Title
A phage display selection of engrailed homeodomain mutants and the importance of residue Q50

Journal
Nucleic Acids Research, 32(12)

ISSN
0305-1048

Authors
Simon, Matthew D.
Sato, Ken
Weiss, Gregory A.
et al.

Publication Date
2004-07-09

Published online July 9, 2004

Nucleic Acids Research, 2004, Vol. 32, No. 12 **3623–3631**
doi:10.1093/nar/gkh690

A phage display selection of engrailed homeodomain mutants and the importance of residue Q50

Matthew D. Simon[1], Ken Sato[2], Gregory A. Weiss[2] and Kevan M. Shokat[1,3,*]

[1]Department of Chemistry, University of California, Berkeley, CA 94720, USA, [2]Department of Chemistry, University of California, Irvine, CA 92697, USA and [3]Department of Cellular and Molecular Pharmacology, University of California, San Francisco, CA 94143, USA

Received May 7, 2004; Revised and Accepted June 18, 2004

ABSTRACT

Mutants of engrailed homeodomain (HD) that retain DNA-binding activity were isolated using a phage display selection. This selection was used to enrich for active DNA-binding clones from a complex library consisting of over a billion members. A more focused library of mutant homeodomains consisting of all possible amino acid combinations at two DNA-contacting residues (I47 and Q50) was constructed and screened for members capable of binding tightly and specifically to the engrailed consensus sequence, TAATTA. The isolated mutants largely recapitulated the distribution of amino acids found at these positions in natural homeodomains thus validating the *in vitro* selection conditions. In particular, the unequivocal advantage enjoyed by glutamine at residue 50 is surprising in light of reports that minimize the importance of this residue. Here, the subtle contributions of residue Q50 are demonstrated to play a functionally important role in specific recognition of DNA. These results highlight the complex subtlety of protein–DNA interactions, underscoring the value of the first reported *in vitro* selection of a homeodomain.

INTRODUCTION

The complexity of protein–DNA interfaces presents a challenge to protein engineering; no simple set of rules is available to predictably alter the DNA-binding activity of a given transcription factor (1). Despite this limitation, transcription factors with altered DNA-binding behavior have been generated by selecting the desired mutants from large libraries of mutant transcription factors (2). For example, phage display systems have been developed to isolate mutant zinc finger transcription factors with altered DNA-binding specificities (2–4). From these selections, mutant zinc fingers have been isolated that bind selectively to desired DNA sequences (5). These engineered transcription factors hold promise for numerous biomedical applications [see Jamieson *et al*. (6) and references therein] and provide valuable insight into the subtle and complex rearrangements possible at the protein–DNA interface (7).

Despite the successful engineering of zinc fingers, *in vitro* selection systems for most other important families of DNA-binding domains, including the homeodomains (HDs), have not been described. This may be due to the relatively high affinity of zinc fingers (K_d values on the order of 10 pM) compared with other DNA-binding domains (e.g. homeodomains have K_d values on the order of 1 nM). Homeodomains, the DNA-binding domains of the homeotic proteins, play an important role in the development of all metazoans (8). Over 1000 protein sequences of HDs are currently available, of which over 100 are from humans, and some mutations in these human HDs are known to cause disease (9). In order to facilitate the study and engineering of this important family of DNA-binding domains, we have developed a phage display system to select functional mutants of engrailed HD (EnHD).

Structural (10–14) and biochemical (15–17) studies of HDs reveal that these domains fold into a highly conserved structure composed of three alpha-helices in which the C-terminal helix makes sequence-specific contacts in the major groove of DNA (18). The other two helices pack against the third, and the N-terminus of the HD makes additional important contacts in the DNA minor groove (Figure 1). EnHD, the prototypical and widely studied HD, binds to the consensus sequence TAATTA (15,19). In the major groove, base-specific contacts are made primarily by residues I47, Q50 and N51 (10,11,20). Residue N51 is highly conserved though all HDs and the N51A mutation in EnHD abrogate binding to DNA (15). The contributions of Q50 and I47 to binding are less pronounced; I47A and Q50A mutations lead to ~20- and ~2-fold reductions in binding affinity, respectively (15,16). Despite their relatively small contributions to binding affinity, residues I47 and Q50 remain important to HD–DNA recognition. For example, the subtle I47L missense mutation in the HD of HOXD13 causes limb malformation in humans (21).

The elusive role of homeodomain residue Q50 has received considerable attention (14,15,17,22–26). Early investigations focused on a Q50K mutation that alters homeodomain DNA-binding specificity from TAATTA to TAATCC (23–26). Surprisingly, alanine-scanning experiments revealed that the EnHD Q50A mutant has only a slight loss of DNA-binding affinity and specificity compared with the wild-type protein (15). Furthermore, recent structural comparisons between the Q50 (WT) and Q50A crystal structures demonstrate only subtle changes at the protein–DNA interface (22). Together, these data suggest that Q50 plays only a modest role in

*To whom correspondence should be addressed. Tel: +1 415 514 0472; Fax: +1 415 514 0822; Email: shokat@cmp.ucsf.edu

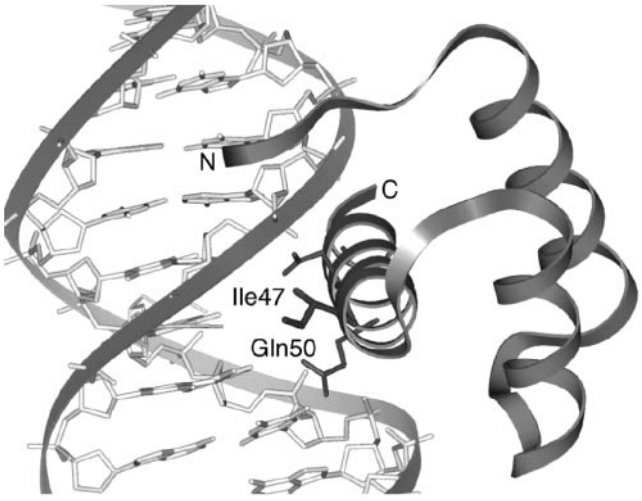

Figure 1. Ribbon representation of EnHD bound to TAATTA based on the reported crystal structure (11). Residues labeled are those examined by phage display in this study (I47 and Q50). Residue N51 is also shown.

determining the sequence affinity and specificity of EnHD. These results are in contrast to the observed strong conservation of Q50 in naturally occurring homeodomains, and its relationship, if any, to DNA-binding affinity and specificity.

To directly probe the amino acid tolerance of EnHD at positions 47 and 50, we developed an *in vitro* selection capable of selecting active homeodomains from pools of mutants. This selection is based on the display of EnHD on the P8 protein of M13 phage particles. Mutagenesis and selection conditions were developed that allow selective purification of EnHD mutants retaining target-specific DNA-binding activity. Surprisingly, the isolated mutants largely recapitulated the distribution of amino acids found at these positions in natural homeodomains, demonstrating that the functional importance of these residues is encoded directly in the DNA–protein interaction. Given that biochemical data on Q50 has suggested that this residue plays a minor role in DNA binding, these results suggest that *in vitro* selections may be able to detect subtle residue contributions that are also captured by evolution but masked in typical biochemical measurements. The sensitivity of this phage display system to subtle, functionally important properties of the EnHD makes it a valuable groundwork toward the engineering of the EnHD–DNA interface.

MATERIALS AND METHODS

Plasmid construction

Residues 1 to 60 of EnHD were amplified from pMal-En (17) using primers with the sequence 5′-GGCTATCGGATGCATCGGACGAGAAGCGTCCAC-3′ (EnHD-P8 without hexa-histidine) or 5′-GGCTATCGGATGCATCGCACCATCACCATCACCATGACGAGAAGCGTCCAC-3′ (for H_6-EnHD-P8), and 5′-TTTTTTCTTAAGCGGATTTTTGGAGCCCGTCGACTTCTTGATCTT-3′, the product was digested with NsiI and AflII and then ligated into the phagemid pSO1148a (27). The stop template was constructed using QuikChange mutagenesis with the primer 5′-AAGATCTGGTTCCAGTAATAACGGGCCAAGATCAAG-3′ and its complement. To construct EnHD-P3 and H_6-EnHD-P3, the respective P8 constructs were subjected to Kunkel mutagenesis (28) with primer 5′-GCGGGCCAAGATCAAGAAGTCGACGGATTTTGATTATGAAAAGATGGC-3′.

Phage production

To obtain purified phage stocks, phages were propagated and precipitated essentially as described previously (29). Briefly, the phagemid constructs described above were transformed into XL1-Blue Subcloning Efficiency Cells and plated on LB Agar supplemented with 100 μg/ml ampicillin. Individual colonies were picked and inoculated into 1 ml of 2YT media supplemented with ampicillin and tetracycline. After 6–8 h shaking at 37 °C, K07 helper phage were added to 10^{10} phage per milliliter. After 20 min the cultures were diluted into at least 30 ml of 2YT media supplemented with ampicillin and subsequently incubated with shaking over night. The bacteria were pelleted by centrifugation (10 min, 4°C, 16 000 *g*). To the supernatant, 1/5 volume of 20% PEG/2.5 M NaCl solution was added. After 5 min of incubation at room temperature, the phage particles were pelleted by centrifugation (10 min, 4°C, 16 000 *g*). The supernatant was decanted and the pellet was further dried by centrifugation (5 min, 4°C, 16 000 *g*). The pellet was resuspended in 1/20 volume stock buffer (100 mM KCl, 20 mM HEPES–KOH pH 7.4, 5 mM $MgCl_2$) supplemented with protease inhibitors (Complete Mini, EDTA Free, Roche). The phage solution was cleared by centrifugation (5 min, 4°C, 27 000 *g*) and the phage were either used immediately or stored at −80 °C for later use. Typically, this protocol yields titers above 10^{13} infectious particles per milliliter.

Analysis of phage DNA by agarose gel electrophoresis

To analyze the ssDNA encapsulated by the phage particles, 10 μl of phage stock was added to 10 μl of DNA Loading Buffer (8 M urea, 20 mM EDTA, 5 mM Tris–HCl pH 7.4) and the solution heated to 95°C for 2 min. After cooling to room temperature, the sample was electrophoresed on a 1% agarose gel supplemented with ethidium bromide.

Phage ELISA

DNA-binding behavior of individual phage populations was examined using an ELISA in which phage bound to DNA-coated wells were detected using an HRP-conjugated anti-M13 antibody. Polysorp 96-well plates (NUNC) were coated with neutravidin (Pierce, 5 μg/ml, 0.1 M $NaHCO_3$, 100 μl per well). The plates were incubated overnight at 4°C with rocking. The neutravidin solution was decanted, and the plates rinsed three times with wash buffer (100 mM KCl, 20 mM HEPES pH 7.4, 5 mM $MgCl_2$, 0.05% Tween-20) using a wash bottle and tapping the plate on paper towels to remove residual buffer between each rinse. The plates were then blocked for at least 1 h with 10 mg/ml BSA in stock buffer. Alternatively, pre-blocked, neutravidin-coated plates were purchased from Pierce and used without further treatment. Biotinylated-DNA solutions ($TAATTA_{20}$ = dsCGCAGTGTAATTACCTCGAC-biotin or $TATATA_{20}$ = dsCGCAGTGTATATACCTCGAC-biotin) were added to each well (100 μl per well, 30 nM DNA, in stock buffer) and incubated for 1 h at room temperature. The DNA solutions were decanted and the wells rinsed with wash

buffer (3×) as above. Phage solutions were added immediately after rinsing.

Individual colonies of interest were grown at 37°C with shaking in 1 ml of 2YT with ampicillin and tetracycline for 6–8 h in individual wells of a 2 ml 96-well assay block (Costar). K07 helper phage were added to a concentration of 10^{10} phage per milliliter and the blocks were incubated for another 24 h at 37°C. The bacteria were removed by centrifugation (15 min, rt, 5000 *g*) and the supernatant was combined with 1/10 volume of 10× T (1 M KCl, 200 mM HEPES pH 7.4, 50 mM $MgCl_2$, 0.15% Tween-20) in a V-bottom 96-well plate. These samples were mixed extensively before addition to the DNA-coated ELISA plate described above. After addition of the phage samples, the plates were incubated with rocking at room temperature for 2 h. The phage solutions were decanted, and the plate rinsed with wash buffer (5×) as above.

A 1:10 000 dilution of HRP-conjugated anti-M13 antibody (Amersham Life Science, 150 µl per well) was added in antibody buffer (stock buffer with 5% glycerol and 0.1 mg/ml BSA) and the plates were incubated with shaking at room temperature for 30 min. The antibody solution was decanted, and the plates were rinsed with wash buffer (5×) as before. The plates were developed for at least 10 min using ABTS solution (5 mg 2,2′-Azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) diammonium salt, 22 ml 50 mM citrate pH 5, 36 µl 30% H_2O_2) and the absorbance was recorded at 405 nm.

Assaying purified phage was achieved using an identical protocol except the purified phages were diluted to the desired concentration in 2YT before addition to the 10× T.

Phage titering experiments

Phage binding and recovery efficiencies were examined by capturing DNA-bound phage on magnetic beads, eluting the phage and then titering the eluant. Phage solutions (10 µl, 10^{13} phage particles/ml) were incubated with biotinylated DNA (usually 30 nM) in 100 µl binding buffer (100 mM KCl, 20 mM HEPES–KOH pH 7.5, 5 mM $MgCl_2$, 5% glycerol, 0.1 mg/ml BSA) with or without competitor salmon sperm DNA (500 ng/ml) and supplementary unbiotinylated target DNA. The incubations were mixed gently in 1.5 ml tubes on a roller for 1.5 h and then added to 0.1 mg pre-rinsed magnetic beads (Dynal M270) suspended in 400 µl selection wash buffer (binding buffer with additional 0.1% Triton-X) in a fresh 1.5 ml microcentrifuge tube. The beads were incubated with the phage solution on a roller for 30 min at room temperature. Using a magnetic stand (Dynal), the beads were captured and the buffer removed by aspiration. The beads were rinsed with selection wash buffer (10×), resuspending the beads during each rinse by flicking the tubes removed from the magnetic stand (each rinse requires ~4 min).

The DNA-bound phages were eluted by resuspending the beads in 40 µl DNase I solution (RNase-free DNase I 10 U/µl, Roche, 1:20 in selection wash buffer) for 10 min. The beads were captured and the eluant used to infect 1 ml of rapidly growing XL1-Blue cells (OD_{600} 0.2–1) for 20 min. Serial dilutions (1:5) of the infected cells into LB media were plated on LB–Amp plates in 10 µl spots. After incubation at 37°C overnight, the resulting colonies were counted to determine the titer of the eluant.

The dynamic range of this biopanning protocol was determined by diluting wild-type EnHD bearing phage (H6-EnHD-P8) into K07 helper phage at various ratios and performing the titering experiment described above.

Mutagenesis

Similar to previous work with the parent phagemid (29), a phagemid library was constructed using a modified form of Kunkel mutagenesis (28). CJ236 cells were grown to 0.5 OD_{600} in 2YT media supplemented with chloramphenicol (to select for the F′ episome), and infected with stop-template bearing phage (10^6 phage per milliliter). The resulting cells were titered onto LB–Amp plates. A single colony from a fresh infection was used to inoculate 2YT (1 ml) supplemented with chloramphenicol and ampicillin. After 6 h shaking at 37°C, K07 helper phage (10^{10} particles per milliliter final concentration) were added. After 20 min, the culture was diluted into 2YT supplemented with ampicillin and uridine (0.25 µg/ml). The phage were allowed to propagate overnight at 37°C with shaking, and precipitated as described above. The dU-containing ssDNA was purified from the phage using a Qiaprep Spin M13 kit (Qiagen).

For library A, the modified Kunkel reaction (29) was performed using a primer degenerate at the positions encoding amino acid 47 and 50 (5′-CGAGGCGCAGAT-CAAGNNSTGGTTCNNSAACAAGCGGGCCAAGATC-3′ where N = A, T, C or G, and S = C or G). For maximum transformation efficiency, the covalently closed-circular DNA (CCC-DNA) is commonly electroporated into SS320 cells; however, in this case, due to the small library size (400 members), 10 µl of the reaction was used to transform 100 µl of XL1-Blue subcloning efficiency chemically competent cells (Stratagene). After transformation, the cells were grown at 37°C with shaking in 2YT media supplemented with ampicillin and helper phage (K07, 10^{10} phage per milliliter). The phage library was isolated via precipitation as described above.

For library B, a higher efficiency electroporation protocol using *Escherichia coli* SS320 cells (29) was employed using the primer 5′-GCGAGTTGGGCCTGAACGAGNNSNNS-NNSNNSRSTTGGNTTNNSNNSAAGCGGGCCAAGATCA-AGAAG T-3′ where N = A, T, C or G; R = A or G, and S = C or G.

Selections

Each round of selection was performed following the same protocol as the titering experiment described above (exceptions are noted below). After each round, the infected cells were used to inoculate 2YT media (50 ml) supplemented with ampicillin (50 µg/ml) and helper phage (K07, 10^{10} particles per milliliter). The phages were propagated overnight and isolated by precipitation as described above. Each round of the selection differed as follows. For library A, the first round included salmon sperm DNA (500 ng/ml) and the DNase I elution was allowed to proceed for 5 min. The second round proceeded as the first except that unbiotinylated target DNA (45 nM) was added to each reaction, and the DNase I elution was allowed to proceed for only 1 min before capturing the beads. In the third round, 450 nM unbiotinylated DNA was included and the phages were eluted with DNase I for 10 min. In rounds 2 and 3, to ensure the library was not dominated by non-specific binders, two negative controls were run in parallel

with the desired selection: one control contained a biotinylated, scrambled DNA sequence (30 nM), and the other contained biotin (30 nM) without DNA. Excess phage isolated from each round could be stored at −80°C for later use with only minimal loss of activity or increase in non-specific binding.

Selections with library B were the same as A except that a negative selection was performed by pre-incubating the phage solution (100 µl, approximately 10^{13} phage per milliliter) with magnetic beads (1 mg) overnight at 4°C to remove the tightest bead binders. Furthermore, to maintain the high diversity of the library, the first round was performed in triplicate with less rinsing (only five rinses of the magnetic beads after capturing the phage) and the three pools of selected phage were pooled to increase the overall yield. The next two rounds were carried out as described for library A.

Analysis of selected phage

Individual phage clones from library A after rounds 2 and 3 were analyzed for specific DNA-binding activity by phage ELISA as described above. Binding was compared between the natural EnHD target DNA, $TAATTA_{20}$, and the scrambled version of this DNA, $TATATA_{20}$. Clones that bound to $TAATTA_{20}$, but not $TATATA_{20}$, were sequenced. The EnHD-encoding portion of the selected phage clones was sequenced using two sequential PCR reactions as described previously (30).

Expression of mutant homeodomains

Mutant homeodomains were expressed off-phage as maltose-binding protein fusions, and purified to near homogeneity using amylose resin. Mutations were introduced using either QuikChange or Kunkel mutagenesis. Proteins were expressed and purified from *E.coli* BL21(DE3) cells essentially as described (17). The protein was eluted from the amylose column using elution buffer (200 mM KCl, 20 mM HEPES pH 7.6, 1 mM EDTA, 10 mM maltose and 0.001% NaN_3). Protein concentrations were determined using the calculated extinction coefficient, 74 250 $M^{-1}cm^{-1}$, at 280 nm.

Electrophoretic mobility shift assays

Binding of the protein to DNA off-phage was determined essentially as described (15,17). The DNA probe, $TAATTA_{20}$, was 5′ end-labeled using T4 polynucleotide kinase and [γ-^{32}P]ATP. After annealing, excess nucleotide was removed from the reaction using a Bio-Spin 6 column (Bio-Rad). The radiolabeled probe (1 nM) was combined with increasing concentrations of protein in assay buffer (50 mM KCl, 20 mM HEPES pH 7.6, 1 mM EDTA, 5% glycerol, 0.02% NP-40). The binding reactions were incubated for at least 90 min at rt and were then run on 8% 0.5× TBE polyacrylamide gels (pre-run for at least 30 min at 200 V, 4°C). After loading, the gels were run for 3 min at 200 V, and then the voltage was reduced to 150 V for 20 min. After drying, the gels were exposed to a phosphor screen for at least 30 min and analyzed using a phosphorimager (Molecular Dynamics).

For analysis and comparisons involving the Q50A mutant, rapid loading of the gel was essential for repeatable results. To this end, gels were loaded using a multi-channel pipette and appropriately spaced nine-well combs.

Apparent equilibrium dissociation constants were determined using the equation: $\Theta = B_{max}/(1 + K_d/[EnHD_{mut}])$ where $\Theta = cpm_{bound}/(cpm_{bound} + cpm_{free})$, B_{max} is an adjustable parameter representing the maximum value of Θ at high $[EnHD_{mut}]$. The reported error for each K_d represents the standard error from the curve fit.

To determine the specificity of binding, competition experiments were performed essentially as above except that 50 nM of protein was combined with 1 nM radiolabled DNA and increasing concentrations of salmon sperm DNA. The reactions proceeded for >6 h at 4°C and then warmed to RT before loading.

RESULTS

Development of an engrailed HD-P8 fusion with specific DNA-binding activity

To display EnHD on phage, a phagemid (EnHD-P8) was constructed encoding resides 1 to 60 of EnHD fused to the N-terminus of the major M13 coat protein, P8, via a short linker using a construct previously optimized for high-copy display of other proteins (27). Despite precedent for displaying these residues of EnHD on a different phage system (17), EnHD-P8 did not display significant DNA-binding activity in a phage titering experiment (Figure 2A). The previously reported construct in which EnHD was fused to the minor phage-coat protein, P3, retained DNA-binding activity. This phage construct was used to study homeodomain mutants on a one at a time basis, but no selection was reported. Encouraged by the phage ELISA results from the reported construct (17), we made an analogous phagemid that encoded for EnHD fused to P3 (EnHD-P3). Unfortunately, EnHD-P3 had minimal DNA-binding activity as observed in a phage titering experiment (Figure 2A).

In order to develop a viable phage display construct, modifications to the N-terminus of EnHD were explored. Surprisingly, a hexa-histidine tag rescued the DNA-binding activity for H_6-EnHD-P8 but not for H_6-EnHD-P3 (Figure 2A). A phage ELISA was used to address whether the observed DNA binding was specific for the engrailed target sequence, $TAATTA_{20}$, or rather an artifact of additional positive charge from the hexa-histidine tag. The binding of H_6-EnHD-P8 was highly specific for $TAATTA_{20}$ as compared with a non-target DNA sequence, $TATATA_{20}$, with the same overall base composition (Figure 2B). This confirms that the H_6-EnHD-P8 construct successfully displays EnHD in a viable conformation, enabling target-specific binding to DNA.

Optimization of biopanning conditions

In order to specifically capture phagemids encoding active HDs, a biopanning procedure was developed wherein streptavidin-coated magnetic beads (Dynal M270) were used to capture phage bound to biotinylated DNA. Several experimental parameters were unexpectedly critical to the success of this strategy. Initial attempts using Neutravidin-coated 96-well Polysorp plates in lieu of the magnetic beads did not lead to significant enrichment over background, despite attempts using various buffers, rinses and blocking protocols. Magnetic beads (Dynal 270) proved superior, initially leading to a five-fold

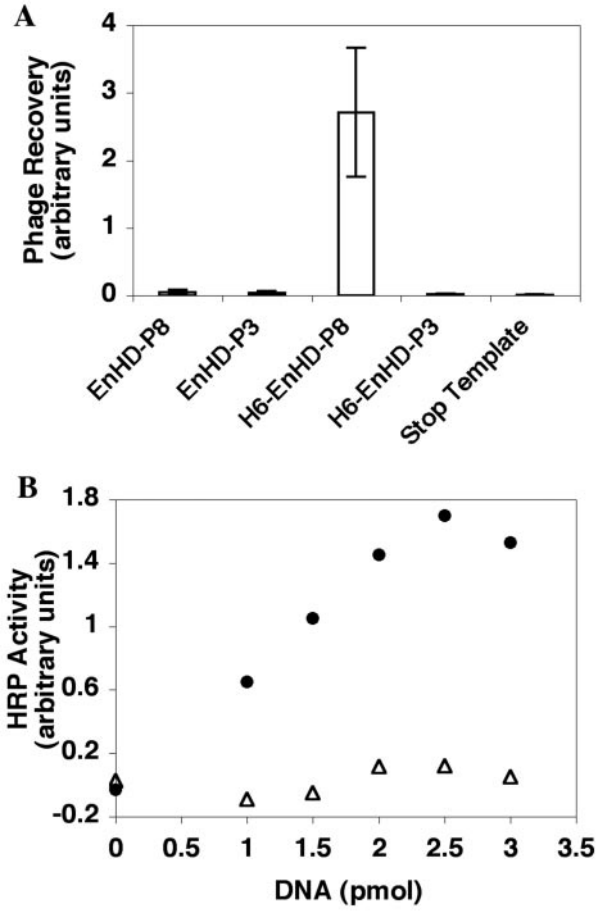

Figure 2. Development of an EnHD phage display construct with specific DNA-binding activity. (**A**) The DNA-binding activity of EnHD phage constructs as examined by phage capture titering experiments measured by the number of colony forming units recovered after elution. (**B**) Binding of H_6-EnHD-P8 to $TAATTA_{20}$ (closed circles) and $TATATA_{20}$ (open triangles).

enrichment relative to a negative control experiment applying identical conditions, but lacking DNA. After optimizing conditions, including biotinylated-DNA concentration, the quantity of beads, buffer composition, rinse time and number, enrichment factors of greater than 500-fold were observed. The inclusion of BSA and glycerol in the assay and wash buffer were essential for successful biopanning; without these additives, wild-type EnHD-displaying phage rinsed off the DNA-coated beads easily (less than 6 rinses), whereas after buffer optimization, the desired phage remained bound despite extensive rinsing (more than 12 rinses). These optimized selection conditions have a dynamic range appropriate for the selection of active EnHD, as demonstrated by titering the phage yields from a single round using phage displaying wild-type protein diluted into helper phage (Figure 3).

Elution of the bound phage could be accomplished with high salt (>1 M KCl), low pH (0.1 M HCl) or DNase I (data not shown). In individual rounds of selection, these different elution conditions lead to similar results. However, through multiple rounds of selection, elution with DNase I was found to be essential to minimize the accumulation of phage that bound to the bead matrix ('bead-binders') rather than DNA (discussed below).

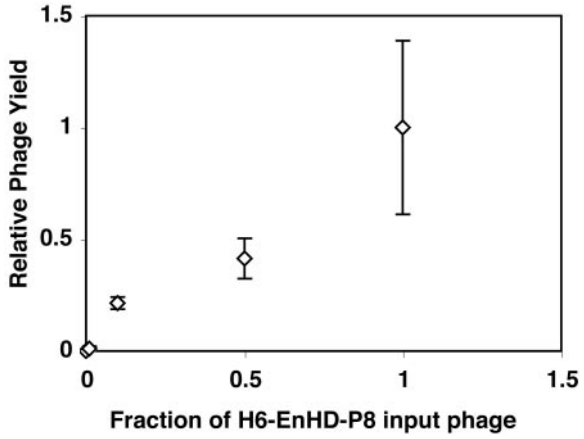

Figure 3. Examination of the dynamic range of a phage pull-down titering experiment. The input phage consisted of H6-EnHD-P8 diluted at various concentrations into helper phage. Yields are reported relative to undiluted H6-EnHD-P8.

Mutagenesis and selection of DNA-binding engrailed HD mutants

In order to probe the protein–DNA interface of EnHD bound to its target DNA, a phage library was constructed with residues 47 and 50 randomized (library A). To construct library A, a modified form of Kunkel mutagenesis (28) was performed. The starting template for this reaction is an inactive template containing two ochre stop codons (TAA) that block the production of active fusion phage protein. This stop template guarantees that any wild-type clones recovered from the selection originate from the mutant pool and are not contamination from the starting template.

Mutagenesis and transformation led to titers of 5.6×10^4 ampicillin-resistant colonies. Ten clones were chosen at random and sequenced. Out of the sequenced clones, 30% encoded mutant EnHD ORFs (N47;V50, G47;Q50 and F47;L50), whereas the rest contained stop codons from the starting template. At this level of mutagenesis efficiency, the titer is sufficient to include on average approximately 40 copies of each member (with all possible combinations of the 20 amino acids at both positions) suggesting that a high degree of coverage was achieved in this library.

Initial attempts to select active DNA binders using a highly complex library (more than 10^9 members, library B) of engrailed mutants displayed on phage led instead to an accumulation of 'bead binding' clones (results not shown). To suppress the accumulations of these clones, an elution strategy was developed in which the DNA-bound phages were selectively eluted by hydrolyzing the biotinylated target DNA with DNase I. Under these conditions, undesirable phage that bind to the beads independently of the DNA should not elute, and will not be carried onto the next round of the selection. Indeed, the use of DNase I led to selective enrichment of DNA-binding phage after three rounds of selection using both library A (>1000-fold enrichment) and library B (>500-fold enrichment) as reported in Table 1. These phages bound to the desired DNA sequence ($TAATTA_{20}$) but not a scrambled sequence ($TATATA_{20}$) or biotin and beads alone. Both selections included high concentrations of salmon sperm DNA to ensure that only sequence-specific binders were enriched.

Table 1. Phage yields[a] from each round of selection against $TAATTA_{20}$

	$TAATTA_{20}$	$TATATA_{20}$	No DNA[b]
Library A			
Round 1	4.4×10^6	ND	6.0×10^5
Round 2	1.5×10^6	7.3×10^5	5.3×10^5
Round 3	1.1×10^8	1.5×10^5	7.5×10^3
Library B			
Round 1	8.8×10^4	ND	1.1×10^5
Round 2	4.7×10^4	1.0×10^4	8.8×10^4
Round 3	7.8×10^6	3.8×10^4	1.0×10^4

The selection stringency was increased in each round. For experimental details see materials and methods.
[a]Extrapolated from the number of ampicillin resistant colonies recovered.
[b]Biotin alone was used as a negative control. ND = Not determined.

Table 2. Sequences of clones from library A that bind specifically to $TAATTA_{20}$

Number[a]	Sequence
Wt	...I K **I**$_{47}$ W F **Q**$_{50}$ N...
13	...I K **I**$_{47}$ W F **Q**$_{50}$ N...
7	...I K **N**$_{47}$ W F **Q**$_{50}$ N...
7	...I K **V**$_{47}$ W F **R**$_{50}$ N...
5	...I K **V**$_{47}$ W F **Q**$_{50}$ N...
5	...I K **T**$_{47}$ W F **Q**$_{50}$ N...
4	...I K **F**$_{47}$ W F **Q**$_{50}$ N...
2	...I K **T**$_{47}$ W F **A**$_{50}$ N...
1	...I K **I**$_{47}$ W F **M**$_{50}$ N...
1	...I K **T**$_{47}$ W F **M**$_{50}$ N...

[a]Number of isolates of each clone.

Analysis of engrailed HD mutants

In order to identify active DNA-binding clones from enriched library A, phages were grown from individual colonies and examined for specific DNA-binding using a phage ELISA. Out of 92 clones tested from the second and third rounds, 53 bound specifically to the desired $TAATTA_{20}$ sequence (as compared with binding to the scrambled version, $TATATA_{20}$). Most of the remaining clones displayed low signal for both targets. Only one clone (Y47;Q50) was a non-specific binder, showing significant signal for both sequences. The clones that demonstrated the desired specific DNA-binding activity in the spot assay were sequenced. Out of the 53 clones, 45 gave unambiguous sequencing results (the remaining 8 primarily suffered from cross-contamination), and are listed in Table 2. All of the clones sequenced from library A contained mutations only at positions 47 and 50, as programmed in the original library.

Similarly, after three rounds of selection and amplification, individual clones from library B were assayed for binding to $TAATTA_{20}$ and $TATATA_{20}$ by phage ELISA. Many of the clones bound specifically to $TAATTA_{20}$; the remaining clones yielded low signal for binding to both DNA sequences. Out of the specific binders, sequences for several of the mutants were determined and listed in Table 3.

Comparison of natural HDs with the selected EnHD mutants

To compare the pressures imposed by this selection with those of natural evolution, the sequences of the clones recovered from library A were compared with the sequences of natural HDs in The Homeodomain Resource (31,32). To avoid biases associated with disproportionately represented orthologs in the HD Resource, we chose to compare the selected mutants to the 129 human HDs previously examined by Banerjee-Basu *et al*. (33). The majority of sequences recovered were either similar to, or identical to the natural HDs (Figure 4); the two most frequently encountered residues at position 47 are valine and isoleucine, both from our selection and in human HD proteins (compare Figure 4A and C). The third most common mutant found at this position, I47N, also displays significant representation in the natural homeodomains. In the recovered clones, position 50 was predominantly glutamine, as is most common in natural HDs (compare Figure 4B with D). These results support the conclusion that the 47I/V and 50Q are the optimal combination for binding $TAATTA_{20}$ in this sequence context. However, it is notable that some of the less abundant residues found in our selection (e.g. T47 and F47) also have precedent in natural HDs.

Table 3. Sequences of clones from library B that bind specifically to $TAATTA_{20}$[ab]

...N$_{41}$ E A Q I K I W F Q N K R A K$_{55}$...
...N$_{41}$ E A Q I K **T** W F Q N **T** R A K$_{55}$...
...N$_{41}$ **A** A Q I K **V** W F Q N K R A K$_{55}$...
...N$_{41}$ E A Q I K **N** W F Q N **T** R A K$_{55}$...
...N$_{41}$ E A Q I K **T** W F Q N K R A K$_{55}$...

[a]Mutations are shown in bold.
[b]The high proportion of mutations at position 47 is due to the codon (RST) employed in the mutagenesis.

The residues that were not recovered from this selection are also noteworthy. There are many natural HDs with a lysine at residue 50. However, Q50K engrailed is known to bind TAATCC preferentially over TAATTA so the absence of these mutants in the recovered clones was expected. It is similarly encouraging that the conservative mutation, I47L, was not recovered. It is known that, in the context of HOX11D, a I47L mutation weakens DNA-binding activity *in vitro* and causes developmental defects in humans (21). Therefore, we conclude that recovery of the expected mutants and absence of others validates the functional significance of our selection conditions.

Residue 50 and its contribution to DNA binding

Residue Q50 has attracted the attention of many researchers (see Grant *et al*. (22) and references therein), partially because a Q50K mutation binds to TAATCC instead of the native TAATTA, suggesting this residue may play an important role recognizing bases 5 and 6 of the consensus site. Given the reported role of Q50 in sequence specificity, it was surprising that Q50A also displays preference for TAATTA (16), thus challenging notions that the interactions of the glutamine side chain are of primary importance. However, the preponderance of Q50 in our recovered clones demonstrates that Q50 is the optimal amino acid for binding to $TAATTA_{20}$ under our selection conditions.

Despite the prevalence of clones containing Q50 that were recovered in this selection, one frequently encountered clone

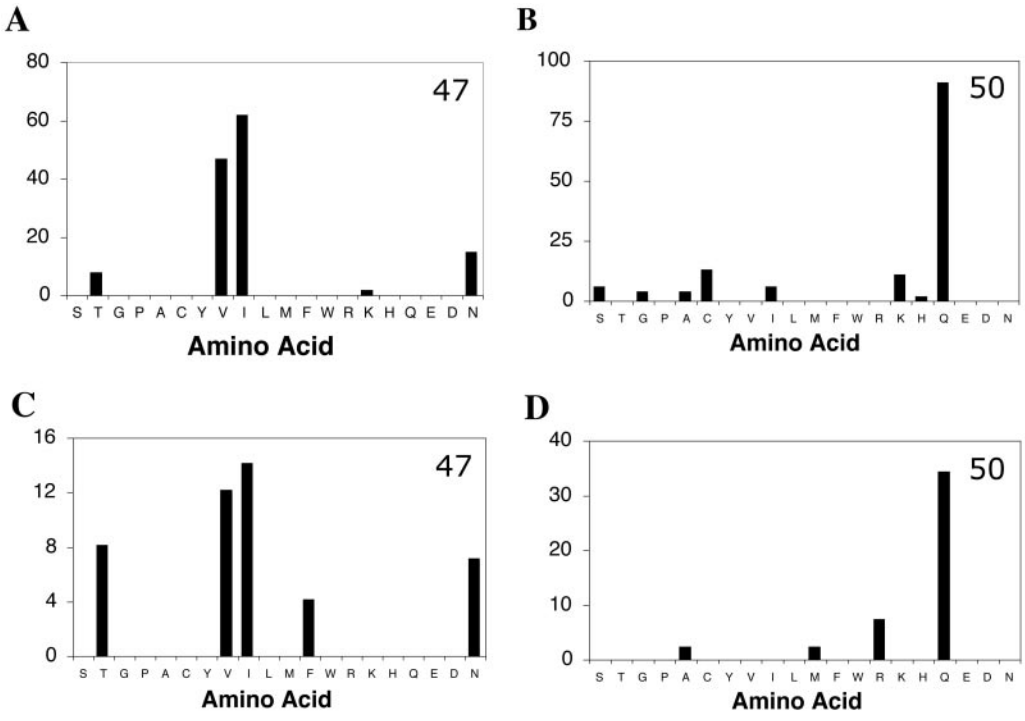

Figure 4. Amino-acid distribution in 129 human HDs (33) found (**A**) at residue 47 and (**B**) at residue 50, compared with the distribution of 45 clones recovered from a phage selection of EnHD (**C**) at residue 47 and (**D**) at residue 50.

(7/45) contained Q50R instead (I47V;Q50R). It is surprising that a mutant with a bulky arginine side chain at position 50 would specify a preference for the final 2 bp of the target DNA. Using an electrophoretic mobility shift assay (EMSA) the specificity and affinity of the mutant were demonstrated to be similar to wild-type EnHD (Figure 5A). The I47V;Q50R mutant has an apparent K_d of 18 ± 2 nM (compared to wild type with an apparent *K*d of 7 ± 1 nM under the same EMSA conditions). Attempts to analyze the single mutant with only Q50R (without I47V) were complicated by poorly behaved shifts (data not shown) possibly caused by higher order binding of the mutant to the probe. Structural studies will be necessary to elucidate the role of R50.

It is notable that the EnHD Q50A single mutant, despite its reasonably high affinity and specificity for TAATTA-containing sites, was not recovered in this selection. The only Q50A mutant recovered from this selection also contained I47T (Table 2). Similarly, in nature, homeodomains with A50 are very rare; one of the few exceptions, exemplified by IRX-2 in humans, has I47T in addition to Q50A. Thus our selection experiments support the finding from HD sequence analysis that I47T may act as a second site suppressor for the Q50A mutation.

To explore the biochemical consequences of these mutations singly and in combination, the DNA-binding activities of the Q50A and the I47T;Q50A mutants were compared by EMSA (Figure 5B and C). Rapid-loading EMSA conditions were required to ensure repeatable shifts of the Q50A mutant. Under these conditions, wild-type EnHD binds with an apparent K_d of 5 ± 1 nM. As previously reported (15), the Q50A single mutant binds with ~2-fold decreased affinity compared to wild type (apparent K_d of 10 ± 3 nM). The I47T;Q50A mutant recovered from our selection binds more tightly than EnHD Q50A, and nearly as well as wild type (apparent K_d of 6 ± 1 nM). Both mutants (Q50A and I47T;Q50A) bind to $TAATTA_{20}$ with similar specificity (only 4-fold below WT), as demonstrated in an EMSA experiment with increasing concentrations of competitor salmon sperm DNA (Figure 5C).

How does I47T compensate for the Q50A mutation? In a crystal structure of the Q50A mutant, ordered water molecules produce a clathrate-like structure in the space created by the missing carboxamide-bearing side chain of Q50 (22). It is tempting to speculate that the hydroxyl group from I47T may anchor the local water structure in a fashion that compensates for the loss of subtle yet important Q50-DNA contacts. These results demonstrate how *in vitro* selections can uncover unexpected mutants that invite further study.

DISCUSSION

Despite many advances in biochemical, structural and computational approaches, even the best understood protein–DNA interactions have proven resistant to simple predictive models. Here, using an *in vitro* selection, we have illustrated how a residue (Q50) that appears unimportant by structural and biochemical characterization, in fact has a significant functional role. The significance of this selective advantage is supported by previous studies with Pho2, the engrailed homolog in yeast. It was found that Pho2 Q50A binds to an upstream promoter site at ~70% of wild-type levels *in vitro* but only activated expression from this promoter at ~5% of wild-type levels *in vivo* (34). As found for EnHD in the current study, it appears the importance of Q50 in Pho2 is greater than expected from biochemical experiments alone. These findings challenge

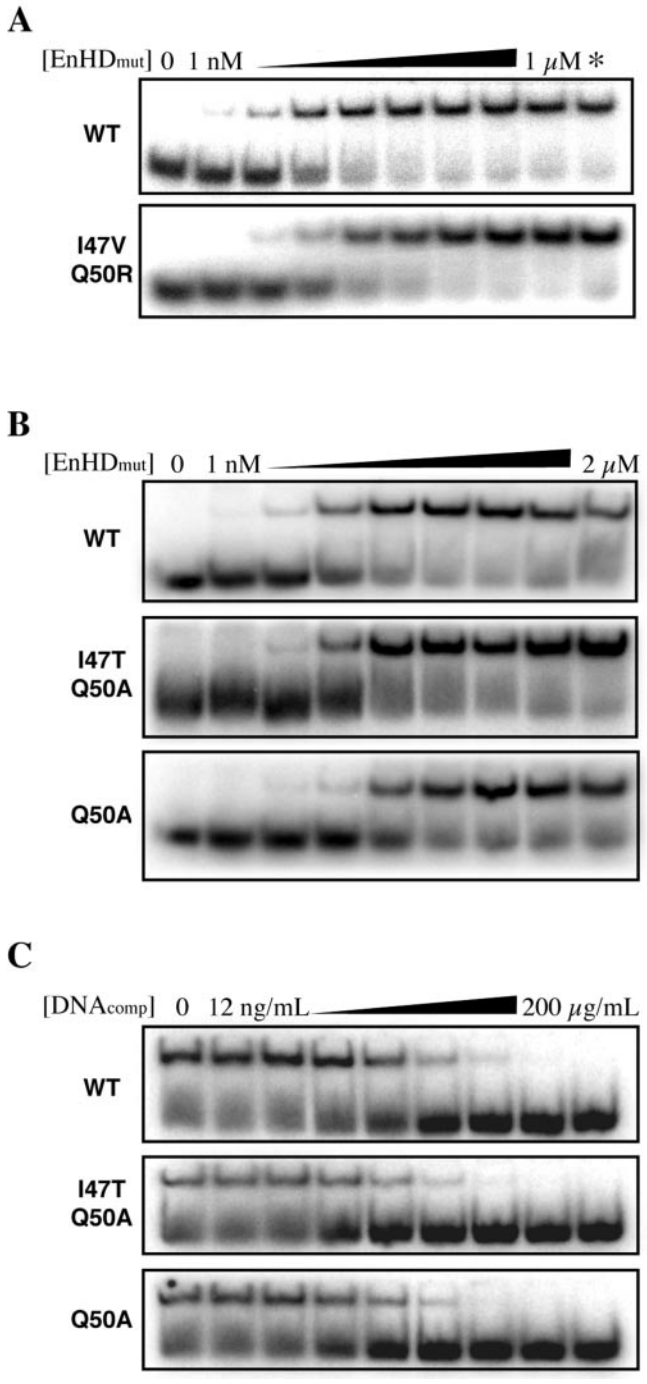

Figure 5. Examination of the affinity (**A** and **B**) and specificity (**C**) of various EnHD mutants. (A) EMSA comparing the affinity between the EnHD I47V;Q50R and EnHD WT. The asterisk indicates addition of excess competitor salmon-sperm DNA to demonstrate the specificity of binding. (B) EMSA comparing the affinities of EnHD WT, EnHD I47T;Q50A and EnHD Q50A binding to $TAATTA_{20}$. (C) EMSA competition experiment with increasing concentrations of salmon sperm DNA comparing the specificities of binding. Experiments in (B) and (C) were performed under rapid loading conditions.

our modest understanding of protein–DNA interactions; alterations with effects well under an order of magnitude (such as the change in affinity and specificity observed for EnHD Q50A) can clearly have important functional consequences.

We were able to examine the importance of these subtle effects in EnHD by developing a robust phage display selection capable of isolating active HD mutants. This selection is not limited to small libraries—a similar selection using a highly complex library (1.5×10^9 unique transformants) led to enrichment factors of >500-fold (Table 1, library B). In light of the results from this more complex library, this work serves as the foundation to study HDs with rapid functional epitope mapping techniques, such as shotgun scanning (30,35,36). Furthermore, this phage display system has been instrumental toward isolating HDs that bind specifically to unnatural, synthetic nucleotides (37). We conclude that the selection system reported herein provides powerful new opportunities for the study and engineering of this important class of transcription factors, the homeodomains.

ACKNOWLEDGEMENTS

We would like to thank Professor Christopher S. Francklyn for generously providing the pMal-En construct and Zack Knight for helpful comments on this manuscript. This work was supported by graduate fellowships from the Organic Division of the American Chemical Society and the National Science Foundation to M.D.S. and NIH RO1 EB-001987 to K.M.S.

REFERENCES

1. Pabo,C.O. and Nekludova,L. (2000) Geometric analysis and comparison of protein–DNA interfaces: why is there no simple code for recognition? *J. Mol. Biol.*, **301**, 597–624.
2. Pabo,C.O., Peisach,E. and Grant,R.A. (2001) Design and selection of novel Cys(2)His(2) zinc finger proteins. *Annu. Rev. Biochem.*, **70**, 313–340.
3. Segal,D.J., Dreier,B., Beerli,R.R. and Barbas,C.F. (1999) Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5′-GNN-3′ DNA target sequences. *Proc. Natl Acad. Sci. USA*, **96**, 2758–2763.
4. Greisman,H.A. and Pabo,C.O. (1997) A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. *Science*, **275**, 657–661.
5. Isalan,M., Klug,A. and Choo,Y. (2001) A rapid, generally applicable method to engineer zinc fingers illustrated by targeting the HIV-1 promoter. *Nat. Biotechnol.*, **19**, 656–660.
6. Jamieson,A.C., Miller,J.C. and Pabo,C.O. (2003) Drug discovery with engineered zinc-finger proteins. *Nature Rev. Drug Discov.*, **2**, 361–368.
7. Miller,J.C. and Pabo,C.O. (2001) Rearrangement of side-chains in a zif268 mutant highlights the complexities of zinc finger-DNA recognition. *J. Mol. Biol.*, **313**, 309–315.
8. Gehring,W.J., Affolter,M. and Burglin,T. (1994) Homeodomain Proteins. *Annu. Rev. Biochem.*, **63**, 487–526.
9. D'Elia,A.V., Tell,G., Paron,I., Pellizzari,L., Lonigro,R. and Damante,G. (2001) Missense mutations of human homeoboxes: a review. *Human Mutat.*, **18**, 361–374.
10. Kissinger,C.R., Liu,B.S., Martinblanco,E., Kornberg,T.B. and Pabo,C.O. (1990) Crystal-structure of an engrailed homeodomain–DNA complex at 2.8-a resolution—a framework for understanding homeodomain–DNA interactions. *Cell*, **63**, 579–590.
11. Fraenkel,E., Rould,M.A., Chambers,K.A. and Pabo,C.O. (1998) Engrailed homeodomain-DNA complex at 2.2 angstrom resolution: a detailed view of the interface and comparison with other engrailed structures. *J. Mol. Biol.*, **284**, 351–361.
12. Li,T., Stark,M.R., Johnson,A.D. and Wolberger,C. (1995) Crystal-structure of the Mata1/Mat-Alpha-2 homeodomain heterodimer bound to DNA. *Science*, **270**, 262–269.
13. Piper,D.E., Batchelor,A.H., Chang,C.P., Cleary,M.L. and Wolberger,C. (1999) Structure of a HoxB1-Pbx1 heterodimer bound to DNA: role of

the hexapeptide and a fourth homeodomain helix in complex formation. *Cell*, **96**, 587–597.
14. TuckerKellogg,L., Rould,M.A., Chambers,K.A., Ades,S.E., Sauer,R.T. and Pabo,C.O. (1997) Engrailed (Gln50→Lys) homeodomain–DNA complex at 1.9 angstrom resolution: structural basis for enhanced affinity and altered specificity. *Structure*, **5**, 1047–1054.
15. Ades,S.E. and Sauer,R.T. (1994) Differential DNA-binding specificity of the engrailed homeodomain—the role of residue-50. *Biochemistry*, **33**, 9187–9194.
16. Ades,S.E. and Sauer,R.T. (1995) Specificity of minor-groove and major-groove interactions in a homeodomain–DNA complex. *Biochemistry*, **34**, 14601–14608.
17. Connolly,J.P., Augustine,J.G. and Francklyn,C. (1999) Mutational analysis of the engrailed homeodomain recognition helix by phage display. *Nucleic Acids Res.*, **27**, 1182–1189.
18. Gehring,W.J., Qian,Y.Q., Billeter,M., Furukubotokunaga,K., Schier,A.F., Resendezperez,D., Affolter,M., Otting,G. and Wuthrich,K. (1994) Homeodomain–DNA recognition. *Cell*, **78**, 211–223.
19. Draganescu,A. and Tullius,T.D. (1998) The DNA binding specificity of engrailed homeodomain. *J. Mol. Biol.*, **276**, 529–536.
20. Clarke,N.D., Kissinger,C.R., Desjarlais,J., Gilliland,G.L. and Pabo,C.O. (1994) Structural studies of the engrailed homeodomain. *Protein Sci.*, **3**, 1779–1787.
21. Caronia,G., Goodman,F.R., McKeown,C.M.E., Scambler,P.J. and Zappavigna,V. (2003) An 147L substitution in the HOXD13 homeodomain causes a novel human limb malformation by producing a selective loss of function. *Development*, **130**, 1701–1712.
22. Grant,R.A., Rould,M.A., Klemm,J.D. and Pabo,C.O. (2000) Exploring the role of glutamine 50 in the homeodomain-DNA interface: Crystal structure of engrailed (Gln50→Ala) complex at 2.0 angstrom. *Biochemistry*, **39**, 8187–8192.
23. Percivalsmith,A., Muller,M., Affolter,M. and Gehring,W.J. (1990) The interaction with DNA of wild-type and mutant Fushi-Tarazu Homeodomains. *EMBO J.*, **9**, 3967–3974.
24. Treisman,J., Gonczy,P., Vashishtha,M., Harris,E. and Desplan,C. (1989) A single amino-acid can determine the DNA-binding specificity of homeodomain proteins. *Cell*, **59**, 553–562.
25. Hanes,S.D. and Brent,R. (1989) DNA specificity of the bicoid activator protein is determined by homeodomain recognition helix residue-9. *Cell*, **57**, 1275–1283.
26. Hanes,S.D. and Brent,R. (1991) A genetic model for interaction of the homeodomain recognition helix with DNA. *Science*, **251**, 426–430.
27. Sidhu,S.S., Weiss,G.A. and Wells,J.A. (2000) High copy display of large proteins on phage for functional selections. *J. Mol. Biol.*, **296**, 487–495.
28. Kunkel,T.A., Roberts,J.D. and Zakour,R.A. (1987) Rapid and Efficient Site-Specific Mutagenesis without Phenotypic Selection. *Methods Enzymol.*, **154**, 367–382.
29. Sidhu,S.S., Lowman,H.B., Cunningham,B.C. and Wells,J.A. (2000) Phage display for selection of novel binding peptides. *Methods Enzymol.*, **328**, 333–363.
30. Avrantinis,S.K., Stafford,R.L., Tian,X. and Weiss,G.A. (2002) Dissecting the streptavidin-biotin interaction by phage-displayed shotgun scanning. *ChemBioChem*, **3**, 1229–1234.
31. Banerjee-Basu,S., Sink,D.W. and Baxevanis,A.D. (2001) The Homeodomain Resource: sequences, structures, DNA binding sites and genomic information. *Nucleic Acids Res.*, **29**, 291–293.
32. Banerjee-Basu,S., Moreland,T., Hsu,B.J., Trout,K.L. and Baxevanis,A.D. (2003) The Homeodomain Resource: 2003 update. *Nucleic Acids Res.*, **31**, 304–306.
33. Banerjee-Basu,S. and Baxevanis,A.D. (2001) Molecular evolution of the homeodomain family of transcription factors. *Nucleic Acids Res.*, **29**, 3258–3269.
34. Justice,M.C., Hogan,B.P. and Vershon,A.K. (1997) Homeodomain-DNA interactions of the Pho2 protein are promoter-dependent. *Nucleic Acids Res.*, **25**, 4730–4739.
35. Weiss,G.A., Watanabe,C.K., Zhong,A., Goddard,A. and Sidhu,S.S. (2000) Rapid mapping of protein functional epitopes by combinatorial alanine scanning. *Proc. Natl Acad. Sci. USA*, **97**, 8950–8954.
36. Sato,K., Simon,M.D., Levin,A.M., Shokat,K.M. and Weiss,G.A. (2004) Dissecting the engrailed homeodomain-DNA interaction by phage-displayed alanine shotgun scanning. *Chem. Biol.*, in press.
37. Simon,M.D. and Shokat,K.M. (2004) Adaptability at a protein-DNA interface: re-engineering the engrailed homeodomain to recognize an unnatural nucleotide. *J. Am. Chem. Soc.*, in press.