Nucleotide sequence of mouse SCIP cDNA, a POU-domain transcription factor

Elisabeth C.Zimmerman, C.Michael Jones, Victor Fet, Brigid L.M.Hogan and Mark A.Magnuson*
Departments of Molecular Physiology and Biophysics, Medicine and Cell Biology, Vanderbilt University Medical School, Nashville, TN 37232, USA

Submitted January 8, 1991

EMBL accession no. X56959

SCIP (Tst-1) is a POU domain transcription factor expressed in neural tissue and testes. Tst-1 was first cloned from rat testes and found to be expressed both there and in a restricted subset of neurons in embryonic and adult brain (1). Subsequently a cDNA termed SCIP (for suppressed cAMP inducible POU) was isolated from a rat sciatic nerve cDNA library (2). DNA sequence analysis has shown that both the Tst-1 and SCIP cDNAs are products of the same gene (2, 3). A partial amino acid sequence for rat Tst-1 (1) and a full amino acid sequence for rat sciatic nerve SCIP (3) have been reported. The rat protein is 451 amino acids in length and contains both POU-specific and POU-homeodomain-specific regions that are similar to the equivalent regions of Pit-1, Oct-1, Oct-2 and other POU proteins (4, 5, 6). The highest degree of similarity, however, is to the putative neural transcription factors Brn-1 and Brn-2 (1).

We obtained a POU-homeo domain DNA fragment by PCR amplification using degenerate primers and used this DNA fragment to screen a 6.5 day mouse embryo cDNA library in λ GT10. Analysis of six isolates showed they each contained a cDNA fragment of approximately 1.6 kb. Fragments of one cDNA clone (#2-4c) were ligated into Bluescript M13(-) and pEMBL 18 and 19 for DNA sequence analysis. The fragment was 1516 bp in length and contained an open reading frame of 1347 bp. Overall the sequence had a high GC content (72%), and this was even more pronounced at the 5' end (85% of the first 100 bases). The protein encoded was 449 amino acids in length and was 99.1% homologous with rat SCIP. Only one amino acid was different (position 97, serine to glycine) but there was a deletion of two alanines in the poly-alanine rich region near the N-terminus, thus accounting for the shorter length. The POU-domain and POU homeodomain-specific regions of the mouse protein are identical to those of rat SCIP. The cDNA cloned therefore appears to encode mouse SCIP.

Several features of the protein encoded by this cDNA are illustrated in the Figure. The POU-domain (amino acids 240-319) and the POU homeo-domain (amino acids 335-396) are located towards the C-terminus. A poly-alanine repeat is located near the N-terminus followed by 6 poly-glycine repeats, each of which contains 4-6 glycines. A poly-proline and poly-histidine repeat (6 amino acids each) are located near the C-

terminus. In addition to these repetitive regions, the overall composition of the protein is biased in favor of these four amino acids. In total there are 67 alanines (14.9%), 84 glycines (18.7%), 37 histidines (8.2%), and 39 prolines (8.7%). Indeed, just over half (50.5%) of the protein is composed of these four residues. The functional significance of these repetitive domains in SCIP is unknown but other transcription factors are known to contain similar repetitive regions.

ACKNOWLEDGEMENTS

We thank David Weng, Johns Hopkins University School of Medicine, for the 6.5 day mouse embryo cDNA library. These studies were supported by grants to MAM (DK42612 and DK42502) and to BLMH (HD25580) from the National Institutes of Health.

REFERENCES

1. He, X., Treacy, M.N., Simmons, D.M., Ingraham, H.A., Swanson, L.W. and Rosenfeld, M.G. (1989) *Nature* **340**, 35-42.
2. Monuki, E.S., Weinmaster, G., Kuhn, R. and Lemke, G. (1989) *Neuron* **3**, 783-793.
3. Monuki, E.S., Kuhn, R., Weinmaster, G., Trapp, B.D. and Lemke, G. (1990) *Science* **249**, 1300-1303.
4. Herr, W., Sturm, R.A., Clerc, R.G., Corcoran, L.M., Baltimore, D., Sharp, P.A., Ingraham, H.A., Rosenfeld, M.G., Finney, M., Ruvkun, G. and et al. (1988) *Genes Dev.* **2**, 1513-1516.
5. Sturm, R.A., Das, G. and Herr, W. (1988) *Genes Dev.* **2**, 1582-1599.
6. Gerster, T., Balmaceda, C.G. and Roeder, R.G. (1990) *EMBO J.* **9**, 1635-1643.

Diagrammatic illustration of the mouse SCIP cDNA and the open reading frame it contains (boxed region). Unique restriction sites are shown above. Shaded boxes are used to show the POU-specific and POU homeodomain-specific regions and areas containing at least four alanines, glycines, histidines, or prolines.

*To whom correspondence should be addressed at 708 Light Hall, Vanderbilt University School of Medicine, Nashville, TN 37232, USA