Vol. 175, No. 1, 1991 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
February 28, 1991 Pages 333-338

AMINO ACID SUBSTITUTIONS IN THE SP1 ZINC FINGER DOMAIN ALTER THE DNA BINDING AFFINITY TO COGNATE SP1 TARGET SITE

HANS-JÜRGEN THIESEN AND BJÖRN SCHRÖDER

Basel Institute for Immunology, Grenzacherstrasse 487. CH-4005 Basel, Switzerland

Received January 14, 1991

Non conserved amino acids, which are located in the postulated α-helical region of the third zinc finger in human transcription factor SP1, have been replaced by amino acids, which occur at the analogous zinc finger position in human protein Kox 15. This helical domain was mutated from SDHLSKH to SSHLIQH (SP1-M3). Aspartic acid (D), serine (S) and lysine (K) were substituted by serine (S), isoleucine (I) and glutamine (Q). The DNA binding of the mutated SP1-M3 protein to the SP1 cognate target site GGG GCG GGG was significantly impaired, indicating that the amino acids, aspartic acid, serine and lysine play a pivotal role in DNA recognition. The mutated SP1 finger cannot imitate the function of the wild type SP1 finger in interacting with the cognate SP1 target site. This structure-function analysis indicates that the third SP1 zinc finger participates in sequence-specific DNA recognition. Thus, the affinity of zinc finger domains can be altered by substituting a limited number of amino acids. This observation is consistent with the notion that zinc finger domains are positioned in the major groove of the DNA and wrap around the DNA. Structure-function analysis of this kind might lead to the description of a zinc finger specific recognition code.

Several hundred genes encoding Cys_2/His_2 zinc finger domains have been identified from *Drosophila* to man, based on the zinc finger repeat of 28-30 amino acids (1-7). The functions of the encoded proteins are poorly understood. Miller, McLachlan and Klug (8) predicted the classical zinc finger structure based on the observation that *Xenopus* transcription factor III A coordinates zinc (9). By interactive computer modelling strategies the existence of an α-helical zinc finger domain was postulated encompassing the conserved leucine and the conserved histidine (10,11). The presence of an α-helical zinc finger region was demonstrated by nuclear magnetic resonance (NMR) analysis (12, 13). This region is thought to be positioned in the major groove of the DNA making sequence-specific contacts (9-14). The sequence analysis of zinc finger regions of about 70 individual zinc fingers, present in Kox1 to Kox 30, led to the hypothesis of a zinc finger specific recognition code (7). It is postulated that each zinc finger interacts with DNA in a sequence-specific manner independently of the binding specificity of the neighbouring zinc fingers. In turn, the sequence specificity determined for one specific finger should be retained independently of its position within a zinc finger region.

In order to identify amino acid positions involved in contacting the DNA in a sequence-specific manner, amino acids were replaced in the third SP1 finger (4) by amino acids occurring at analogous positions in Kox 15, a human zinc finger protein (7). If protein folding of individual zinc fingers occurs independently of each other, the helical zinc finger conformation of the mutated SP1 zinc finger should resemble the native finger conformation in the Kox 15 protein. Thus, it is

0006-291X/91 $1.50

less likely that the helical conformation of the DNA binding domain is impaired. However, replacing amino acids which are likely to participate in sequence-specific DNA recognition should have a significant impact on SP1 DNA recognition. In this study, the amino acid region SSHLIQH which occurs in several zinc finger domains, e.g. Kox15, was substituted for the SDHLSKH finger region of SP1. The Kox 15 gene has recently been mapped to chromosome 10cen-q24 (15). By using electrophorectic mobility shift assays (EMSA) the DNA binding of mutated SP1-M3 is shown to be strongly reduced in comparison to the affinity of wild type SP1 protein.

Materials and Methods

Oligonucleotides: Oligonucleotides for site-directed mutagenesis and for electrophoretic mobility shift assays (EMSA) were synthesized on 380A Applied Biosystems DNA Synthesizer. The oligonucleotide 1892 used in EMSA contains one SP1 binding site TTGGGGCGGGGCTT surrounded by cassette sequences (16), which harbour the appropriate primer annealing sites for primer A and primer B (17,18). For site-directed mutagenesis oligonucleotide 1688 (GTTGTATCAGGTGGGAACTCCTCATGAAGCGCTTAGGACAC) and oligonucleotide 1689 (CATGAGGAGTTCCCACCTGATACAACATATCAAGACCCACCAG) were synthesized encoding serine at zinc finger position 17, isoleucine at position 20 and glutamine at position 21 (For assignment of amino acid positions, see reference 11). The oligonucleotides 393 (GTAAAACGACGGCCAGTG) and 392 (AAACAGCTATGACCATG), which are universal primers of Bluescript plasmid (Stratagene), have been used together with the oligonucleotides 1688 and 1689 in PCR mutagenesis (19). Oligonucleotide 1956 (CAGCCCGGGAGATCTGCC-ACCTGCATGAC) introduces a Bgl II site at the 3' end of the SP1 fragment in pB-516c.

Site-directed mutagenesis: The Bam H1 - Bgl II fragment derived from pP_{ac} Sp1-516c (20), encoding three zinc fingers of the human transcription factor SP1 (4), was cloned into the BamH1 site of Bluescript (Stratagene) to yield pB-516c. Two polymerase chain reactions (PCR) were performed using oligonucleotide pairs 393/1688 and 1689/392 together with pB-516c generating Sp1 fragments A and B (Fig.1). Each fragment harbours the introduced mutations either at the 3' or 5' site. They were isolated from a 6% PAA gel (18). The complete SP1 fragment encoding the desired mutations and a restored Bgl II site was generated by performing a second PCR using primer 393, 1956 on SP1 fragments A and B. The PCR product was extracted by phenol/chloroform, digested with BamH1 and Bgl II, gel purified and cloned into pAR 3039 (21) to yield pAR-SP1M3. Standard PCR conditions were applied (17,18). Introduced mutations were verified by sequence analysis.

E.coli expression: Mutated Sp1 protein was expressed and purified according to the procedure described for the analogous wild type SP1 protein (17,18). Mutated SP1 protein was subjected to FPLC Mono S chromatography, eluted with an increasing salt gradient, peak fractions were collected and analysed on 15% polyacrylamide-SDS gel in comparison with recently purified SP1 (17,18). Fractions containing the mutated SP1 protein were pooled. Protein concentrations were determined by the method of Lowry (22).

Electrophoretic Mobility Shift Assay (EMSA): Double-stranded oligonucleotide 1892 was generated by PCR amplification using primer A and primer B (17) and labelled as previously described (18). A 30 µl standard EMSA reaction mixture consists of 10 µl of 3 x band shift buffer (15 mM NaCl, 150 mM KCl, 36 mM Hepes pH 7.9, 36% glycerol, 5 mM $MgCl_2$, 300µM $ZnCl_2$), 6µl H_20, 3 µl DTT (10 mM) and 10 µl labelled oligonucleotide 1892 (10000-20000 Cerenkov cpm), 1 µl SP1 protein. Equal amounts of SP1 and SP1-M3 proteins were diluted in protein dilution buffer (100 mM KCl, 10 mM Tris-HCL, 1 mM dithiothreitol, 1mM $ZnCl_2$). The band shift reactions were incubated for 30 min at R.T. and loaded onto a 6% polyacrylamide band shift gel (acrylamide/bisacrylamide 30%:0.8%) containing 100µM $ZnCl_2$ and 0.25x Tris-borate electrophoresis buffer (TBE)(18). The gel electrophoresis was run in 0.25x TBE (100µM $ZnCl_2$) at 10 V/cm and a current not exceeding 20 mA. The gels were dried onto Whatman 3MM paper and exposed to Kodak XAR.

Results and Discussion

The zinc finger domains of several Cys_2/His_2 type zinc finger proteins have been shown to form sequence-specific contacts with DNA. Mutants of ADR1 (23) and of the Drosophila Krüppel protein (24) had lost their sequence-specific binding once evolutionarily conserved amino acids such as cysteines, histidines or amino acids of the histidine/cysteine finger link were mutated. It is likely that in most of these cases the general conformation of the zinc finger is distorted, e.g. if one of the amino acids involved in coordinating zinc is replaced, the rigidity and structural constraints of the zinc finger are impaired (24). If the conserved amino acids phenylalanine and leucine of the zinc finger consensus ($YxCx_2Cx_3Fx_5Lx_2Hx_3HTGEKP$) are mutated, the formation of the postulated hydrophobic core in the putative finger tip might be effected (25). Interactive computer modelling (10,11) and NMR studies on zinc finger peptides (12, 13) suggest that an alpha helical region might encompass amino acid position 16 to 24 including the conserved leucine and histidine (for assignment of amino acid positions, see reference 11). In comparison, α-helical regions of DNA binding domains, analysed in helix-turn-helix motifs (26, 27) and in homeo-domains (28) have been shown to be positioned in the major groove of the DNA, favouring sequence-specific contacts. Therefore, it seems likely that the sequence specificity of zinc finger proteins is determined by specific amino acids which are positioned in the putative helical region. In order to show that this region is responsible for sequence-specific DNA recognition, the helical region of the third SP1 finger, SDHLSKH, was replaced by the amino acids SSHLIQH, which are present at identical positions in one of the fingers in Kox 15, a human zinc finger protein (7).

Oligonucleotides with mutations in the third finger of SP1 were made by performing PCR based site-directed mutagenesis on pB-516c (Fig. 1). The non conserved amino acids, aspartic acid, serine and lysine at zinc finger position 17, 20 and 21, were replaced by serine, isoleucine and glutamine. Recombinant mutated SP1-M3 protein was purified by FPLC Mono S chromatography. Peak fractions containing mutated SP1-M3 protein were identified by sodium dodecyl sulfate (SDS)-polyacrylamide electrophoresis (29) (Fig.2).

Recently, the DNA recognition of recombinant wild type SP1 was characterized (17). High affinity binding sites containing GCG nucleotides in their middle seemed to be preferentially selected (17, 30). The DNA binding activity of mutated SP1-M3 protein was assessed and compared with SP1(WT) in electrophoretic mobility shift assays (Fig.3). Comparing the binding activity of both proteins demonstrates that the mutated SP1-M3 protein binds to the SP1 target site GGG GCG GGG with an at least five fold reduced affinity (Fig.3). Thus, the structure-function analysis of mutated SP1 and wild type SP1 identifies the introduced mutations in the third finger as a loss of function mutation. Only at higher protein concentrations the mutated SP1-M3 protein binds to the Sp1 target site (data not shown).

Footprint analysis, methylation interference studies and mutational analysis of SP1 target sites demonstrated that the SP1 target site encompasses 9-10 nucleotides (4, 16). The original SP1 consensus site GGG GCG GGG consists of a stretch of guanines, only interrupted by one cytidine. Theoretically, each of the three SP1 zinc fingers can be assigned to three nucleotides of

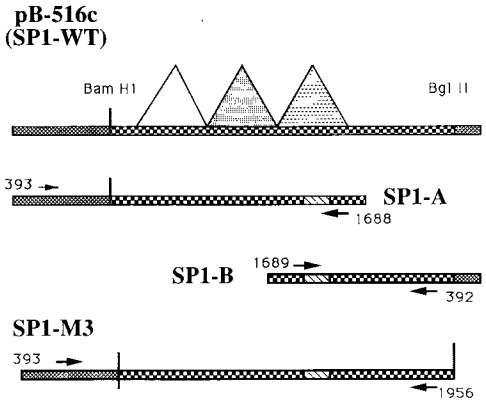

Fig.1 PCR based site-directed mutagensis (19).
The three SP1 zinc fingers are represented by three pyramids. The third finger was mutated by using oligonucleotides 1688 and 1689.

the target site. Assuming that the sequential zinc finger representation 1, 2 and 3 is also repeated at the level of SP1 target site (8), then the second finger should always bind to the nucleotides GCG in the middle of the consensus site (Fig.4). The third finger should only participate in binding

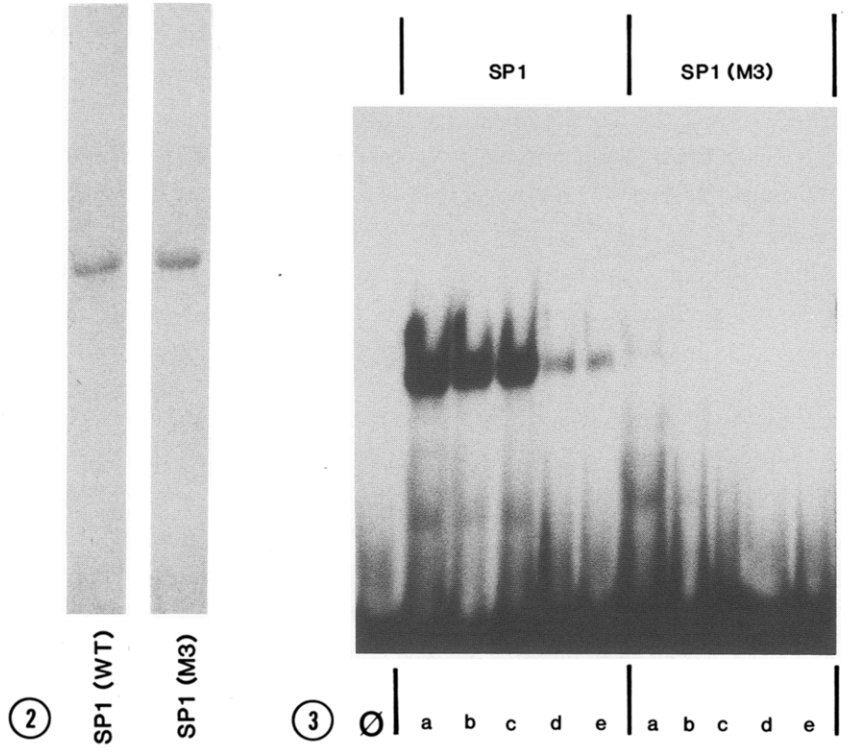

Fig.2 SDS polyacrylamide electrophoresis (22) of SP1 (WT) and mutated SP1-M3.

Fig.3 Electrophoretic mobility shift assay (EMSA).
Wild type SP1 and mutated SP1-M3 protein dilutions were incubated with labelled double-stranded oligonucleotide 1892 that harbours one cognate SP1 binding site.
(Lane ø: no protein; lane a: 3300 pg; lane b: 1650 pg; lane c: 1237 pg; lane d: 825pg; lane e: 618 pg). Protein DNA complexes were analysed on a 6% polyacrylamide gel.

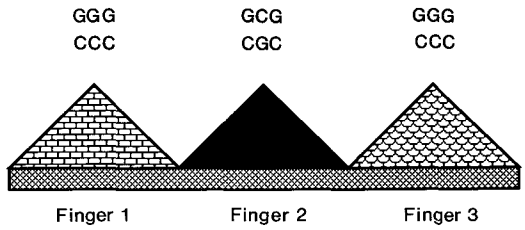

Fig.4 Model of DNA protein interactions of SP1 protein with its target site GGG GCG GGG. Each SP1 zinc finger has been assigned to three nucleotides of its cognate target site. This representation implies that the SP1 protein has to wrap around the DNA.

either to the 5´ or 3´ ends of the binding site. In each case, the DNA recognition of the third zinc finger would require nucleotide interactions with GGG and/or CCC on the opposite strand. Protein DNA interactions occurring between the SP1 target site and the third SP1 finger are significantly impaired by the mutated finger encompassing SSHLIQH (Fig.3). Thus, the nucleotides GGG or CCC do not seem to be an optimal recognition site for a zinc finger encompassing SSHLIQH.

Though direct DNA contacts have not yet been established for the third finger of SP1 it can be envisaged that the strength of electrostatic interactions is lessened by the replacement of basic amino acids. By the introduction of isoleucine, a non polar hydrophobic amino acid, the possibilities of forming appropriate hydrogen bonds to the GGG/CCC motif are significantly impaired (Table 1). It should be highly interesting in assigning sequence-specific and non sequence-specific DNA interactions to individual amino acids once crystallographic data of zinc fingers are available. Since adjacent nucleotides in the SP1 target sites seem to be recognized by the three zinc fingers (Fig.4), it is highly suggestive that the zinc finger domains are positioned in the major groove of the DNA and thus coil around the DNA.

In ongoing experiments the target detection assay (TDA)(17) is used to assess whether the mutated SP1 protein has gained new binding preferences. The mutations in the third finger of SP1-M3 represents the first example where the binding affinity of SP1 has been altered by site-directed mutagenesis of amino acids that are not highly conserved in the zinc finger consensus (11). Taking into account that the human genome might encode several hundred zinc finger proteins (5,6), numerous non identical zinc fingers should be capable of recognizing identical DNA target sites. Because of this high redundancy in forming putative DNA binding domains, the existence of a zinc finger specific recognition code will depend on the capability of individual zinc fingers in discriminating between the nucleotides they interact with. Structure-function analysis of the kind reported here will reveal whether DNA recognition rules can be obtained and later on be applied for designing DNA binding domains with desired specificities (7).

Table 1. The DNA recognition of mutated SP1-M3

Mutants of Finger III	GGG/CCC	Proteins
S D H L S K H	+++	SP1
S S H L I Q H	-	Kox 15

Amino acid substitutions in the third finger of SP1 significantly reduce the DNA binding to the cognate SP1 target site.

Acknowledgments

Michael V.Wiles and Gek-Kee Sim are thanked for critical reading of the manuscript. R.Tjian is thanked for supplying the plasmid pP_{ac} Sp1-516c (20). The Basel Institute for Immunology was founded and is supported by F. Hoffmann-La Roche Ltd., CH-4005 Basel, Switzerland.

References

1. Ginsberg, A.M., King, B.O. and Roeder, R.G. (1984) Cell 39, 479-489.
2. Rosenberg, U.B., Schroeder, C., Priess, A., Kienlin, A., Côte, S., Riede, I. and Jäckle, H. (1986) Nature 319, 336-339.
3. Chowdhury, K., Deutsch, U. and Gruss, P. (1987) Cell 48, 771-778.
4. Kadonaga, J.T., Carner, K.R., Masiar, F.R. and Tjian, R. (1987) Cell 51,1079-1090.
5. Bellefroid, E.J., Lecocq, P.J., Benhida, A., Poncelet, D.A., Belayew, A. and Martial, J.A. (1989) DNA 8:377-387.
6. Bray, P. and Thiesen, H.J. (1990) N.Biol. 2, 325-327
7. Thiesen, H.J. (1990) N.Biol. 2, 363-374
8. Miller, J., McLachlan, A.D. and Klug, A. (1985) EMBO J. 4, 1609-1614.
9. Diakun, G.P., Fairall, L. and Klug, A. (1986) Nature 324, 689-691
10. Berg, J. (1988) Proc. Natl. Acad. Sci. USA 85, 99-102.
11. Gibson, T.J., Postma, J.P.M., Brown, R.S. and Argos, P. (1988) Protein Engineering 2, 209-218.
12. Lee, M.S., Gippert, G.P., Soman, K.V., Case, D.A. and Wright P.E. (189) Science 245, 635-637.
13. Parraga,G., Horvath, S.J., Eisen, A., Taylor, W.E., Hood, L., Young, E.T. and Klevit, R.E. (1988) Science 241,1489-1492.
14. Gottesfeld, J.M., Blanco, J. and Tennant, L.K.L. (1987) Nature 329, 460-462.
15. Huebner, K., Druck, T., Croce, C.M. and Thiesen, H.J. (1991) Am. J. Hum. Genet. (in press).
16. Thiesen, H.J. (1990) DNA and Protein Engineering Techniques 2, 92-95.
17. Thiesen, H.J. and Bach, C. (1990) Nucleic Acids Res. 18, 3203-3209
18. Thiesen, H.J. (1990) In Immunological Methods (I. Lefkovitz and B.Pernis, Eds.) Vol. IV, pp. 61-74 , Academic Press, Orlando, Fl.
19. Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K., and Pease, L.R. (1989) Gene 77, 51-59.
20. Courey, A.J. and Tjian, R.(1988) Cell 55, 887-898.
21. Rosenberg, A.H., Lade, B.N., Chui, D.S., Lin, S.W., Dunn, J.J. and Studier, F.W. (1987) Gene 56, 125-135.
22. Lowry, O.H., Rosebrough, N.J., Fare, A.L. and Randall, R.J. (1951) J. Biol. Chem. 193, 265-275.
23. Blumberg, H., Eisen, A., Sledziewski, A., Bader, D. and Young, E.T. (1987) Nature 328, 443-445.
24. Redemann, N., Gaul, U. and Jäckle, H. (1988) Nature 332, 90-92
25. Gaul, U., Redemann, N. and Jäckle, H. (1989) Proc. Natl. Acad. Sci. (USA) 86, 4599-4603
26. Wharton, R.P. and Ptashne, M. (1985) Nature 316, 601-605
27. Anderson, W.F., Ohlendorf, D.H., Takeda, Y. and Mathews, B.W. (1981) Nature 290, 754-758.
28. Kissinger, C.R., Liu, B., Martin-Blanco, E., Kornberg, T.B. and Pabo, C.O. (1990) Cell 63, 579-590.
29. Lämmli, U.K. (1970) Nature 227, 680-685
30. Letovsky, J. and Dynan, W.S. (1989) Nucleic Acids Res. 17, 2639-2653.