FEBS Letters 407 (1997) 320–324 FEBS 18513

Functional interference between contacting amino acids of homeodomains

Lucia Pellizzari, Gianluca Tell, Dora Fabbro, Carlo Pucillo, Giuseppe Damante*

Dipartimento di Scienze e Tecnologie Biomediche, Università di Udine, P. le Kolbe 1, 33100 Udine, Italy

Received 26 February 1997; revised version received 11 March 1997

Abstract In a protein, the function of an amino acid at some position depends on the amino acids at other positions. Here we demonstrate a functional interference between base-contacting amino acids (at positions 50 and 54) of homeodomains. When, in the context of Antennapedia or Goosecoid homeodomains, Lys^{50} is paired to Tyr^{54} or Ala^{54} and Gln^{50} is paired to Met^{54}, the resulting proteins efficiently discriminate among different DNA sequences. In contrast, in the presence of the pair Lys^{50}-Met^{54}, both homeodomains show a reduced capability to discriminate among different DNA sequences. Sequence selection experiments performed in the context of the Goosecoid homeodomain suggest that the presence of Met^{54} precludes the base-discriminating function of Lys^{50}. These results may explain why the pair Lys^{50}-Met^{54} is never found in natural homeodomains.

Key words: DNA binding; Homeodomain; Amino acid

1. Introduction

An amino acid that contributes to the function of a protein in one sequence context may be unfavorable in another. In terms of protein evolution, the 'fitness' of an amino acid at some position depends on the amino acids at other positions [1]. Studies of protein families may be useful to reveal such a kind of functional interference. In this view, the homeodomain (HD) protein family is a powerful system. The HD is the DNA-binding domain of a large number of transcription factors controlling differentiation and development [2]. It consists of a 60 amino acids-long structure, conserved along phylogenesis [3]. The way how HDs interact with DNA is also conserved [4]. Base-contacting amino acids are located in the N-terminal arm as well as in the recognition helix. Contacting amino acids of the recognition helix are located at positions 47, 50, 51 and 54. Residues at positions 47 and 51 do not contribute to the differential binding specificity existing among different HDs. In contrast, residues at position 50 and 54 play a discriminatory role. Most HDs recognize sequences having the 5′-TAAT-3′ core motif or a variation of it [4]. The residue at position 50 discriminates the dinucleotide located after the 3′ end of the 5′-TAAT-3′ motif [5]. Differently, amino acid at position 54 recognizes the nucleotide at the 3′ end of the 5′-TAAT-3′ core motif, therefore it is one of the critical determinants between HDs recognizing or not the 5′-TAAT-3′ core motif [6]. Hence, the combination of amino acid existing at positions 50 and 54 should be a chief determinant for the DNA-binding specificity of HD. Amino acids 50 and 54 side chains lie very close to each other in the recognition helix and are part of a fluctuating, short-lived bonding network [7,8]. These structural features suggest that epistatic interactions may occur between these two residues, so that only particular combinations may ensure correct binding properties to HDs. In fact, about 400 HDs from different species have been thus far sequenced and a bias in the combinations of amino acids 50–54 could be observed [3]. Most HDs contain Gln at position 50 but few contain Lys at the same position. Among the Gln^{50}-containing HDs, more than 50% contain Met at position 54. However, none of the Lys^{50}-containing HDs contains Met^{54}. Previous data obtained in the context of TTF-1HD would suggest that the DNA-discriminating function of HDs containing the pair Lys^{50}-Met^{54} could be reduced. However, the peculiar DNA-binding specificity of TTF-1HD (it recognizes sequences with the core motif 5′-CAAG-3′) precludes to extend this concept to typical HDs, recognizing the 5′-TAAT-3′ core motif.

Aim of this work is to test the effect of the pair Lys^{50}-Met^{54} in the context of two typical 5′-TAAT-3′-recognizing HDs, Antennapedia (AntpHD) and Goosecoid (GscHD). The results indicate an inefficient DNA-binding function for Lys^{50}-Met^{54}-containing HDs. Moreover, sequence selection experiments performed in the context of GscHD suggest that the presence of Met at position 54 impairs the base-discriminating function of Lys^{50}.

2. Materials and methods

2.1. Plasmids and materials

The plasmid encoding for Antp and Gsc HDs have been already described [9,10]. In these plasmids the transcription of the HD-coding sequence is driven by the T7 RNA polymerase promoter. Mutants were constructed by a PCR-based procedure [11], cloned in the bacterial expression vector pT7.7 and verified by nucleotide sequence (Sequenase, Amersham). The protein sequence of the mutants is shown in panel A of Fig. 1. Wherever not specified, all materials were purchased from Sigma.

2.2. Protein expression and purification

Proteins were expressed using the BL21 (DE3) *E. coli* strain [6,12]. After protein expression, the bacterial pellet was resuspended in lysis buffer A (Tris-HCl 20 mM pH 7.4, EDTA 0.1 mM, glycerol 10%, PMSF 1 mM, NaCl 400 mM) and cells were disrupted by a Labsonic U sonicator (Braun) (4 strokes of 10 s at 4°C). The lysed pellet was spun (10000×*g*, 20 min, 4°C) and the supernatant was loaded onto a Econo Pac S Cartridge (Bio-Rad) equilibrated with buffer A. After washing the column with 10 volumes of buffer A, the proteins were eluted with buffer B (the same as buffer A but with NaCl 800 mM). From SDS-PAGE analysis, the achieved purity was 20% for AntpHD and relative mutants and 10% for GscHD and relative mutants (Fig. 1B). These differences have no influence on the quality of the results, because comparisons of the binding activity and specificity are made between members of the same set of mutants (see Figs. 2 and 3). Several lines of evidence demonstrate that these degrees of purity are sufficient to correctly measure the binding activity and specificity of HDs: (i) the contaminating bacterial DNA (present in the crude bacterial extract and able to interfere in the gel-retardation assay) is

*Corresponding author. Fax: (39) (432) 547210.

PII S0014-5793(97)00368-2

completely removed; (ii) two of the proteins, AntpHD and AntpHD(Y^{54}), purified to homogeneity by the procedure of Müller et al. [9], display a binding activity and specificity identical to those of the partially purified proteins; (iii) extracts from BL21, transformed with empty pT7.7 vector and subjected to the partial purification procedure with the Econo Pac S cartridge, do not possess DNA-binding activity to all oligonucleotides used in this study neither modify the binding activity and specificity of homogeneity-purified HDs (data not shown).

The concentration of the active protein was measured by oligonucleotide saturation assay [6,13]. A gel-retardation assay without the presence of calf thymus DNA (see below) was performed, using increasing amounts of oligonucleotides (0.3–50 nM). The protein-bound and free oligonucleotide concentrations values were quantitated as described below and the data were subjected to the Scatchard plot analysis [14,15].

2.3. Gel-retardation assay and quantitation of the binding activity

Oligonucleotides were labeled at the 5′-end by polynucleotide kinase (Boehringer). The binding activity and specificity of HDs were measured as published [6,13]. Briefly, gel-retardation assays were performed incubating protein and double-stranded, ^{32}P-labeled, oligonucleotides in a buffer containing 20 mM Tris-HCl pH 7.6, 75 mM KCl, 0.25 mg/ml bovine serum albumin (BSA), 5 μg/ml calf thymus DNA, 5mM dithiothreitol (DTT), 10% glycerol for 30 min at room temperature. Protein-bound DNA and free DNA were separated on native 7.5% polyacrylamide gel, run in 0.5×TBE for 1.5 h at 4°C. Gels were dried, exposed to X-ray films and bands were quantitated by densitometric scanning of the autoradiogram using a LKB laser densitometer. Results were expressed as fraction of the value of the bound/free ratio obtained for wild-type HDs, considered arbitrarily as 100.

2.4. Sequence selection

Base preference of Gsc and Gsc(M^{54}) HDs was determined by a modified version of the procedure described by Blackwell and Weintraub [16]. The oligonucleotides N:5 (5′-GCTAGCTCCAAGGNTTAACTGGCGTCATAGCTGTTTCCTG-3′) and N:6–7 (5′-GCTAGCTCCANNATTAACTGGCGTCATAGCTGTTTCCTG-3′) (in both sequences N indicates the presence of all four bases at that position), were made double stranded by Klenow enzyme (Boehringer) elongation of a primer complementary to the underlined sequence. 40 pmol of both oligonucleotides were used in preparative gel-retardation assays with the proteins GscHD and GscHD(M^{54}). The amount of protein was calibrated to bind only 2–4% of the total oligonucleotides. The protein-bound oligonucleotides were eluted from gels and sequenced by a PCR-based procedure (Taq DNA Cycle Sequencing Kit; Boehringer, Mannheim).

3. Results and discussion

AntpHD contains Gln and Met at positions 50 and 54, respectively; at the same positions GscHD contains Lys and

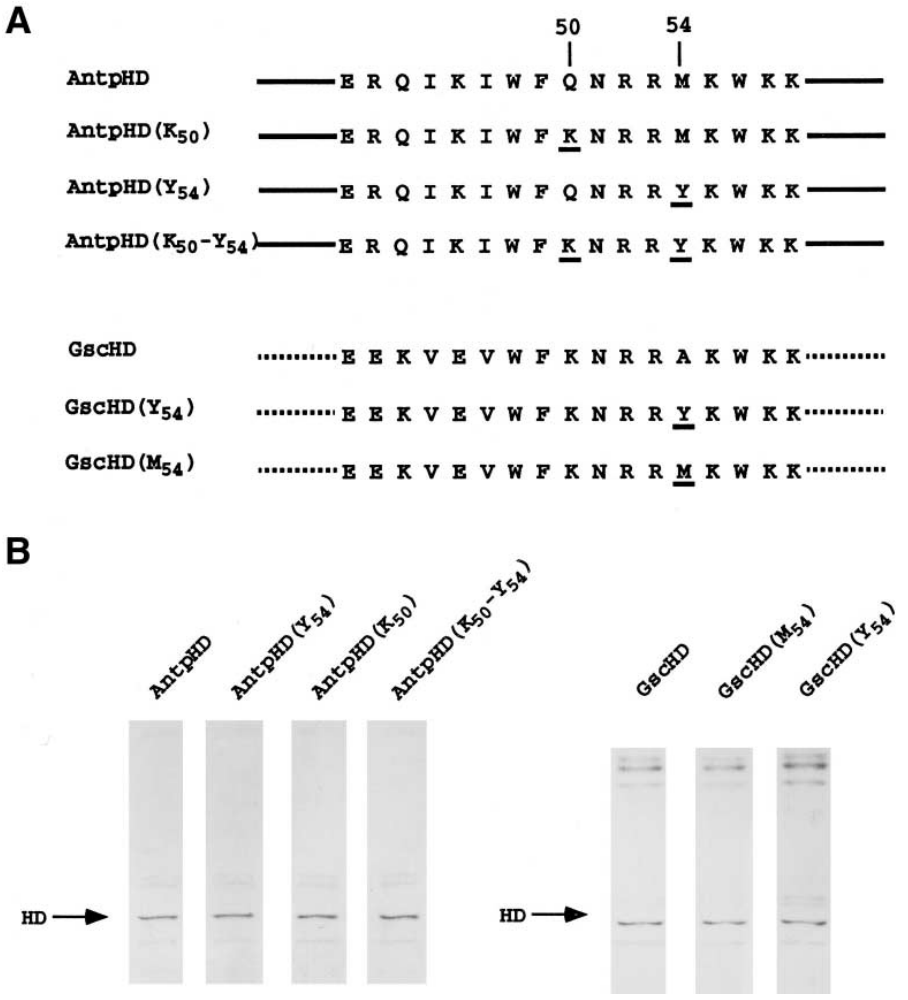


Fig. 1. A: Primary structure of the recognition helix of HDs used in this study; amino acids mutated with respect to those present in wild-type HDs (AntpHD and GscHD) are underlined. B: SDS-PAGE of the partially purified proteins. 10 μl of each sample after the Econo Pac S Cartridge purification were loaded on a 18% SDS-PAGE in reducing conditions. Proteins were then detected by silver staining and the degree of purification was measured by densitometric scanning of the gel.

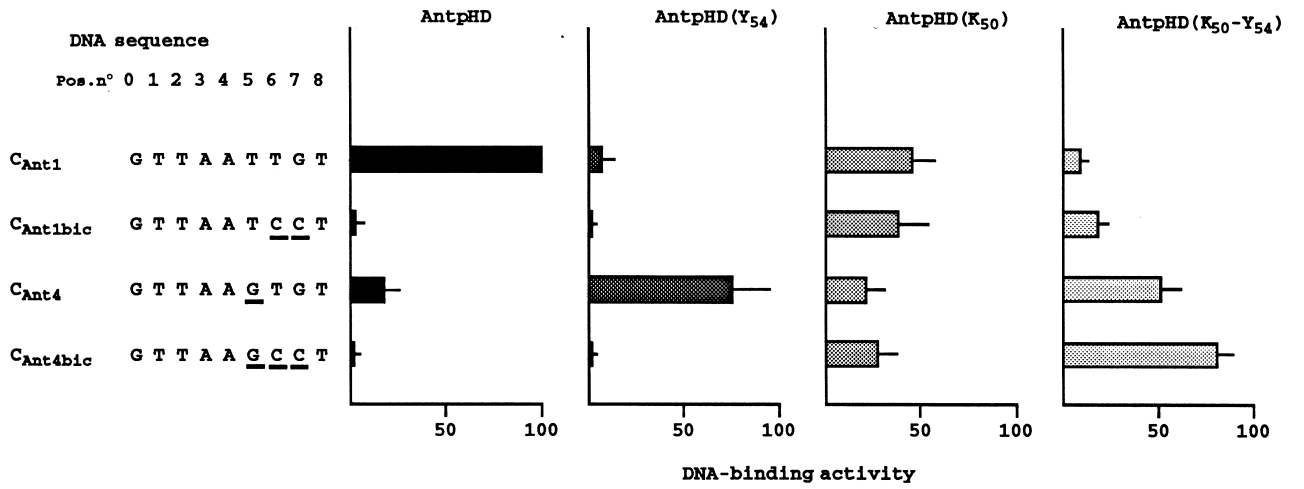


Fig. 2. Relative DNA-binding activity of AntpHD and mutants of it. Sequences of the DNA sites are shown on the left. The DNA-binding activity is expressed as fraction of that measured for the AntpHD/C_{Ant1} interaction, arbitrarily considered as 100. Each bar represents the mean value of four independent determinations. Lines above bars indicate the standard deviation of mean values.

Ala. The sequence of the recognition helix of wild-type HDs and relative mutants used in this study is shown in panel A of Fig. 1.

In a first set of experiments we tested the effect of changing amino acids at positions 50 and 54 in the context of AntpHD. The sequences used to test the binding specificity contain the various combinations of the nucleotides recognized by the amino acids located at positions 50 and 54 and results are shown in Fig. 2. As expected, the favorite site of AntpHD is the oligonucleotide C_{Ant1} which contains T at position 5 (recognized by Met^{54}) and TG at positions 6–7 (recognized by Gln^{50}). C_{Ant4}, though recognized at lower affinity than C_{Ant1}, is preferred over $C_{Ant1bic}$ and $C_{Ant4bic}$, indicating that the presence of the dinucleotide CC at position 6–7 is absolutely incompatible with an efficient binding by the wild-type AntpHD. The mutant AntpHD(Y^{54}) prefers the C_{Ant4} oligonucleotide, according to the preference of Tyr^{54} for G at position 5 [6]. The double mutant in which Gln^{50} and Met^{54} have been changed to Lys and Tyr respectively, AntpHD(K^{50}-Y^{54}), prefers the oligonucleotide $C_{Ant4bic}$, in agreement with the notion that Lys^{50} prefers sites with the dinucleotide CC at position 6–7 [5]. This protein prefers $C_{Ant4bic}$ over $C_{Ant1bic}$, indicating that the Tyr at position 54 discriminates at the level of base-pair 5 also in the presence of Lys^{50}. AntpHD(K^{50}) contains the pair Lys^{50}-Met^{54}. The binding data indicate that the DNA-recognition function of this protein is inefficient. In fact, (i) the capability of discriminating different DNA sequences is greatly reduced (the difference in relative binding affinity between the strongest and the weakest bound sequence is only 2-fold); (ii) the maximal binding affinity, obtained with the oligonucleotide C_{Ant1}, is reduced compared to the ones observed for other proteins of

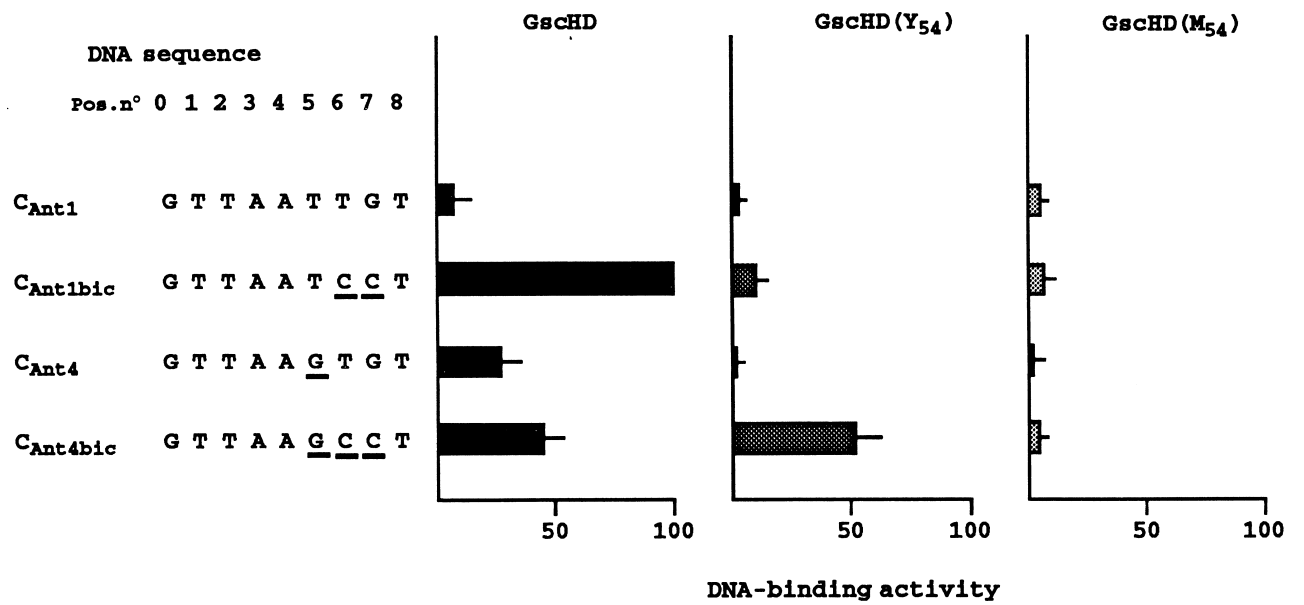


Fig. 3. Relative DNA-binding activity of GscHD and mutants of it. Sequences of the DNA sites are shown on the left. The DNA-binding activity is expressed as fraction of that measured for the GscHD/$C_{Ant1bic}$ interaction, arbitrarily considered as 100. Each bar represents the mean value of three independent determinations. Lines above bars indicate the standard deviation of mean values.

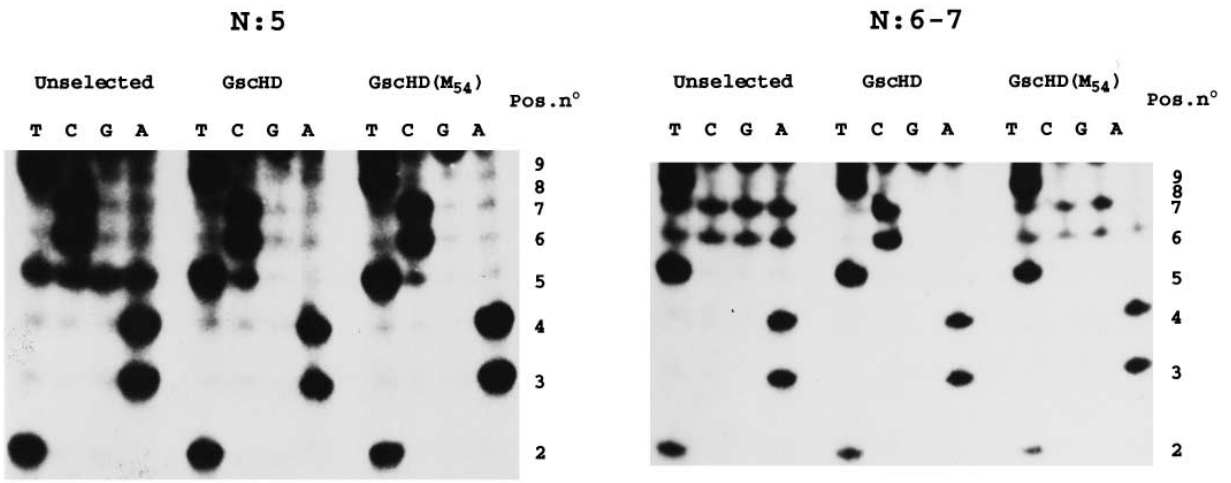
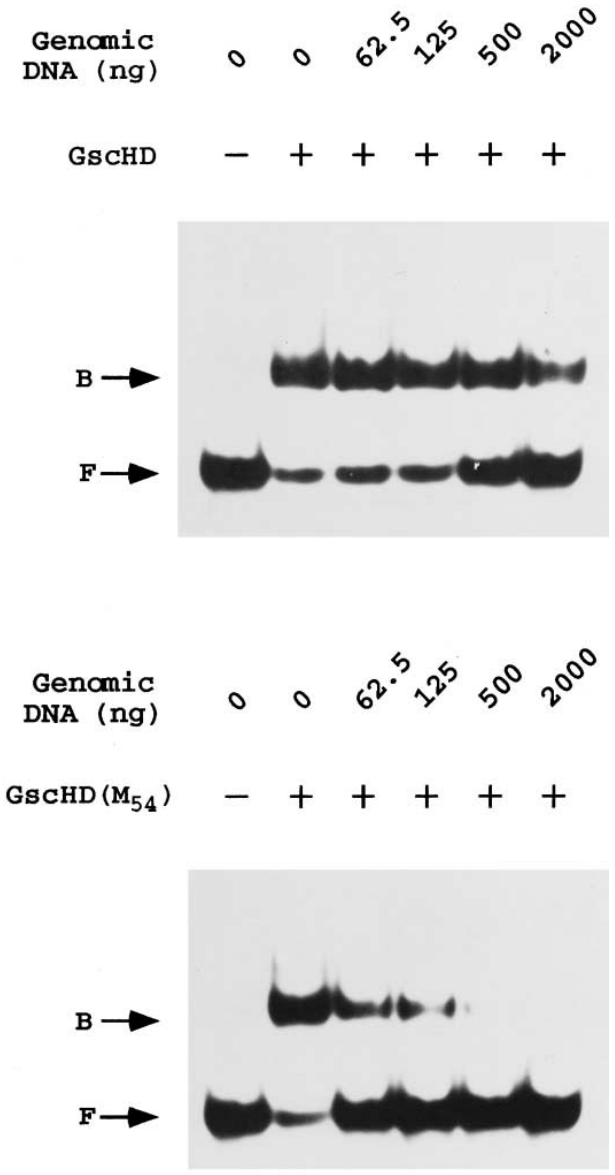


Fig. 4. Base preference of GscHD and GscHD(M^{54}). For each oligonucleotide subjected to selection (N:5 and N:6–7), three DNA sequences are shown: the unselected (on the left), the GscHD-selected (in the middle), the GscHD(M^{54})-selected (on the right). The number of each base is indicated at the extreme right of each experiment and corresponds to the numbering of Figs. 2 and 3.



the Antp series with respective preferred sites; (iii) the favorite site (C_{Ant1}) does not contain the dinucleotide CC at position 6–7, indicating that the discriminating function of Lys^{50} is significantly impaired.

GscHD naturally contains Lys at position 50 [17]. As expected, the oligonucleotide $C_{Ant1bic}$ is the favorite sequence of this protein. In fact, Lys^{50} would select CC over TG at positions 6–7 and Ala^{54} allows a better recognition of T over G at position 5 (Fig. 3). For the mutant GscHD(Y^{54}), where Ala^{54} is changed to Tyr, the preferred site is $C_{Ant4bic}$, indicating that also in the context of GscHD the discriminating function of Tyr^{54} is preserved in presence of Lys^{50}. The mutant GscHD(M^{54}) contains the pair Lys^{50}-Met^{54}. This protein shows both a reduced binding activity and a reduced discriminatory power for the tested oligonucleotides. Thus, in the context of GscHD the replacement of Ala^{54} with Met impairs the DNA-binding function of this protein, indicating again an inefficient DNA-binding function when a HD contains the pair Lys^{50}-Met^{54}. Structural studies with Engrailed HD have revealed that the side chain of Ala^{54} does not establish contact with DNA [18]. This evidence suggests that the reduced binding function of GscHD(M^{54}) mutant could be due to a negative effect of Met^{54} on the Lys^{50}-DNA interaction. This possibility was tested by site selection experiments (Fig. 4). The N:5 oligonucleotide contains a degeneration at position 5, the base-pair recognized by amino acid 54. The oligonucleotide N:6–7 contains degenerations at positions 6 and 7, the base-pairs recognized by amino acid 50. Selected DNAs from N:5 oligonucleotide show that GscHD or GscHD(M^{54}) strongly prefer T at position 5 (Fig. 4), indicating that these proteins do not differ in their capability to discriminate at this position. In contrast, selected DNAs from N:6–7 oligonucleotide indicate that only GscHD, but not GscHD(M^{54}), is able to discriminate at level of base-pairs 6 and 7. In fact, GscHD strongly prefers the dinucleotide CC at position 6–7, while

←

Fig. 5. GscHD discriminates specific from non-specific DNA sequences more efficiently than GscHD(M^{54}). The indicated amounts of genomic DNA (from calf thymus) were added in the binding reaction simultaneously to the HDs and the $C_{Ant1bic}$ oligonucleotide. After 30 min the reactions were loaded on polyacrylamide gels.

GscHD(M^{54}) is by no means able to discriminate at position 6, and only a weak preference for G is observed at position 7. Since the dinucleotide at position 6–7 is contacted by amino acid at position 50 (Lys in the Gsc context), we conclude that in the mutant GscHD(M^{54}) the discriminating function of Lys^{50} is severely reduced.

The impaired DNA-discriminating function of Lys^{50} by the presence of Met^{54} would suggest that the effects shown in Figs. 2 and 3 could be mainly due to a reduced discriminating function between specific and non-specific DNA sequences of the whole Lys^{50}-Met^{54}-containing HDs (DNA-binding assays of Figs. 2 and 3 are performed in presence of non-specific DNA, see Section 2). This possibility was tested in the experiment shown in Fig. 5. The GscHD/$C_{Ant1bic}$ or the GscHD(Met^{54})/$C_{Ant1bic}$ complexes were incubated either in the absence or in the presence of genomic DNA (used at different concentrations). In the absence of competitor DNA, GscHD(Met^{54}) is able to interact with the oligonucleotide $C_{Ant1bic}$ as well as the wild-type protein. However, the complex GscHD(Met^{54})/$C_{Ant1bic}$ is, at least, 30-fold less resistant than the complex GscHD/$C_{Ant1bic}$ to the competition with genomic DNA. These results indicate that GscHD(Met^{54}) is not able to discriminate among different DNA sequences as efficiently as the wild-type GscHD. Therefore, in presence of the pair Lys^{50}-Met^{54}, the HD capability of discriminating different DNA sequences is severely reduced.

In a binding assay without competitor DNA performed with fushi tarazu (ftz) HD, the pair Gln^{50}-Met^{54} prefers GG over CC at positions 6–7, while the pair Lys^{50}-Met^{54} shows the opposite binding preference [19]. However, the ftz HD which contains the pair Lys^{50}-Met^{54} shows a differential binding affinity between GG- and CC-containing sequences 2.5-fold lower than that of the protein containing the pair Gln^{50}-Met^{54} [19]. Accordingly, a ftz protein which contains the pair Lys^{50}-Met^{54} activates promoters with a binding site containing either CC or GG at position 6–7 [20], suggesting a reduced DNA-discriminatory function also in a in vivo situation.

These data indicate that HDs with the pair Lys^{50}-Met^{54} are not able to discriminate different DNA sequences as efficiently as those possessing other combinations and, therefore, may provide the explanation why this combination is never present in natural HDs [3]. Our sequence selection experiment strongly suggests that the presence of Met^{54} precludes an efficient base-discriminating function by Lys^{50}. Structural studies are required to understand the molecular details of this epistatic effect. The concept that, during evolution, interferences on the DNA-binding function may have limited the combinations of contacting amino acids is strengthened by the recent finding of a network of covariant amino acids, which includes residues at positions 50 and 54, on the outside surface of HD-recognition helix [21].

Acknowledgements: We acknowledge M. Müller and D. Wilson for providing the expression vectors encoding Antp and Gsc HDs. We thank Andrea Damante for the help during protein expression and purification. This work is supported by a grant of the Associazione Italiana Ricerca sul Cancro (A.I.R.C.) to G.D. L.P. is supported by the Ph.D. program 'Diagnosi e terapie cellulari e molecolari' of Udine University.

References

[1] U. Gobel, C. Sander, R. Schneider, A. Valencia, Proteins Struct Funct Genet 18 (1994) 309–317.
[2] W.J. Gehring, Science 236 (1987) 1245–1252.
[3] W.J. Gehring, M. Affolter, T. Bürglin, Annu Rev Biochem 63 (1994) 487–526.
[4] W.J. Gehring, Y.Q. Qian, M. Billeter, K. Furukubo-Tokunaga, A.F. Schier, D. Resendez-Perez, M. Affolter, G. Otting, K. Wüthrich, Cell 78 (1994) 211–223.
[5] S.D. Hanes, R. Brent, Science 251 (1991) 426–429.
[6] G. Damante, L. Pellizzari, G. Esposito, F. Fogolari, P. Viglino, D. Fabbro, G. Tell, S. Formisano, R. Di Lauro, EMBO J 15 (1996) 4992–5000.
[7] C. Wolberger, A.K. Vershon, B. Liu, A.D. Johnson, C.O. Pabo, Cell 67 (1991) 517–528.
[8] M. Billeter, P. Günter, P. Luginbühl, K. Wüthrich, Cell 85 (1996) 1057–1065.
[9] M. Müller, M. Affolter, W. Leupin, G. Otting, K. Wuthrich, W.J. Gehring, EMBO J 7 (1988) 4299–4304.
[10] D. Wilson, G. Sheng, T. Lecuit, N. Dostatni, C. Desplan, Genes Dev 7 (1993) 2120–2134.
[11] S.N. Ho, H.D. Hunt, R.M. Horton, J.K. Pullen, L.R. Pease, Gene 77 (1989) 51–60.
[12] F.W. Studier, B.A. Moffat, J Mol Biol 189 (1986) 113–130.
[13] D. Fabbro, G. Tell, A. Leonardi, L. Pellizzari, C. Pucillo, R. Lonigro, S. Formisano, G. Damante, Nucleic Acids Res 24 (1996) 3283–3288.
[14] G. Scatchard, Ann NY Acad Sci 51 (1949) 660–672.
[15] M. Affolter, A. Percival-Smith, M. Müller, W. Leupin, W.J. Gehring, Proc Natl Acad Sci USA 87 (1990) 4093–4097.
[16] T.K. Blackwell, H. Weintraub, Science 250 (1990) 1104–1110.
[17] B. Blumberg, C.V. Wright, E.M. De Robertis, K.W. Cho, Science 253 (1991) 194–196.
[18] C.R. Kissinger, B. Liu, E. Martin-Blanco, T.B. Kornberg, C.O. Pabo, Cell 63 (1990) 579–590.
[19] A. Percival-Smith, M. Müller, M. Affolter, W.J. Gehring, EMBO J 9 (1990) 3967–3974.
[20] A.F. Schier, W.J. Gehring, Proc Natl Acad Sci USA 90 (1993) 1450–1454.
[21] N.D. Clarke, Protein Sci 4 (1995) 2269–2278.