307A,B Ser1 C $\alpha$ H–Ser1 C $\beta$ H<sub>2</sub>. Connectivity 401 appears to be due to spin-diffusion between Asn11 C $\alpha$ H and Gln12 C $\alpha$ H, while 402 may be due to spin-diffusion between Gln6 C $\alpha$ H–His7 C $\alpha$ H and/or Gln14 C $\alpha$ H–His15 C $\alpha$ H.

range connectivities  $d_{\alpha N}(i, i+3)$  or  $d_{\alpha\beta}(i, i+3)$  expected for a fully formed helix (Dyson *et al.*, 1988).

It may be seen from Figures 3c and 4d that each Gln–Gln–His–Pro tetrapeptide shows some  $d_{NN}$  connectivities consistent with nascent helix formation. In particular, the Gln residues immediately following the  $\beta\sigma$ -turn in each unit (i.e. Gln5 and Gln13) show strong  $d_{NN}$  connectivities to their N-terminal neighbour. A few medium-range NOE connectivities consistent

with at least transient helix formation were also observed involving these particular Gln residues, as summarized in Figure 4d; these comprise cross-peaks from the main chain NH of Gln5 to Asn3 C $\alpha$ H, Asn3 C $\beta$ H<sub>2</sub> and Pro2 C $\alpha$ H and the corresponding connectivities in unit II. Also, the connectivities within each  $\beta\sigma$ -turn linking Ser(*i*) C $\beta$ H<sub>2</sub> to Gln(*i*+3) C $\beta$ H<sub>2</sub> and C $\gamma$ H<sub>2</sub> may suggest that Gln(*i*+3) has helical characteristics.

Transient NOE enhancements in the <sup>15</sup>N-labelled peptides,

**a**

	NH	C <sub>α</sub> H		C <sub>β</sub> H		C <sub>γ</sub> H		C <sub>δ</sub> H		Others
		A	B	A	B	A	B	A	B	
<sup>1</sup> Ser	---	3.97	4.38	4.29	---	---	---	---	---	---
<sup>2</sup> Pro	---	4.47	2.44	2.02	2.16	2.16	3.75	3.75	---	---
<sup>3</sup> Asn	8.28	4.63	2.89	2.81	---	---	---	---	N <sub>δ</sub> H	7.39/6.64
<sup>4</sup> Gln	7.94	4.30	2.21	2.00	2.38	2.38	---	---	N <sub>ε</sub> H	7.29/6.46
<sup>5</sup> Gln	7.90	4.34	2.22	2.07	2.41	2.41	---	---	N <sub>ε</sub> H	7.33/6.45
<sup>6</sup> Gln	7.94	4.30	2.10	2.00	2.38	2.38	---	---	N <sub>ε</sub> H	7.23/6.36 (8.05)
<sup>7</sup> His	7.91	4.99	3.32	3.17	---	---	7.34	---	C <sub>ε</sub> H	8.44
<sup>8</sup> Pro	---	4.60	2.36	1.97	2.08	2.03	3.76	3.54	---	---
<sup>9</sup> Ser	8.29	4.99	4.10	3.96	---	---	---	---	---	---
<sup>10</sup> Pro	---	4.48	2.39	2.02	2.13	2.13	3.86	3.86	---	---
<sup>11</sup> Asn	8.17	4.67	2.85	2.85	---	---	---	---	N <sub>δ</sub> H	7.33/6.58
<sup>12</sup> Gln	8.05	4.27	2.25	2.25	2.43	2.43	---	---	N <sub>ε</sub> H	7.31/6.51
<sup>13</sup> Gln	8.00	4.31	2.21	2.07	2.41	2.41	---	---	N <sub>ε</sub> H	7.29/6.39
<sup>14</sup> Gln	7.95	4.30	2.13	2.00	2.38	2.38	---	---	N <sub>ε</sub> H	7.23/6.36
<sup>15</sup> His	7.85	4.99	3.34	3.19	---	---	7.34	---	C <sub>ε</sub> H	8.44
<sup>16</sup> Pro	---	4.49	2.39	2.02	2.09	2.09	3.76	3.51	---	---
C-terminal NH <sub>2</sub>		7.46	6.63							

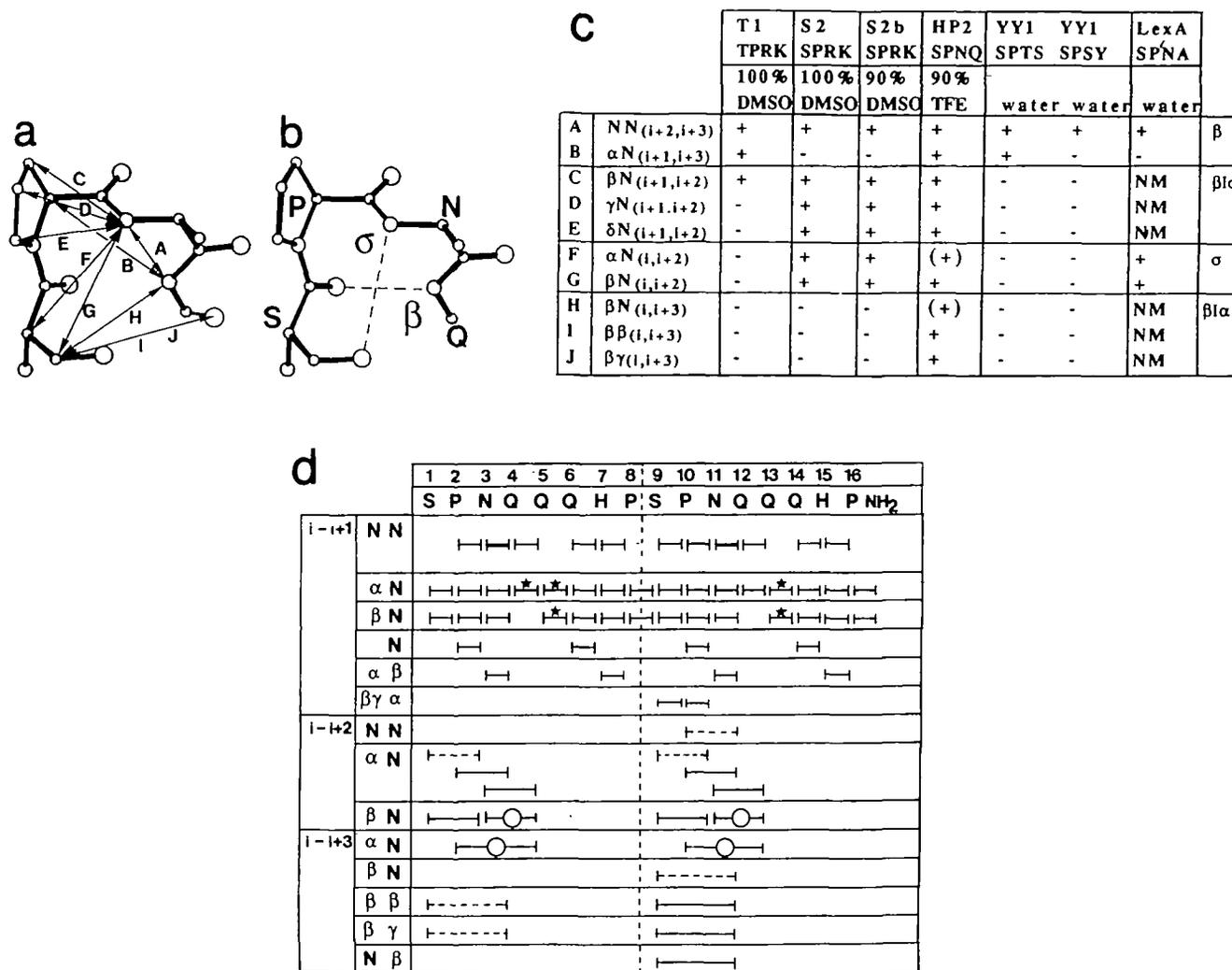
**b**

	NH	C <sub>α</sub> H		C <sub>β</sub> H		C <sub>γ</sub> H		C <sub>δ</sub> H		Others
		A	B	A	B	A	B	A	B	
<sup>1</sup> Ser	ND	3.95	4.27	4.09	---	---	---	---	---	---
<sup>2</sup> Pro	---	4.36	2.29	1.87	2.05	2.05	3.65	3.65	---	---
<sup>3</sup> Asn	8.26	4.55	2.75	2.75	---	---	---	---	N <sub>δ</sub> H	7.43/6.70
<sup>4</sup> Gln	8.06	4.26	2.30	---	---	---	---	---	N <sub>ε</sub> H	7.40-7.28/ 6.56-6.46
<sup>5</sup> Gln									N <sub>ε</sub> H	7.40-7.28/ 6.56-6.46
<sup>6</sup> Gln	7.97	4.18	---	---	---	1.86	---	---	N <sub>ε</sub> H	7.40-7.28/ 6.56-6.46
<sup>7</sup> His	8.03	4.89	3.19	3.05	---	---	7.22	---	C <sub>ε</sub> H	8.35
<sup>8</sup> Pro	---	4.47	2.22	1.82	1.95	1.95	3.43	3.65	---	---
<sup>9</sup> Ser	8.30	4.88	3.83	3.95	---	---	---	---	---	---
<sup>10</sup> Pro	---	4.36	2.25	1.90	1.98	1.98	3.75	3.75	---	---
<sup>11</sup> Asn	8.14	4.58	2.74	2.74	---	---	---	---	N <sub>δ</sub> H	7.38/6.64
<sup>12</sup> Gln	8.02	4.26	2.31	2.31	2.07	2.07	---	---	N <sub>ε</sub> H	7.37/6.54
<sup>13</sup> Gln	8.06	4.26	2.30	---	---	---	---	---	N <sub>ε</sub> H	7.40-7.28/ 6.56-6.46
<sup>14</sup> Gln	7.97	4.18	---	---	---	1.86	---	---	N <sub>ε</sub> H	7.40-7.28/ 6.56-6.46
<sup>15</sup> His	7.97	4.89	3.10	3.06	---	---	7.22	---	C <sub>ε</sub> H	8.35
<sup>16</sup> Pro	---	4.37	2.25	1.90	1.98	1.98	3.63	3.42	---	---
C-terminal NH <sub>2</sub>		7.50	6.74							

**c**

NOE	Number	90% TFE		75% TFE		NOE	Number	90% TFE		75% TFE	
		I	II	I	II			I	II	I	II
S(j)NH-P(j+1)C <sub>δ</sub>	[1]	NA	+	NA	+	Q(j+3)C <sub>α</sub> -Q(j+4)NH	[40]	NI	++	NI	NI
S(j)C <sub>α</sub> -P(j+1)C <sub>β</sub> A	[2A]	-	+	-	+	Q(j+5)NH-H(j+6)NH	[28]	+	+	NI	NI
S(j)C <sub>α</sub> -P(j+1)C <sub>β</sub> B	[2B]	-	+	-	+	Q(j+5)C <sub>α</sub> -H(j+6)NH	[29]	+	+	NI	NI
S(j)C <sub>α</sub> -P(j+1)C <sub>γ</sub>	[3]	-	+	-	+	Q(j+5)C <sub>α</sub> -H(j+6)C <sub>β</sub> A	[26A]	+	+	+	+
S(j)C <sub>α</sub> -P(j+1)C <sub>δ</sub>	[4]	++	++	+	+	Q(j+5)C <sub>α</sub> -H(j+6)C <sub>β</sub> B	[26B]	+	+	+	+
S(j)C <sub>β</sub> A-P(j+1)C <sub>δ</sub>	[5A]	+	+	+	+	Q(j+5)C <sub>β</sub> A-H(j+6)NH	[30A]	+	+	NI	+
S(j)C <sub>β</sub> B-P(j+1)C <sub>δ</sub>	[5B]	+	NI	+	NI	Q(j+5)C <sub>β</sub> B-H(j+6)NH	[30B]	+	+	NI	+
<u>S(j)C<sub>α</sub>-N(j+2)NH</u>	[6]	-	-	+	+	Q(j+5)C <sub>γ</sub> -H(j+6)NH	[31]	+	+	NI	+
<u>S(j)C<sub>β</sub>A-N(j+2)NH</u>	[7A]	+	+	+	+	H(j+6)NH-P(j+7)C <sub>δ</sub> A	[32A]	+	+	+	+
<u>S(j)C<sub>β</sub>B-N(j+2)NH</u>	[7B]	+	+	+	+	H(j+6)NH-P(j+7)C <sub>δ</sub> B	[32B]	-	-	+	+
<u>S(j)NH-Q(j+3)C<sub>β</sub></u>	[8]	NA	+	NA	+	H(j+6)C <sub>α</sub> -P(j+7)C <sub>β</sub> A	[33A]	+	+	+	+
<u>S(j)NH-Q(j+3)C<sub>γ</sub></u>	[9]	NA	--	NA	+	H(j+6)C <sub>α</sub> -P(j+7)C <sub>β</sub> B	[33B]	+	+	+	+
<u>S(j)C<sub>β</sub>A-Q(j+3)NH</u>	[10A]	-	-	-	+	H(j+6)C <sub>α</sub> -P(j+7)C <sub>γ</sub>	[34]	+	+	+	+
<u>S(j)C<sub>β</sub>B-Q(j+3)NH</u>	[10B]	-	-	-	+	H(j+6)C <sub>α</sub> -P(j+7)C <sub>δ</sub> A	[35A]	++	++	+	+
<u>S(j)C<sub>β</sub>A-Q(j+3)C<sub>β</sub></u>	[11A]	-	-	+	-	H(j+6)C <sub>α</sub> -P(j+7)C <sub>δ</sub> B	[35B]	++	++	+	+
<u>S(j)C<sub>β</sub>B-Q(j+3)C<sub>β</sub></u>	[11B]	NI	+	NI	+	H(j+6)C <sub>β</sub> A-P(j+7)C <sub>δ</sub> A	[36A]	++	++	+	+
<u>S(j)C<sub>β</sub>A-Q(j+3)C<sub>γ</sub></u>	[12A]	-	-	+	+	H(j+6)C <sub>β</sub> A-P(j+7)C <sub>δ</sub> B	[36B]	++	++	+	+
<u>S(j)C<sub>β</sub>B-Q(j+3)C<sub>γ</sub></u>	[12B]	NI	+	NI	+	H(j+6)C <sub>β</sub> B-P(j+7)C <sub>δ</sub> A	[36C]	++	++	+	+
P(j+1)C <sub>α</sub> -N(j+2)NH	[13]	++	++	+	+	H(j+6)C <sub>β</sub> B-P(j+7)C <sub>δ</sub> B	[36D]	++	++	+	+
P(j+1)C <sub>β</sub> A-N(j+2)NH	[14A]	+	+	+	+	P8C <sub>α</sub> -S9NH	[37]	++	NA	+	NA
P(j+1)C <sub>β</sub> B-N(j+2)NH	[14B]	++	++	+	+	P16C <sub>α</sub> -CONH <sub>2</sub> A	[137A]	NA	++	NA	+
P(j+1)C <sub>β</sub> A-N(j+2)C <sub>α</sub>	[15A]	NI	+	NI	+	P16C <sub>α</sub> -CONH <sub>2</sub> B	[137B]	NA	+	NA	+
P(j+1)C <sub>β</sub> B-N(j+2)C <sub>α</sub>	[15B]	NI	+	NI	+	P8C <sub>β</sub> A-S9NH	[38A]	+	NA	+	NA
P(j+1)C <sub>γ</sub> -N(j+2)NH	[16]	++	+	+	+	P8C <sub>β</sub> B-S9NH	[38B]	+	NA	+	NA
P(j+1)C <sub>δ</sub> -N(j+2)NH	[17]	++	+	+	+	P16C <sub>β</sub> A-CONH <sub>2</sub> A	[138A]	NA	+	NA	+
<u>P(j+1)C<sub>α</sub>-Q(j+3)NH</u>	[18]	++	+	+	+	P16C <sub>β</sub> A-CONH <sub>2</sub> B	[138B]	NA	+	NA	+
<u>P(j+1)C<sub>β</sub>-Q(j+3)NH</u>	[19]	-	-	-	+	P16C <sub>β</sub> B-CONH <sub>2</sub> A	[138C]	NA	+	NA	+
N(j+2)NH-Q(j+3)NH	[20]	++	++	+	+	P16C <sub>β</sub> B-CONH <sub>2</sub> B	[138D]	NA	+	NA	+
N(j+2)C <sub>α</sub> -Q(j+3)NH	[21]	++	++	+	+	P16C <sub>γ</sub> -CONH <sub>2</sub> A	[139A]	NA	-	NA	+
N(j+2)C <sub>α</sub> -Q(j+3)C <sub>β</sub>	[22]	+	+	-	+						
N(j+2)C <sub>β</sub> -Q(j+3)NH	[23]	NA	++	+	+						
N(j+2)C <sub>β</sub> A-Q(j+3)NH	[23A]	+	NA	NA	NA						
N(j+2)C <sub>β</sub> B-Q(j+3)NH	[23B]	+	NA	NA	NA						
N(j+2)C <sub>α</sub> -Q(j+4)NH	[25]	+	+	-	+						

Fig. 3. Chemical shift assignments of resonances of peptide HP2, together with NOE connectivities observed in the 2-D NMR spectra. Chemical shift assignments of the major conformation in (a) 90% TFE and (b) 75% TFE are shown. (c) NOE connectivities of the first (I) and second (II) units in 90 and 75% TFE are shown. The connectivities between non-neighbouring residues are underlined. Here the amino acid residues are numbered, for convenience; Ser(j), Asn(j+1) etc., in each unit. The NOE connectivities are numbered so that the numbers of the NOE connectivities of the first and second units become respectively  $n$  and  $100+n$ . NI, not identified; NA, not applicable.



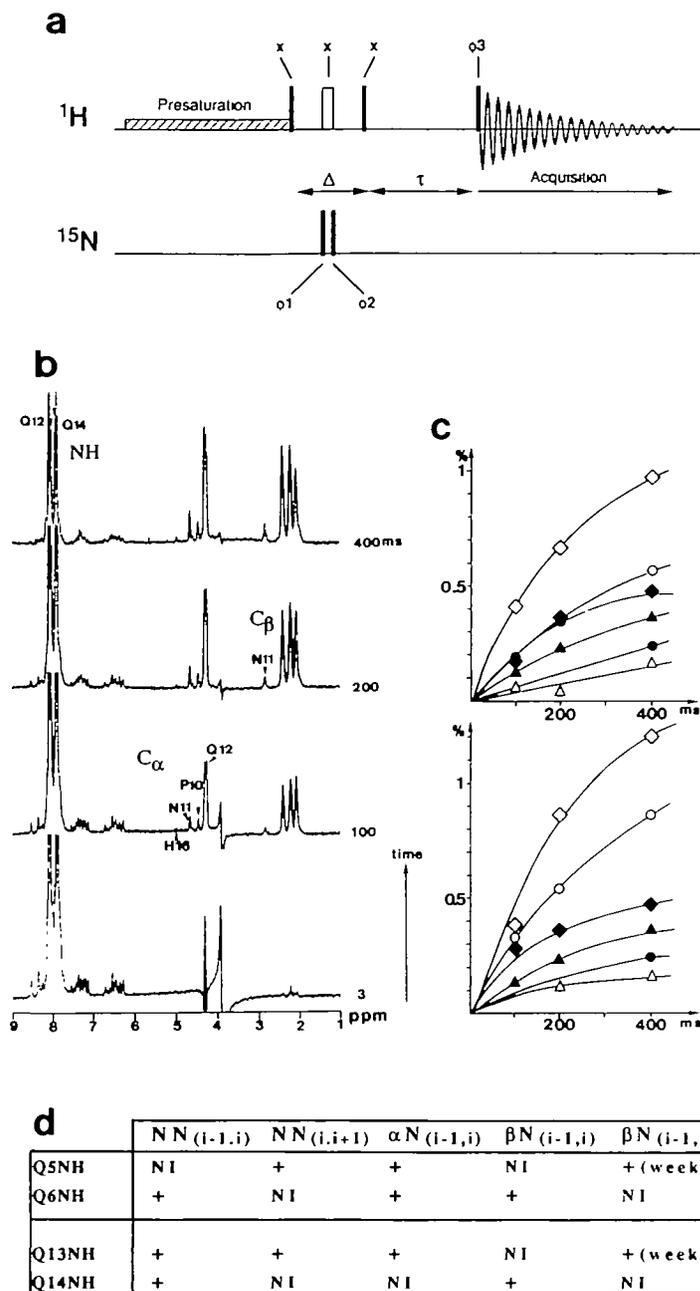
**Fig. 4.** Summary of observed NOE connectivities. (a) NOE connectivities expected for turn structures are shown (A–J). These are for a  $\beta$ -turn (A and B), a  $\beta$ -turn (I) and/or a  $\sigma$ -turn (C–E), for a  $\sigma$ -turn (F and G) and for a  $\beta$ -turn (I) followed by a helical conformation (H–J) (see also c). (b) Hydrogen bonds of the  $\beta$ - and  $\sigma$ -type are shown. (c) NOE connectivities observed for TPRK of T1 peptide in DMSO at 27°C (Suzuki *et al.*, 1993), SPRK of S2 peptide in DMSO at 27°C (Suzuki *et al.*, 1993), SPRK of S2b peptide in 90% DMSO–10% water at 5°C (Suzuki *et al.*, 1993), SPNQ of HP2 at 5°C (this study), SPTS and SPY of YY1 peptide in water (Harding, 1992) and SPNA of LexA in water (Lamerichs *et al.*, 1989) are shown. NOE connectivities of SPNQ shown in parentheses are observed only in 75% TFE, others are observed in 90% TFE. NM, the NOE connectivity was not mentioned in the original report (Lamerichs *et al.*, 1989). (d) Solid lines and dotted lines show respectively NOE connectivities found in 90% TFE and those found only in 75% TFE. For Pro residues, connectivities depicted as involving the NH actually refer to  $C\beta H_2$ . The two NN( $i, i+1$ ) connectivities emphasized by bold lines are the two strongest NOE connectivities among those of this type. The solid lines with stars and those with open circles respectively show comparatively strong and weak NOE connectivities identified by transient NOE enhancement measurements using  $^{15}N$ -labelled peptides.

HP2[Q5], HP2[Q6], HP2[Q13] and HP2[Q14], were measured as a function of NOE mixing time using the heteronuclear spin-echo edited 1-D pulse sequence shown in Figure 5a (in each case, the position of the  $^{15}N$ -label is indicated in square brackets); spectra obtained with this pulse sequence show only those NOE enhancements that originate at the proton directly bonded to the labelled nitrogen atom. An example of the spectra obtained from such a series of NOE experiments is shown in Figure 5b for the peptide HP2[Q13]. It may be seen that the detection level for NOE enhancements using this technique is considerably lower than with conventional NOESY experiments and this raises the difficult problem of interpreting very small enhancements in conformationally mobile systems. Rather than tackle this issue here, we have only used the somewhat larger NOE enhancements in our analysis. Thus, for example, very small enhancements (of  $\sim 0.2\%$  at  $\tau = 400$  ms) are observed from His7  $C_\alpha H$  to Gln5 NH and Gln6 NH (and corresponding

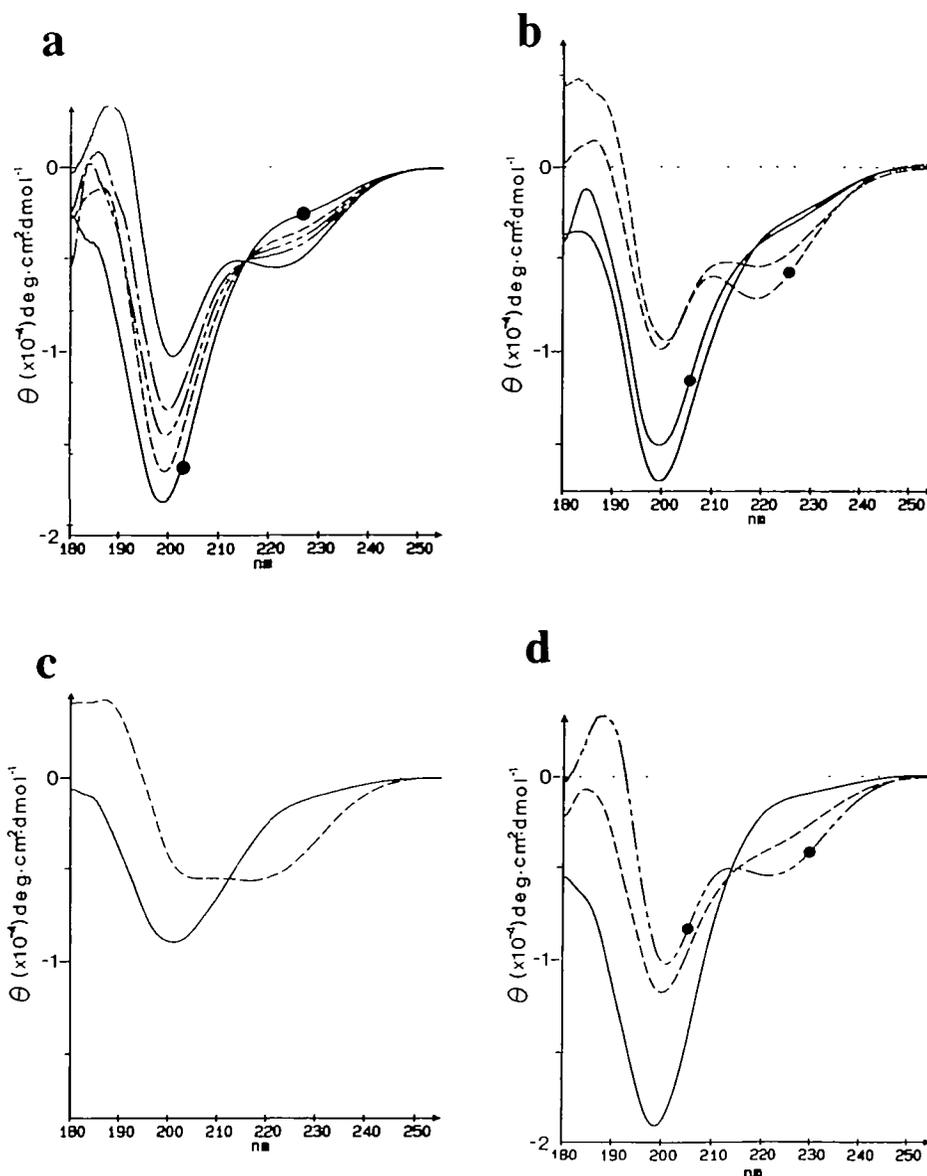
enhancements in unit II), but these interactions are not expected to correspond to short distances in any structure and we believe that they possibly arise through spin-diffusion. Figure 5c shows the time course for the intensities of some of the NOE enhancements in four specifically  $^{15}N$ -labelled peptides.

Several of the sequential connectivities involving the labelled amide positions were identified using the  $^{15}N$ -edited 1-D experiments, including most of the  $d_{\alpha N}(i-1, i)$ ,  $d_{\beta N}(i-1, i)$ ,  $d_{NN}(i-1, i)$  and  $d_{NN}(i, i+1)$  connectivities (where in each case residue  $i$  is the labelled Gln) and these are summarized in Figure 5d. Generally, the enhancements involving Gln5 and Gln13 are stronger than those involving Gln6 and Gln14, suggesting that the latter are less structured.

Some further indication of the extent to which each residue is structured may be obtained by comparing the intensity of the sequential connectivity  $d_{\alpha N}(i-1, i)$  with the intrasidial connectivity  $d_{\alpha N}(i, i)$ . For residues that are in an extended conforma-



**Fig. 5.** Transient NOE enhancement measurements in specifically  $^{15}\text{N}$ -labelled peptides. **(a)** Pulse sequence used to record transient NOE enhancements in specifically  $^{15}\text{N}$ -labelled peptides is shown. Filled vertical bars represent  $90^\circ$  pulses, open vertical bars  $180^\circ$  pulses. The phase cycles used were as follows:  $\phi_1 = x, -x, x, -x$ ,  $\phi_2 = x, x, -x, -x$ ,  $\phi_3 = x, x, x, x, y, y, y, y, -x, -x, -x, -x, -y, -y, -y, -y$  and receiver =  $x, -x, -x, x, y, -y, -y, -y, -x, x, x, -x, -y, y, y$ . This sequence is the 1-D analogue of a doubly half-filtered NOESY experiment (Otting *et al.*, 1986) and also shares features of the heteronuclear spin-echo-edited sequence of Doddrell *et al.* (1983). Proton magnetization coupled to  $^{15}\text{N}$  is first selected by the heteronuclear spin-echo sequence, then rotated onto the  $z$ -axis and NOE enhancements allowed to evolve during the variable period  $\tau$ , before finally rotating the magnetization back into the transverse plane for detection. Only enhancements originating at the proton directly bound to the  $^{15}\text{N}$ -label are detected. The total refocusing delay in the heteronuclear spin-echo ( $\Delta$ ) was set to  $1/|J(^{15}\text{N}, ^1\text{H})| = 5.46$  ms and the NOE mixing times ( $\tau$ ) used were 10  $\mu\text{s}$ , 100, 200 and 400 ms. Presaturation of the  $\text{H}_2\text{O}$  signal was applied at very low power using an irradiation field at the same frequency as and phase-coherent with the  $^1\text{H}$  hard pulses (Zuiderweg *et al.*, 1986). NOE enhancements in the 1-D  $^{15}\text{N}$ -edited NOE spectra were quantified by integrating the enhanced signals and were normalized relative to the signal of the proton directly attached to the  $^{15}\text{N}$ -label, measured at very short (10  $\mu\text{s}$ ) NOE mixing time. **(b)** Spectra for peptide HP2[Q13], with mixing times of 10  $\mu\text{s}$ , 100, 200 and 400 ms. The large signal from the directly  $^{15}\text{N}$ -coupled proton of Gln13 is clipped in this presentation. Other signals in the 10  $\mu\text{s}$  spectrum arise from Gln13  $\text{C}_\alpha\text{H}$  (at 4.31 p.p.m.) and from a low background level of other  $^{15}\text{N}$ -sites, together with a dispersion-mode artefact at the site of the residual water signal ( $\sim 3.9$  p.p.m.). The signal of Gln13  $\text{C}_\alpha\text{H}$  arises partly because this signal has a (small) two-bond  $J$  coupling to the  $^{15}\text{N}$ , so that to a slight extent it is also selected by the heteronuclear spin-echo sequence and partly through the creation of homonuclear zero-quantum coherence between Gln13 NH and Gln13  $\text{C}_\alpha\text{H}$  by the  $90^\circ$  pulse immediately preceding the mixing period  $\tau$ ; this latter coherence is unavoidably retained by the phase cycle. In the longer mixing time spectra, the unwanted contributions from Gln13  $\text{C}_\alpha\text{H}$  and the residual water are largely absent, leaving only the transient NOE enhancements of interest. **(c)** Time course of transient NOE enhancements is shown. The upper panel shows enhancements within unit I from Gln5 NH to Pro2  $\text{C}_\alpha\text{H}$  ( $\circ$ ), Asn3  $\text{C}_\alpha\text{H}$  ( $\diamond$ ) and His7  $\text{C}_\alpha\text{H}$  ( $\Delta$ ) and from Gln6 NH to Pro2  $\text{C}_\alpha\text{H}$  ( $\bullet$ ), Asn3  $\text{C}_\alpha\text{H}$  ( $\blacklozenge$ ) and His7  $\text{C}_\alpha\text{H}$  ( $\blacktriangle$ ). The lower panel shows the corresponding enhancements within unit II; from Gln13 NH to Pro10  $\text{C}_\alpha\text{H}$  ( $\circ$ ), Asn11  $\text{C}_\alpha\text{H}$  ( $\diamond$ ) and His15  $\text{C}_\alpha\text{H}$  ( $\Delta$ ) and from Gln14 NH to Pro10  $\text{C}_\alpha\text{H}$  ( $\bullet$ ), Asn11  $\text{C}_\alpha\text{H}$  ( $\blacklozenge$ ) and His15  $\text{C}_\alpha\text{H}$  ( $\blacktriangle$ ). **(d)** Summary of the relatively strong NOE enhancements found using the labelled peptides HP2[Q5], HP2[Q6], HP2[Q13] and HP2[Q14].



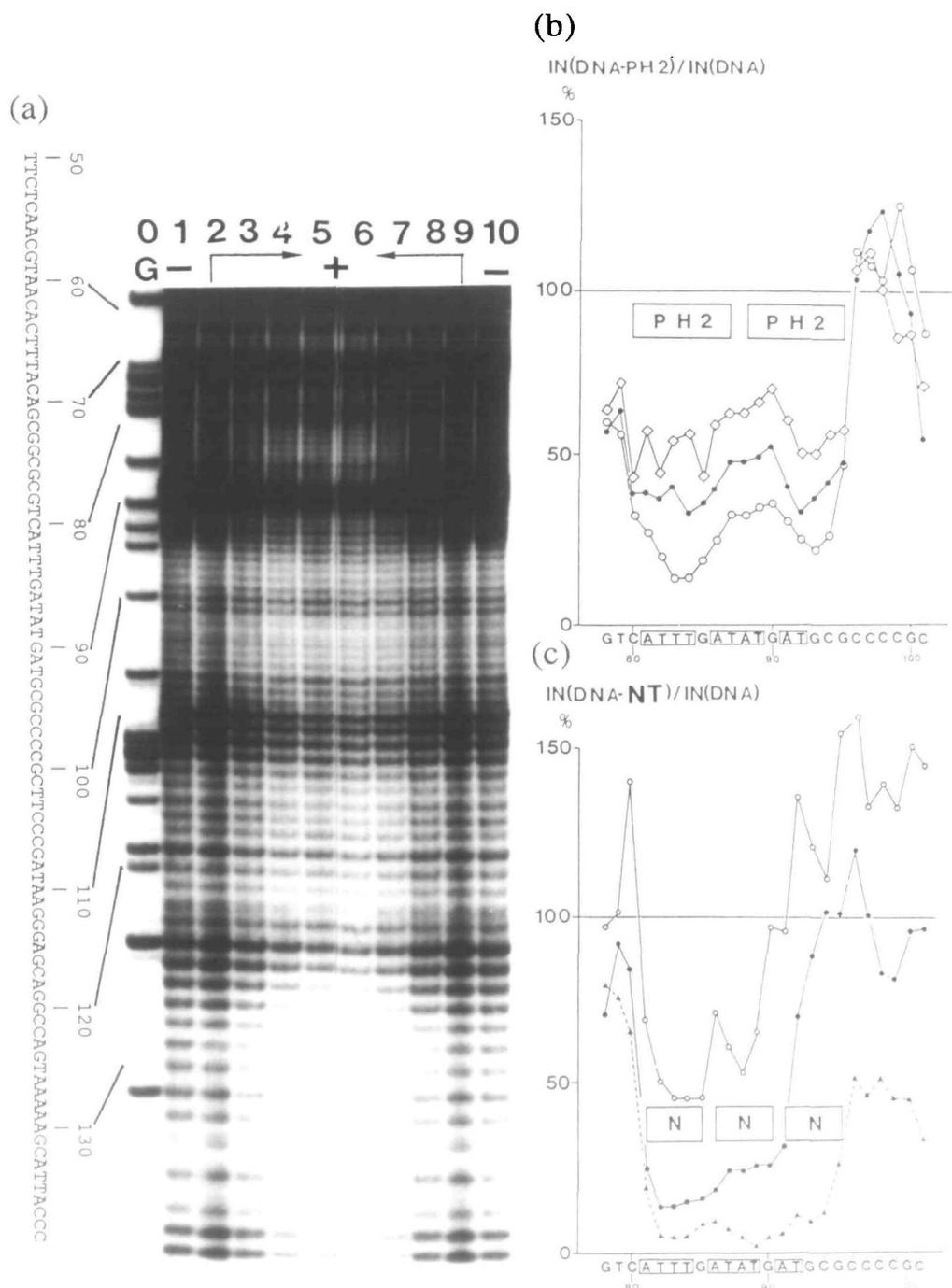
**Fig. 6.** CD spectra of PH2 and HP peptides. (a) The CD spectrum shown as mean residue ellipticity of PH2 at 20°C in the absence (—●—) and presence of 30% (---), 50% (----), 70% (---) and 90% (—) TFE. (b) The CD spectra of HP1, SPNQQQHP (TFE 0%) (—●—), HP1 (TFE 90%) (---●---), HP4, (SPNQQQHP)<sub>4</sub> (TFE 0%) (—), HP4 (TFE 90%) (---), are shown. (c) The CD spectra of QN1, QQQNHPSPN, in the absence (—) and presence of 90% (---) of TFE are shown. (d) The CD spectra of S4, G-(SP[K/R][K/R])<sub>4</sub> in the absence (—) and presence of 90% (---) of TFE are shown. For comparison the CD spectrum of PH2 (TFE 90%) (---●---) is also shown.

tion, the sequential  $d_{\alpha N}(i-1, i)$  connectivity is expected to be significantly stronger than the intraresidual  $d_{\alpha N}(i, i)$  connectivity (Wüthrich, 1986), whereas in a turn or helical conformation this ratio may be reversed. However, the sequential distance  $d_{\alpha N}$  is so much shorter than the intraresidual one for an extended conformation that the intensity ratio is not reversed even in cases where an extended or random coil form of a peptide is in fast exchange with a structured state involving a turn, as was seen for the peptide SPRKSPRK [see Figure 3 of Suzuki *et al.* (1993)]. Using this intensity ratio as an approximate guide, it may be seen that Asn11 and Gln12 are more structured than Gln13 or His15; thus Asn11 C $_{\alpha}$ H–Asn11 NH is as strong as Pro10 C $_{\alpha}$ H–Asn11 NH and Gln12 C $_{\alpha}$ H–Gln12 NH is as strong as Asn11 C $_{\alpha}$ H–Gln12 NH, but Gln13 C $_{\alpha}$ H–Gln13 NH is weaker than Gln12 C $_{\alpha}$ H–Gln13 NH and His15 C $_{\alpha}$ H–His15 NH is weaker than Gln14 C $_{\alpha}$ H–His15 NH.

Thus, in summary, the HP2 peptide folds into two structural units in aqueous TFE solution, within each of which the sequence Ser-Pro-Asn-Gln forms a stable  $\beta\sigma$ -turn, while the remaining residues show some nascent helix characteristics. The observed NOE connectivities are summarized in Figure 3c.

#### CD spectra of the PH2 and HP peptides

The CD measurements, in conjunction with the NMR studies described above, lead to three conclusions in particular. Firstly, the PH2 peptide and the HP peptides (Figure 6) do not have totally 'random' or extended conformations even in the absence of TFE. The CD spectrum of PH2 peptide in aqueous solution (Figure 6a) is characterized by two negative peaks, one at  $\sim 200$  nm, which reflects an extended state and the other at  $\sim 222$  nm, which reflects the presence of a helical conformation (Greenfield and Fasman, 1969). On adding increasing amounts of trifluoroethanol



**Fig. 7.** Cu(II)-*o*-phenanthroline footprintings. (a) Cu(II)-*o*-phenanthroline reaction on a 160 bp DNA fragment (tyrT) in the absence (lanes 1 and 10) and presence (lanes 2–9) of PH2 peptide. The tyrT concentration was kept to  $9.6 \times 10^{-9}$  M. The PH2 concentration used is  $3.2 \times 10^{-4}$  M (lanes 2 and 9),  $6.8 \times 10^{-4}$  M (lanes 3 and 8),  $1.3 \times 10^{-3}$  M (lanes 4 and 7) and  $1.9 \times 10^{-3}$  M (lanes 5 and 6). G denotes the Maxam–Gilbert guanine-specific marker reaction in lane 0. (b) The ratio of band intensity in the presence of PH2, to that of free DNA, at an A/T-rich sequence. The PH2 concentration used is  $6.8 \times 10^{-4}$  M ( $\diamond$ ),  $1.3 \times 10^{-3}$  M ( $\bullet$ ) and  $1.9 \times 10^{-3}$  M ( $\circ$ ). Boxes show the peptide-binding sites. (c) The protection profile for the same sequence as (b) produced by netropsin:  $1.2 \times 10^{-7}$  M ( $\circ$ ),  $2.4 \times 10^{-7}$  M ( $\bullet$ ) and  $2.4 \times 10^{-6}$  M ( $\blacktriangle$ ). As the concentration is increased, the extent of the footprint increases consistent with three molecules successively covering five base pairs each, as is found in the crystal structure.

(TFE), which is known to stabilize peptide secondary structure, the mean residue ellipticity at 222 nm ( $\theta_{222}$ ) becomes larger, while that at 200 nm becomes smaller, as is expected for a coil-to-helix transition (Greenfield and Fasman, 1969). The PH2 peptide is, to some extent, folded even in the absence of TFE, since its CD spectrum in water is different from that of a 'random coil'; a positive peak at  $\sim 220$  nm is expected for a 'random coil' (Greenfield and Fasman, 1969), while the CD of PH2 and

220 nm is negative, indicating that it is somewhat helical. The CD spectra of HP2 are essentially the same as those of PH2 (data not shown).

Secondly, even the longest peptide HP4, i.e. (SPNQQHP)<sub>4</sub>, folds into blocks rather than forming a continuous helix. The CD spectra of peptides containing one (HP1) and four (HP4) repeats are similar to that of PH2, suggesting that these three peptides are structured to an approximately similar extent (Figure 6b).

This contrasts with the case of a continuous  $\alpha$ -helical structure, which becomes more stable the larger the number of amino acid residues involved (Zimm and Bragg, 1959). This effect occurs not only for long polymers but for short peptides as well (Goodman *et al.*, 1969).

Thirdly, the HPSP junction, about which little could be deduced from the NMR study, is not fully 'random' but likely to be in a structured state in the presence of TFE. The CD spectrum of QN1, QQQHPSPN, which has no physical connectivity between SPN and QQQH, is changed upon adding increasing amounts of TFE (Figure 6c). The QN1 peptide lacks the N-terminal core and indeed its CD is different from that of the peptides PH2 and HP's and does not correspond to random coil in TFE. Therefore, the HPSP junction itself has a tendency to fold into a particular structure, which probably is largely restricted by the two Pro residues.

#### *DNA binding of the K10 repeat: a footprinting study*

The peptide PH2 (that is, P-SPNQQQHP-SPNQQQH) was used for a DNA-binding study. The CD spectra of PH2 and HP2 are essentially the same, suggesting that the two structures are also very similar. PH2 has been shown to bind to DNA by its ability to protect a 160 bp DNA fragment (TyrT) from cleavage by Cu(II)-*o*-phenanthroline (Figure 7a and b). To our knowledge, this is the first demonstration of DNA binding of the *K10* gene product. The aim of the present DNA binding study is to determine the number of base pairs bound by one unit in order to help the modelling of peptide–DNA interaction. The phenanthroline reagent was used because the regions of protection observed are much better defined than those obtained using DNase I or Fe-EDTA. As a demonstration of this, we applied the technique to analysis of the DNA–netropsin complex. It has been proved that netropsin binds to four base pairs directly and covers five or six base pairs (Kopka *et al.*, 1985). Our footprinting results on titrating with netropsin show that 15 bp are covered by three molecules and, thus, five base pairs by one molecule (Figure 7c).

The sequence specificity of the PH2 peptide is weaker than that of netropsin and the footprint of PH2 is less striking, but it is still clear enough to conclude that the PH2 peptide mainly binds to two A/T-rich sequences, 5'-C(80)ATTTGATATG(90)ATGCG-3' and 5'-GTAAAAA(130)GC-3' (one of which is shown in Figure 7b), but does not bind to G/C-rich sequences well (see the right-hand side of the protection profile in Figure 7b). We conclude that one PH2 binding site seems to be eight or nine base pairs long (Figure 7b), so that one unit covers approximately four base pairs.

By comparing the concentration of PH2 necessary to give 50% protection at an A/T-rich sequence, residues 119–137, with those of other DNA binding ligands, netropsin (Figure 7c) and Hoechst 33258 (data not shown), the DNA binding constant of PH2 is estimated as  $2.0 \times 10^5 \text{ M}^{-1}$ . This value may be an underestimate, as the competition with Cu-phenanthroline could be more serious when the DNA binding constant of a ligand is smaller. Furthermore, the specificity of the peptide is not as high as that of the drugs and, thus, the peptide also binds to the rest of the sequence in the DNA fragment, which is ignored in this estimation. Since the whole *K10* repeat contains eight units and since it starts with the SPKK sequence, of which the DNA binding constant is higher than that of PH2 (Suzuki, 1989), the DNA binding constant of the whole *K10* repeat is expected to be considerably higher than that estimated for PH2.

## Discussion

It is known that some peptides which are not part of a hydrophobic core become helical only upon binding to DNA [see for example, a discussion on basic domain–leucine zipper proteins by Weiss *et al.* (1990)], whereupon their basic residues adopt conformations suitable for interactions with phosphates of the DNA backbone. It therefore seems reasonable to propose that upon binding to DNA, the three weakly helical Gln residues of each unit in HP2 would become more structured; Gln can interact with phosphates of DNA. Thus, the *K10* repeat is likely to fold into a helical conformation broken into segments by the HPSP connections upon DNA binding. This idea seems to be supported by our footprinting experiments, because a model of SPNQQQHP in a short helical segment (Figure 1b) is slightly shorter than netropsin, in agreement with the fact that netropsin spans five base pairs of DNA while SPNQQQHP spans four.

Such a turn structure adopted by SPXX at the N-terminus of a helical conformation is observed in some proteins. Among the 24 SPX3X4 sequences (where X3 is not Gly) recorded at high resolution in non-identical protein structures in the Protein Data Bank (Brookhaven), ten examples have either of two hydrogen bonds,  $\beta$ - or  $\sigma$ -type and six of these ten are located at the N-terminus of a helix. A serine residue located at the N-terminus of an  $\alpha$ -helix can alter the hydrogen bond profile by making a hydrogen bond from its side chain OH to the main chain NH of the residue  $i+2$ , thereby terminating the  $\alpha$ -helix (N-cap, Richardson and Richardson, 1988). When placed at the N-terminus of a helix, a  $\beta\sigma$ -turn made up by the SPXX sequence can cap the N-terminus perfectly, since the NH groups of residues  $i+2$  and  $i+3$  are capped by the  $\sigma$ -type hydrogen bond and the  $\beta$ -turn-type hydrogen bond respectively, and the Pro of course has no NH.

Furthermore, the HPSP junction can cap the C-terminus of the preceding unit (C-cap). If the SP sequence of HPSP folds into a  $\sigma$ -type  $\beta$ -turn, statistics suggest that His-Pro-Ser should adopt a  $\beta$ -strand-like extended conformation because of the restrictions imposed by two prolines (M. Macarthur and J.M. Thornton, personal communication). The dihedral angles of HPSP are likely to be  $\beta$ - $\beta$ - $\beta$ - $\alpha$ , which actually coincides with the NMR observation that the –His7-Pro8– connection is extended, that is, the NOE connectivity Pro8 C $\alpha$ H–His7 NH is stronger than His7 C $\alpha$ H–His7 NH and the NOE connectivity Pro8 C $\alpha$ H–Ser9 NH is stronger than Ser9 C $\alpha$ H–Ser9 NH.

Models of one unit and two units are made so that they fulfil the NMR measurements and the above discussions (Figure 1b and c). If a long helical structure were to bind to DNA, it would have to be kinked and separated into short blocks (Warrant and Kim, 1978) in order to follow the groove of the DNA (Figure 1d). The model shown in Figure 1c has enough degrees of freedom to follow the groove of DNA.

The model described in this paper for the *K10* repeat might have a further relevance. The helix–turn–helix protein LexA has a DNA recognition helix starting with the sequence SPNA (Lamerichs *et al.*, 1989), resembling the sequence of SPNQ of HP2 (see the summary of NMR studies in Figure 4c). Although the N-cap of the recognition helix of LexA is not well characterized, NOE connectivities between Ser and Asn were observed, suggesting that a structure similar to that described for SPNQ in this paper, is adopted. The SPKK motifs present in the C-terminus of histone H1 are separated by some amino acid residues [see the histone sequences listed in von Holt *et al.* (1989)]. The most frequent sequence among them, SPKKAKKP (Suzuki, 1992) starts with SP and ends with Pro at the eighth

position as in the SPNQQQHP sequence. The synthetic peptide, KSPKKAKKP, shows a CD spectrum similar to that of PH2 (unpublished results). Two groups have proposed different secondary structures for the H1 C-terminus, namely an  $\alpha$ -helix (Clark *et al.*, 1988) and a  $\beta$ -turn I (Erard *et al.*, 1990), but we suggest that the region folds into a combination of both upon binding to DNA.

### Acknowledgements

We thank Professor A.Klug for his reading of the manuscript and critical comments. We thank Dr C.Chothia for his help with the statistical study and Professor J.M.Thornton for helpful discussion. The work of S.A. was supported by the Ministry of Education, Culture, and Science, Japan.

### References

- Brooks,B.R., Bruccoleri,R.E., Olafson,B.D., States,D.J., Swaminathan,S. and Karplus,M. (1983) *J. Comput. Chem.*, **4**, 187–217.
- Brunger,A.T., Kuriyan,J. and Karplus,M. (1987) *Science*, **235**, 458–460.
- Churchill,M.E.A. and Suzuki,M. (1989) *EMBO J.*, **8**, 4189–4195.
- Clark,D.J., Hill,C.S., Martin,S.R. and Thomas,J.O. (1988) *EMBO J.*, **7**, 69–75.
- Dayringer,H.E., Tramontano,A., Sprang,S.R. and Fletterick,R.J. (1986) *J. Mol. Graph.*, **4**, 82–87.
- Doddrell,D.M., Williams,D.H., Reid,D.G., Fox,K.R. and Waving,J.M. (1983) *J.C.S. Chem. Commun.*, **1**, 218–220.
- Drew,H.R. and Travers,A.A. (1984) *Cell*, **37**, 491–502.
- Dyson,H.J., Rance,M., Houghten,R.A., Wright,P.E. and Lerner,R.A. (1988) *J. Mol. Biol.*, **201**, 201–207.
- Erard,M., Lakhdra-Ghazal,F. and Amalric,F. (1990) *Eur. J. Biochem.*, **191**, 19–26.
- Goodman,M., Verdini,A.S., Toniolo,C., Phillips,W.D. and Bovey,F.A. (1969) *Proc. Natl Acad. Sci. USA*, **64**, 444–450.
- Greenfield,N. and Fasman,G.D. (1969) *Biochemistry*, **8**, 4108–4116.
- Harding,M.M. (1992) *J. Med. Chem.*, **35**, 4658–4664.
- Jones,T.A. (1982) In Sayer,D. (ed.), *Computational Crystallography*. Clarendon Press, Oxford, pp. 303–317.
- Kopka,M.L., Pjura,P.E., Goodsell,D.S. and Dickerson,R.E. (1985) *Proc. Natl Acad. Sci. USA*, **82**, 1376–1380.
- Lamerichs,R.M.J.N., Padilla,A., Boelens,R., Kaptein,R., Otteben,G., Rüterjans,H., Granger-Schnarr,M., Oertel,P. and Schman,M. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 6863–6867.
- Lesk,A.M. and Hardman,K.D. (1982) *Science*, **216**, 537–540.
- Otting,G., Senn,H., Wagner,G. and Wüthrich,K. (1986) *J. Magn. Resonance*, **70**, 500–505.
- Prost,E., Deryckere,F., Roos,C., Haenlin,M., Pantesco,V. and Mohier,E. (1988) *Genes Dev.*, **2**, 891–900.
- Richardson,I.S. and Richardson,D.C. (1988) *Science*, **240**, 1648–1652.
- Smith,J.M. and Thomas,D.J. (1990) *Comput. Appl. Biol. Sci.*, **6**, 93–99.
- Spassky,A. and Sigman,D.S. (1985) *Biochemistry*, **24**, 8050–8056.
- Suzuki,M. (1989) *EMBO J.*, **8**, 797–804.
- Suzuki,M. (1991) In Eckstein,F. and Lilley,D.M.J. (eds), *Nucleic Acids and Molecular Biology 5*. Springer-Verlag, Berlin, pp. 126–140.
- Suzuki,M. (1992) In Kyogoku,Y. and Nishimura,Y. (eds), *Molecular Structures and Life*. Japan Science Society Press, Tokyo/CRC Press, Boca Raton, FL, pp. 157–170.
- Suzuki,M. and Yagi,N. (1991) *Proc. R. Soc. Lond.*, **B246**, 231–235.
- Suzuki,M., Gerstein,M. and Johnson,T. (1993) *Protein Engng.*, **6**, 565–574.
- von Holt,C., Brandt,W.F., Greyling,H.J., Lindsey,G.G., Retief,J.D., Rodrigues,J.De A., Schwager,S. and Sewell,B.T. (1989) *Methods Enzymol.*, **170**, 431–523.
- Warrant,R.W. and Kim,S.-H. (1978) *Nature*, **271**, 130–135.
- Weiss,M.A., Ellenberger,T.E., Wobre,C.R., Lee,J.P., Harrison,S.C. and Struhl,K. (1990) *Nature*, **347**, 575–578.
- Wüthrich,K. (1986) *NMR of Proteins and Nucleic Acids*. Wiley-Interscience Publications, New York.
- Zimm,B.H. and Bragg,J.K. (1959) *J. Chem. Phys.*, **31**, 526–535.
- Zuiderweg,E.R.P., Hallenga,K. and Olejniczak,E.T. (1986) *J. Magn. Resonance*, **70**, 336–343.

Received June 28, 1993; revised November 4, 1993; accepted December 22, 1993