Volume 309, number 3, 385–388 FEBS 11540 September 1992


# Synthesis of a large peptide mimicking the DNA binding properties of the sex determining protein, SRY

F. Poulat, G. Guichard, C. Gozé, F. Heitz, B. Calas and P. Berta

*Centre de Recherches de Biochimie Macromoléculaire, CNRS/INSERM, 1919 route de Mende, BP 5051, 34033 Montpellier Cédex, France*

Received 10 July 1992; revised version received 30 July 1992

The sex determining protein, SRY, has been recently described as containing a DNA binding motif, also called the SRY box. This 80 amino acid box was synthesized using the continuous flow solid-phase technique. The product was then purified and tested according to such diverse criteria as its intrinsic structure or its biological activity (DNA binding capacity), and compared to the full-length protein. The data indicate that the peptide is relevant for the properties described so far for the protein.

Sex determining protein; Solid-phase peptide synthesis; Peptide–nucleotide interaction

## 1. INTRODUCTION

The fate of a human embryo to become male or female is under the genetic control of the Y chromosome. Years of speculation about the nature of the signal, located onto the Y chromosome and able to switch a 6-week-old embryo towards a male development, have ended recently [1,2]. A gene called SRY (sex determining region Y) in human was cloned and characterized using various methods. The first was genetic and came from the study of sex reversal in XY females, some of which were shown to result from de novo mutations in SRY [3,4]. Another one came from the ability to produce male transgenic mice on an XX genetical background [5]. The sequence analysis of the SRY gene revealed a striking homology to proteins described in the literature as candidates for DNA binding proteins and even transcription factors [1]. In fact, this homology is restricted to an 80 amino acid sequence motif first recognized in HMG (high mobility group) proteins and named the HMG box [6]. Although the list of proteins belonging to this class of homology is long, some of them, like TCF-1, a T-cell specific DNA binding protein [7], can be distinguished on the basis of their similarity to the DNA binding domains of SRY. These SRY box-containing proteins have been subclassed as SOX proteins (SRY box-containing genes). To analyse the molecular basis of the interaction between the protein, SRY, and DNA, the synthesis of an 80 amino acid peptide (SRY80) was performed using the Fmoc-polyamide strategy in conjunction with an automatic continuous flow apparatus [8], and its DNA binding properties were compared with the bacterially produced full-length protein.

*Correspondence address:* P. Berta, Centre de Recherches de Biochimie Macromoléculaire, CNRS/INSERM, 1919 route de Mende, BP 5051, 34033 Montpellier Cédex, France. Fax: (33) 67 52 1559.

*Abbreviations:* SRY, sex determining region Y; Fmoc, 9-fluorenylmethoxycarbonyl; DMF, *N,N*-dimethylformamide; DCM, dichloromethane; TFA, trifluoroacetic acid; HPCE, high performance capillary electrophoresis; TBE, Tris-borate EDTA; TCF 1, T cell-specific factor 1.

## 2. MATERIALS AND METHODS

### 2.1. *Peptide synthesis*

The sequence of SRY80 is: D[1]RVKRPMNAFIVWSRDQRRKM[21]ALENPRMRNSEISKQLGYQW[41]KMLTEAEKWPFFQEAQKLQA[61]MHREKYPNYKYRPRRKAKM[80].

Continuous flow solid-phase peptide synthesis was carried out automatically using a Milligen 9050 PepSynthesizer. The flow rate of the unit pump was set at 4.6 ml/min. The incorporation of amino acids was performed using systematic double coupling. The following reaction schedule was used: Fmoc (9-fluorenylmethoxycarbonyl) group deprotection with piperidine-*N,N*-dimethylformamide (DMF) (3:7) (7 min), DMF washing (12 min), first amino acid coupling (30 min), DMF washing (4 min), dichloromethane (DCM) washing (4 min), DMF washing (12 min), second amino acid coupling (30 min), DMF washing (4 min), DCM washing (4 min), DMF washing (12 min). The progress of the synthesis was monitored by continuous UV detection of the reaction column effluent. Synthesis was carried out using 1 g of Fmoc-Met-Pepsyn KA (0.092 mMol Met/g, from Milligen). The protected amino acids (three-fold excess) were coupled in the presence of 2-(1*H*-benzotriazolyl-1-yl)-1,1,3,3-tetramethyluronium tetrafluoroborate (TBTU, three-fold excess), diisopropylethylamine (three-fold excess) and 1-hydroxybenzotriazole (HOBt, three-fold excess). Side-chain protection was as follows: 2,2,5,7,8-pentamethylchroman-6-sulfonyl (PMC) for Arg, Trityl (Trt) for His, tert-butyloxycarbonyl (Boc) for Lys, tert-butyl (t-Bu) ether for Ser and Thr and t-Bu ester for Asp and Glu. Synthesis was completed with the Nα deprotection of the final residue (piperidine-DMF 3:7 7 min, DMF 12 min, and DCM 15 min). The peptide resin was removed from the column, washed with

isopropanol, acetic acid, isopropanol, diethylether and dried in vacuo. The protected peptide resin was stirred for 12 h with 85% trifluoroacetic acid (TFA), 5% phenol, 2.5% thioanisole, 2.5% ethanedithiol and 2.5% water (50 ml) at room temperature. The resin was filtered, washed with TFA and the combined filtrates rotatory evaporated without heating. 100 ml of cold diethylether was added to the oily residue. The white precipitate was collected by centrifugation, redissolved in the minimum of neat TFA and re-precipitated with cold diethylether. The crude peptide was dried in vacuo over KOH pellets for 12 h, dissolved in 0.1% ammonium bicarbonate (pH 8.5) and extensively dialyzed against 0.1% ammonium bicarbonate (pH 8.5). After freeze drying, the peptide was purified by semi-preparative HPLC using a Waters system (Waters model 510 high pressure pumps, Waters automated gradient controller model 680, Waters 486 tunable absorbance detector) with Delta Pak C18 300Å (19 × 300 mm). The mobile phase consisted of solution A (0.1% TFA in water) and solution B (0.08% TFA in acetonitrile).

For amino acid analyses, two samples of peptide were hydrolyzed for 24 and 48 h at 110°C in 6 N HCl and in 4 M methanesulfonic acid in evacuated sealed tubes. The analyses were performed on a 6300 Beckman apparatus. The composition of amino acids is in good agreement with the expected structure: Asn 5.26 (6), Thr 1 (1), Ser 2.45 (3), Glu 13.3 (13), Pro 5.27 (5), Gly 1:13 (1), Ala 6.52 (6), Val 1.15 (2), Met 6.09 (6), Ile 1.46 (2), Leu 3.2 (3), Tyr 4.97 (4), Phe 2.96 (3), His 1.22 (1), Lys 10.59 (10), Arg 11.5 (11), Trp 2.6 (3).

High performance capillary electrophoresis (HPCE) were carried out using a P/ACE System 2000 from Beckman. Mass spectra were carried out on a Lasermat from Finnigan Mat. SDS gel electrophoresis was performed on 20% (w/v) polyacrylamide gels using the buffer system of Laemmli.

### 2.2. *Circular dichroism*

CD spectra were recorded on a Jobin-Yvon Mark V dichrograph. Peptide concentration was 0.1 mg/ml, cell path 1 mm in $PO_4H_2Na$ 10 mM, pH 7.4, NaCl 0.1 M or 0.01 mM.

### 2.3. *Gel shift assay*

Annealed oligonucleotides were labelled by Klenow polymerase in a fill-in reaction with [α-$^{32}$P]dCTP. All probes were purified by polyacrylamide gel electrophoresis. For a typical binding reaction, peptide (1 μg), 1 μg poly(dI.dC) and 200 μg of salmon sperm DNA were incubated in a final volume of 20 μl containing 10 mM HEPES, pH 7.9, 60 mM KCl, 50 mM NaCl, 0.1 mM $MgCl_2$ and 4 mM spermidine. After 5 min of pre-incubation at room temperature, a probe (5,000–7,000 cpm) was added and the mixture incubated at room temperature for an additional 20 min. The samples were then subjected to electrophoresis through a non-denaturating 9% polyacrylamide gel run in 0.5× TBE (Tris-borate EDTA) at 4°C. Probes used were:

```
TCF1-3  5'  GGAGACTGAGAACAAAGCGCTCTCA
                CTGACTCTTGTTTCGCGAGAGTAGCT 5'
MUT1-2  5'  GGAGACTGAGACCAAAGCGCTCTCA
                CTGACTCTGGTTTCGCGAGAGTAGCT 5'
MUT3-4  5'  GGAGACTGAGCACAAAGCGCTCTCA
                CTGACTCGTGTTTCGCGAGAGTAGCT 5'
UNT1-2  5'  AGCTGCCACCGCCCCCGCCAGGAGG
                CGGTGGCGGGGGCGGTCCTCCCCTAG 5'
```

## 3. RESULTS AND DISCUSSION

The Fmoc-polyamide strategy used in conjunction with an automatic continuous flow apparatus [8], appears as a very effective method of making large peptides. This is first due to the use of polar gels (Pepsyn K) freely permeated and solvated by dipolar aprotic media such as DMF. In this solvent, PepsynK matrix

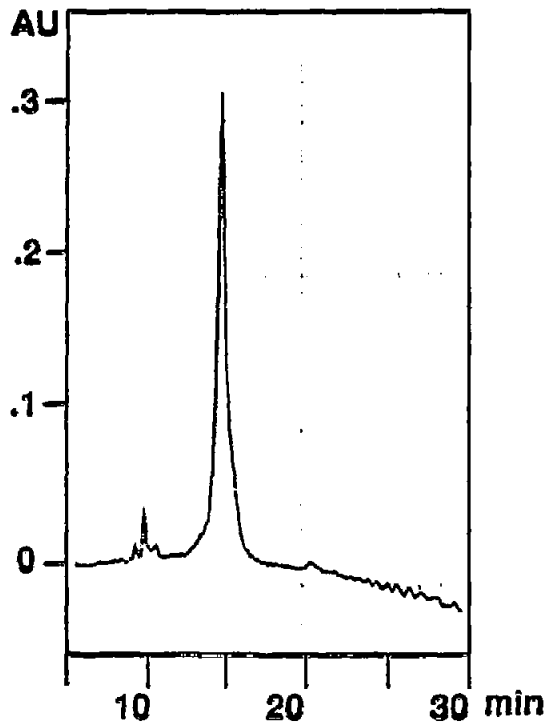


Fig. 1. HPLC of purified SRY80. Column Lichrospher 60 (Merck, Darmstadt; 11 × 0.46 cm). Buffer A, 0.1% TFA in water; buffer B, 0.08% TFA in acetonitrile. Linear gradient, 0–90% B in 30 min. Detection at 214 nm. Flow, 1 ml/min.

is essentially 'transparent' to the reactant molecules; thus, the effect of the polymer on the various chemical processes is minimized, and aggregation within the resin substantially reduced. Second, the use of the Fmoc group allows repetitive deprotection to be performed under extremely mild conditions: 20 or 30% piperidine in DMF. In addition, with 4-alkoxybenzyl linkage agents, the cleavage of the peptide from the support is carried out under mild acidic conditions (trifluoroacetic acid containing various scavengers) which also remove side-chain protections. The preparation of large peptides is also facilitated by the use of the continuous flow apparatus which offers many advantages over discontinuous systems: (i) efficient washings and removal of excess of reagent, (ii) economy of solvent, (iii) continuous analytical control of reaction progress, and (iiii) complete automation.

After purification by HPLC, amino acid analyses of SRY80 revealed the correct amino acid composition (see Materials and Methods) and its purity was further checked by HPLC and HPCE (Fig. 1 and 2). SDS-polyacrylamide gel electrophoresis revealed the expected molecular weight of about 10,000 Da (data not shown) which compares with 10,033 calculated on the basis of the sequence. Furthermore, the mass spectra is in good agreement with the expected structure (SRY80

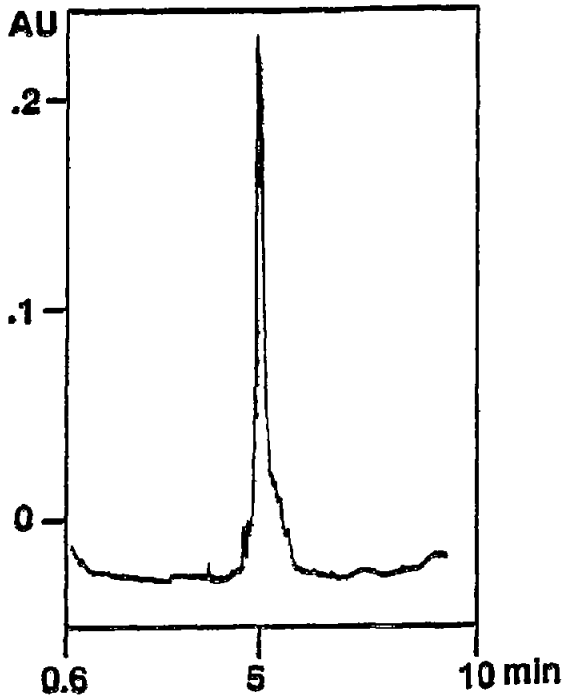


Fig. 2. HPCE of purified SRY80. Capillary, fused silica 57 cm × 75 μm i.d. Buffer, borate 0.1 M, pH 2.5. Temperature, 23°C. Detector, 214 nm. Injection, 5 s. Separation: constant voltage, 30 kV.

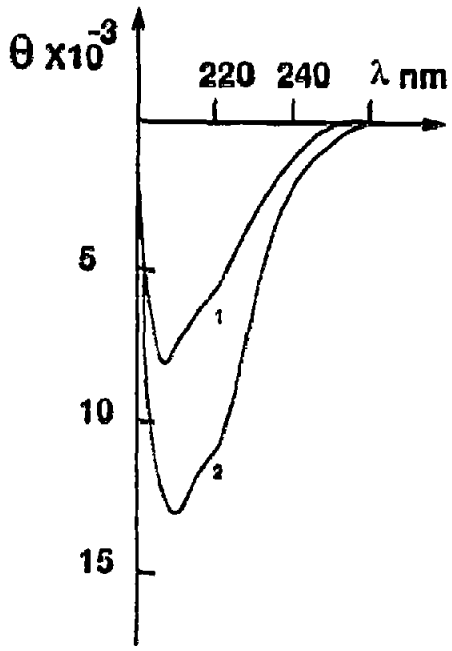


Fig. 3. CD spectra of SRY80 in (trace 1) NaCl 0.01 M, $PO_4H_2Na$ 10 mM, pH 7.4, and (trace 2) NaCl 0.1 M, $PO_4H_2Na$ 10 mM, pH 7.4. $\theta$ is the ellipticity in deg·cm$^2$·mole$^{-1}$ per residue.

measured mass = 10,051, SRY80 expected mass = 10,033).

The CD spectrum of SRY80 (Fig. 3, trace 1) is characterized by a minimum at 207 nm with a shoulder at 221 nm. This type of spectrum indicates a tendency for this peptide to adopt, at least partially, an helical conformation. It must be noted that a decrease in ionic strength decreases the helical content of SRY80 (see Fig. 3, trace 2). This suggests that hydrophobic interactions participate in the stabilisation of the helical structure. Since on the basis of CD spectroscopy it is not possible to define the exact location of the helical regions, especially the length of the perturbation induced by the Pro residue, we are not able to decide which model [9] accounts for the interaction of this peptide with nucleotides.

The interaction of SRY80 to DNA was studied using various oligonucleotide probes. The choice of oligonucleotide sequence is based on a recent report [8] which describes the ability of the SRY protein to bind to the TCF1 (Tcell-specific factor 1) target sequence, AACAAAG. A strong binding of the SRY80 to the probe containing this motif is observed (Fig. 4). Its specificity is assessed using either an unrelated oligonucleotide as control (UNT1-2), or two mutant versions (MUT1-2 and MUT3-4) of the TCF1 recognition motif. As previously described for the full-length protein [10], no binding is observed for the mutant version of TCF1 (MUT1-2) where the second A is replaced with a C, and a partial binding is maintained when the first A is replaced with a C (MUT3-4). It is worth noting that a shorter version of the peptide lacking the first sixteen amino acids loses its ability to interact with the DNA motif.

The present report describes the synthesis and the analysis of an 80 amino acid peptide (SRY80) covering the DNA binding domain of the sex determining protein, SRY. Comparison with the full-length protein reveals a similar ability for the SRY80 to bind to the TCF1 recognition motif, and thus provides us with a good test of activity. This confirms the quality of the

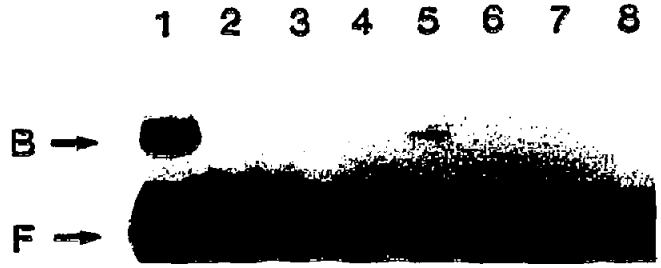


Fig. 4. DNA binding of the peptides and sequence specificity (the 80 amino acids correspond to the 58–138 sequence of the SRY protein). SRY80 or the 64 amino acid peptide corresponding to the 74–138 sequence of the SRY protein were incubated with the TCF1-3 oligonucleotide (lanes 1 and 2, respectively), the MUT1-2 oligonucleotide (lanes 3 and 4), the MUT3-4 oligonucleotide (lanes 5 and 6) and with the unrelated oligonucleotide, UNT1-2 (lanes 7 and 8), according to the protocol described in Materials and Methods. The position of the bound (B) or unbound, free DNA probe (F) are indicated.

peptide as it is known that single amino acid mutations in this sequence abolishes DNA binding ability [10,11]. Furthermore, this result indicates clearly that this 80 amino acid region of the 204 amino acid protein, SRY, is sufficient to display a specific DNA binding. Such a peptide should provide a pure and abundant material source for biophysical investigation (1D or 2D NMR studies, fluorescence spectroscopy, etc.) and the study of its interaction with DNA. Furthermore, SRY80 is a valuable tool to test, without a priori, if the TCF1 recognition motif forms a DNA binding site for the SRY protein.

*Acknowledgements:* This research was supported by an 'Association Française contre les Myopathies' grant to P.B. We thank P. Eustaquio for typing the manuscript Dr. N. Lamb for helpful comments and discussions.

## REFERENCES

[1] Sinclair, A.H., Berta, P., Palmer, M.S., Hawkins, J.R., Griffiths, B.L., Smith, M.J., Foster, J.W., Frischauf, A.M., Lovell-Badge, R. and Goodfellow, P.N. (1990) Nature 346, 240–244.
[2] Gubbay, J., Collingnon, J., Koopman, P., Capel, B., Economou, A., Munsterberg, A., Vivian, N., Goodfellow, P.N. and Lovell-Badge, R. (1990) Nature 346, 245–250.
[3] Berta, P., Hawkins, J.R., Sinclair, A.H., Taylor, A., Griffiths, B.L., Goodfellow, P.N. and Fellous, M. (1990) Nature 348, 448–450.
[4] Jager, R.J., Anvret, M., Hall, K. and Scherer, G. (1990) Nature 348, 452–454.
[5] Koopmans, P., Gubbay, J., Vivian, N., Goodfellow, P.N. and Lovell-Badge, R. (1991) Nature 351, 117–121.
[6] Wright, J.M. and Dixon, G.H. (1988) Biochemistry 27, 576–581.
[7] Van de Wetering, M., Oosterwegel, M., Dooijjes, D. and Clevers, H. (1991) EMBO J. 10, 123–132.
[8] Atherton, E. and Sheppard, R.C. (1989) Solid-phase Peptide Synthesis: A Practical Approach, IRL Press, Oxford.
[9] Anthony-Cahill, S.J., Benfield, P.A., Fairman, R., Wasserman, Z.R., Brenner, S.L., Stafford III, W.F., Altenbach, C., Hubbell, W.L. and DeGrado, W.T. (1992) Science 255, 979–983.
[10] Harley, V.R., Jackson, D.I., Hextall, P.J., Hawkins, J.R., Berkovitz, G.D., Sockanathan, S., Lovell-Badge, R. and Goodfellow, P.N. (1992) Science 255, 453–456.
[11] Nasrin, N., Buggs, C., Fu Kong, X., Carnazza, J., Goebl, M. and Alexander-Bridges, M. (1991) Nature 354, 317–320.