Chemistry & Biology, Vol. 11, 1017–1023, July, 2004, ©2004 Elsevier Ltd. All rights reserved. DOI 10.1016/j.chembiol.2004.05.008

# Dissecting the Engrailed Homeodomain-DNA Interaction by Phage-Displayed Shotgun Scanning

**Ken Sato,[1,5] Matthew D. Simon,[3] Aron M. Levin,[1] Kevan M. Shokat,[3,4] and Gregory A. Weiss[1,2,*]**
[1]Department of Chemistry
[2]Department of Molecular Biology & Biochemistry
University of California
Irvine, California 92697
[3]Department of Chemistry
University of California, Berkeley
Berkeley, California 94720
[4]Department of Cellular and Molecular Pharmacology
University of California, San Francisco
San Francisco, California 94143

## Summary

**Phage-displayed alanine shotgun scanning was used to dissect contributions by engrailed homedomain (En-HD) residues 17 through 46, which indirectly influence recognition of DNA. The relative contributions of such indirect contacts, quantified by shotgun scanning, highlight previously unexplored En-HD residues. Two motifs dominate En-HD function in this region. First, two surface-exposed aromatic residues (F20 and Y25) bracket the hydrophobic core. Second, two sets of turn-forming residues are highlighted, including carboxamide-requiring residues E22/N23 and a leucine/isoleucine splint. The En-HD hydrophobic core exhibits a surprising degree of malleability, as demonstrated by homolog shotgun scanning. Most selectants from in vitro shotgun scanning mirror the consensus human homeodomain sequence. Thus, natural evolution and in vitro selection use similar selection criteria: affinity, specificity, and stability. However, homolog shotgun scanning identified mutations capable of improving the affinity and specificity of En-HD.**

## Introduction

The engrailed homeodomain (En-HD) offers a model protein-DNA interaction for studying eukaryotic gene regulation and DNA binding specificity by proteins. DNA recognition by homeodomains, including En-HD, has been studied by high-resolution X-ray and NMR structures [1–10], mutational studies [11–13], and in vitro selection experiments [14–16]. En-HD folds into three $\alpha$ helices with an unstructured N-terminal arm. Key to DNA recognition, helix-3 (the "recognition helix") fits into the DNA major groove, while the N-terminal arm interacts with the minor groove (Figure 1A). Most homeodomains bind to similar DNA sites having a core sequence of TAAT. The two base pairs trailing the core sequence can add to the sequence specificity of homeodomains [16]. For example, the consensus binding site of En-HD is the DNA sequence TAATTA [1, 17].

Alanine shotgun scanning generates protein libraries with either alanine or the wild-type amino acid in specific positions [18, 19]. Following selection and screens for functional proteins, in this case sequence-specific DNA binding, individual selectants are sequenced. From the sequencing results, distributions of alanine and wild-type in targeted positions are tabulated to identify positions with strong preferences for the wild-type side chain [19]. Shotgun scanning can identify important structural residues by ranking their relative contributions to protein function. In this report, a magnetic bead-based selection for binding to the biotinylated consensus En-HD dsDNA (TAATTA) sequence [20] was used to identify En-HD variants from two 15-residue alanine shotgun scanning libraries. A secondary screen for binding specificity eliminated selectants capable of also binding to a scrambled dsDNA sequence (TATATA).

En-HD is an attractive system for applying shotgun scanning because, although some aspects are understood, important gaps in our understanding of the En-HD•DNA interaction can be clarified. Previous mutational studies [14, 17] of En-HD interactions with DNA have focused largely on the contributions to DNA recognition by residues from helix-3 and the flexible N-terminal arm. Thus, we used phage-displayed shotgun scanning to dissect contributions to molecular recognition by the less well-characterized En-HD residues 17 through 46. The first library consisted of residues 17 to 31 and the second of residues 32 to 46. Such En-HD residues are expected to contribute directly and indirectly to molecular recognition, either through networks of hydrophobic residues to position key residues or through protein stabilization.

## Results

### Library Construction and Selection of Target DNA-Specific Clones

Shotgun scanning applies phage-displayed libraries with alanine, homolog, or wild-type substitutions programmed in specific positions. Though the goal of alanine shotgun scanning is a library composed of substitutions with alanine or wild-type, some positions also encode two additional amino acids, due to degeneracy in the genetic code; these additional substitutions are labeled m2 and m3 in Table 1. Homolog shotgun scanning features libraries of the wild-type and the homolog substitution with degenerate substitutions listed in Table 2. The initial shotgun scanning libraries had diversities approximating the theoretical diversities encoded by each library including the degenerate substitutions ($\approx 10^9$ different En-HD variants for the most diverse alanine shotgun

*Correspondence: gweiss@uci.edu
[5]Current address: First Department of Internal Medicine, Gunma University School of Medicine, 3-39-15 Showa, Maebashi, Gunma 371-8511, Japan.

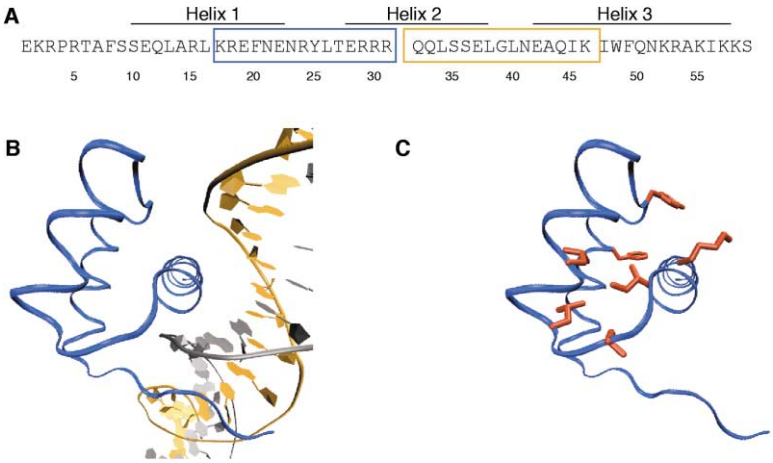


Figure 1. Shotgun Scanning En-HD

(A) The En-HD sequence and library design. Library 1 (blue box) includes residues 16 to 31, and library 2 (yellow box) extends from En-HD residues 32 to 46.

(B) The En-HD structure (blue) bound to DNA [1].

(C) Red En-HD residues contribute significantly to En-HD•DNA affinity and specificity (wt: A > 19).

scanning library). In the naive library, a roughly 1:1 ratio of wild-type to alanine for targeted positions was observed (data not shown). Two alanine shotgun scanning libraries were constructed [18], since a single library including all 30 residues would be unable to include every combination of wild-type and alanine substitutions.

To enrich these libraries for members that bind with high affinity and specificity to TAATTA, we used a phage selection system capable of isolating active mutants of EnHD [20]. From the enriched pools, wild-type or single alanine-substituted En-HD accounted for ≈11% of the sequence-specific DNA binding clones; a total of 93 and 80 unique sequences were found for libraries 1 and 2, respectively. This strong enrichment for the wild-type protein from two different libraries, each with diversities of a billion En-HD variants, demonstrates that the selection successfully identified high-affinity DNA binding proteins.

Table 1. Alanine Shotgun Scanning Results for En-HD Residues 17–46

| wt | wt:A | m2:A | m3:A |
|---|---|---|---|
| **K17** | **7.0** | **85 (E)** | – |
| R18 | 0.2 | 0.5 (G) | 0.1 (P) |
| E19 | 0.6 | – | – |
| **F20** | **>92.0** | – | – |
| N21 | 0.2 | 0.6 (D) | 0.1 (T) |
| **E22** | **5.8** | – | – |
| **N23** | **7.0** | **4.6 (D)** | 1.0 (T) |
| R24 | 0.9 | 0.0 (G) | 1.9 (P) |
| **Y25** | **>92.0** | – | – |
| L26 | 4.0 | 6.8 (V) | 3.7 (P) |
| T27 | 3.0 | – | – |
| E28 | 0.2 | – | – |
| R29 | 0.7 | 2.1 (G) | 0.7 (P) |
| **R30** | **5.4** | 0.4 (G) | 0.2 (P) |
| **R31** | **8.8** | 0.0 (G) | 0.0 (P) |
| Q32 | 0.6 | 0.2 (E) | 0.1 (P) |
| Q33 | 2.1 | 2.1 (E) | 0.1 (P) |
| **L34** | **>80.0** | 0.0 (V) | 0.0 (P) |
| S35 | 0.8 | – | – |
| S36 | 1.0 | – | – |
| E37 | 1.6 | – | – |
| **L38** | **19.0** | **6.3 (V)** | 0.3 (P) |
| G39 | 3.1 | – | – |
| **L40** | **23.0** | **15.0 (V)** | 0.0 (P) |
| N41 | 0.8 | 0.3 (D) | 1.8 (T) |
| E42 | 1.4 | – | – |
| G43 | 0.8 | – | – |
| Q44 | 2.0 | 0.2 (E) | 0.2 (P) |
| **I45** | **30.0** | **49.0 (V)** | – |
| **K46** | **38.5** | 0.5 (E) | – |

Ratio of wild-type (wt) or other programmed mutation (m2, m3) to alanine from En-HD shotgun scanning. Dashes denote positions without m2 or m3 substitutions. Other mutations were found at a low rate (see Supplemental Data). Key residues and substitutions are highlighted in bold. No alanine substitutions were found in positions with wt:A ratios designated >92 or >80; thus, only upper bounds on the ratio of wt:A can be assigned. Such alanine-intolerant positions are expected to make substantial contributions to DNA binding.

## Shotgun Scanning Data

Alanine shotgun scanning data (Table 1) quantifies the relative En-HD residue preferences for wild-type versus alanine. We have previously shown that such values can approximate $\Delta\Delta G_{Ala\text{-}wt}$ values for each position. For an equilibrium binding selection, a ratio of wild-type to alanine of 10 correlates with a $\Delta\Delta G_{Ala\text{-}wt}$ ≈1.4 kcal/mol [19].

Table 2. Homolog Shotgun Scanning Results for En-HD Residues 17–31

| wt | m1:wt | m2:wt | m3:wt |
|---|---|---|---|
| K17 | 1.0 (R) | 0.04 (M) | – |
| R18 | 1.5 (K) | 0.0 (M) | – |
| E19 | 0.4 (D) | 0.0 (Q) | – |
| **F20** | **0.02 (Y)** | – | – |
| N21 | 1.2 (D) | 0.0 (A) | – |
| E22 | 0.2 (D) | 0.02 (Q) | 0.0 (K) |
| N23 | 0.2 (D) | 0.02 (A) | – |
| **R24** | **8.3 (K)** | 0.0 (M) | – |
| Y25 | 1.7 (F) | – | – |
| L26 | 1.3 (I) | – | – |
| T27 | 0.5 (S) | 0.03 (Q) | 0.03 (R) |
| **E28** | 0.0 (D) | **3.0 (R)** | 0.7 (K) |
| R29 | 1.2 (K) | 0.04 (M) | – |
| R30 | 0.3 (K) | 0.0 (M) | – |
| R31 | 0.4 (K) | 0.0 (M) | – |

Ratio of primary mutation (m1) or other mutation (m2, m3) to wild-type (wt) from En-HD homolog shotgun scanning. Dashes denote positions without m2 or m3 substitutions. Key residues and substitutions are highlighted in bold.

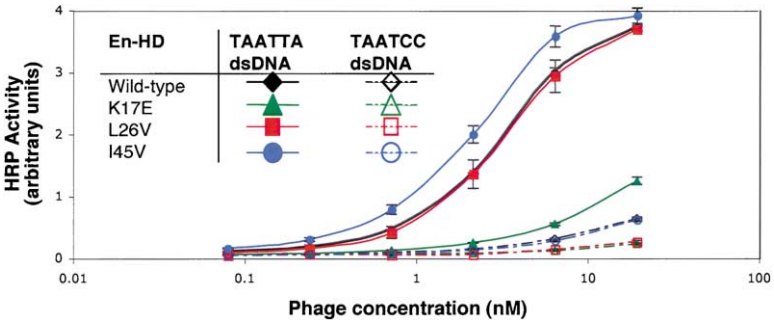


Figure 2. ELISA of Phage-Displayed En-HD Mutants and Wild-Type Protein Assayed for Binding to TAATTA or TAATCC dsDNA

Since non-alanine substitutions are also reported, for consistency, data are presented as ratios of amino acid substitutions.

For example, at the N terminus of the first library, the wild-type side chain of K17 was strongly favored over truncation to alanine (K:A = 7:1). However, due to codon degeneracy, two other possible substitutions were explored, including a glutamic acid replacement for K17. Surprisingly, the mutation K17E was more strongly favored than the wild-type lysine (E:K = >12:1). In a few positions (F20, Y25, and L34), substitution with alanine was not tolerated at all (wt:A = >92 or >80 in Table 1).

To explore the importance of F20 and Y25 to recognition of DNA, we next constructed a library of homologous substitutions (e.g., Tyr to Phe) from residues 17 to 31. The F20Y mutation was selected only once out of 56 homolog scanning selectants (Table 2). In contrast, the equally alanine-resistant Y25 readily tolerated substitution to phenylalanine (F:Y = 1.7). L34, not previously discussed as important to En-HD, resists loss of even a single methylene, as demonstrated by homologous substitutions from alanine shotgun scanning; substitution from leucine to valine at position 34 was disallowed. Additional strongly preferred hydrophobic residues were L26, L38, L40, and I45. However, both L26 and I45 were shown to prefer substitution to valine, and positions L38 and L40 tolerated valine to a lesser extent.

Of the five arginine and two lysine residues in the shotgun-scanned region, four (K17, R30, R31, and K46) were preferred to alanine. Though R18 tolerated substitution to lysine, alanine was actually the preferred substitution in this position. Other polar residues in this central region of En-HD, such as N21 and E28, preferred alanine to wild-type by at least 5-fold; perhaps contributing less to En-HD function, E22, N23, and T27 demonstrated at least a 3-fold preference for the wild-type residue. In the loop region, position 39 preferred the wild-type glycine to alanine (G:A = 3.1). Unprogrammed, spontaneous mutations typically occur infrequently during shotgun scanning. However, a strong preference for mutation of R29 to the unprogrammed substitutions of either S, T, D, or H was observed (mutant:R = 48:1). In other positions, unprogrammed substitutions were unobserved or occurred at negligible rates (<1–2 substitutions).

### Single Point Mutagenesis

To further investigate unexpected substitution patterns, single point substitutions were introduced into the wild-type En-HD before phage ELISA assessment of mutant specificity and affinity (Figure 2). Two homologous substitutions, L26V and I45V, made subtle improvements to En-HD function. L26V, for example, binds similarly to the TAATTA sequence of DNA as wild-type En-HD, but exhibits essentially no binding to TAATCC. Thus, L26V improves the specificity for DNA binding but not the affinity. I45V, however, improved En-HD binding for TAATTA (approximately 3-fold), but retained, like wild-type En-HD, a slight affinity for TAATCC. The charge-switching substitution, K17E, decreases En-HD affinity for DNA.

## Discussion

### Direct En-HD Contributions to DNA Binding

Through quantification of the contributions made by 30 En-HD residues to sequence-specific DNA recognition, data from shotgun scanning identify side chain functionalities critical for protein function. These data complement the wealth of available En-HD structural information [1–10] and other selections with phage-displayed En-HD [20]. Enrichment for a particular substitution pattern, following rounds of selection for binding, can be influenced by expression in the *E. coli* host, improvements to folding, and contributions, direct and indirect, to protein function. Bias due to expression in the *E. coli* host can be uncovered through growth of the naive library; the naive library used in this study retained a roughly 1:1 ratio of wild-type to alanine (or homolog for the homolog shotgun scanning library).

A few shotgun scanned En-HD residues account for either direct or water-mediated contacts to DNA. For example, Y25 appears to form an off-center, though key, $\pi$-cation interaction with R53 at one end of En-HD; this residue is critical for En-HD function, as demonstrated by a total lack of alanine substitution in this position. Shotgun scanning positions completely intolerant of alanine occur only infrequently. This general tolerance for alanine observed during shotgun scanning attests to the plasticity of protein interfaces in general, and specifically to the critical role of Y25. To more carefully explore hydrogen bonding to the DNA backbone by Y25, a homolog shotgun scanning explored the Y25F mutation in this position and homologous substitutions for surrounding residues. In the homolog shotgun scanning library, substitution of Y25 with phenylalanine (Table 2) resulted in essentially no preference for phenylalanine or tyrosine in this position (F:Y = 1.7:1). This result demonstrates the dispensability of a hydrogen bond to the DNA backbone in this position. Thus, the most important contribution to En-HD function by Y25 is to cap one end

Table 3. Comparison of In Vitro Selection with Shotgun Scanning and Natural Evolution

| Residue | En-HD wt | Shotgun Scan Best | Human HD Consensus | Diseases [22] |
|---|---|---|---|---|
| 17 | K | E | E | |
| 18 | R | A | K | • |
| 19 | E | A | E | |
| **20** | **F** | **F** | **F** | • |
| 21 | N | A | H | • |
| 22 | E | E | F | |
| **23** | **N** | **N** | **N** | • |
| 24 | R | K | R | |
| 25 | Y | Y | Y | |
| 26 | L | V | L | |
| 27 | T | T | T | |
| 28 | E | A | R | |
| 29 | R | G | R | |
| **30** | **R** | **R** | **R** | • |
| **31** | **R** | **R** | **R** | ••••• |
| 32 | Q | A | I | |
| 33 | Q | E | E | |
| **34** | **L** | **L** | **L** | • |
| **35** | S | **A** | **A** | • |
| 36 | S | S | H | |
| 37 | E | E | S | • |
| 38 | L | L | L | |
| 39 | G | G | N | |
| 40 | L | L | L | |
| **41** | N | **T** | **T** | • |
| **42** | **E** | **E** | **E** | •• |
| 43 | G | A | R | |
| 44 | Q | Q | Q | |
| **45** | I | **V** | **V** | • |
| **46** | **K** | **K** | **K** | • |

The most preferred substitutions from alanine and homolog shotgun scanning are listed as "shotgun scan best." Each dot represents one human disease. For example, there are five human developmental diseases associated with missense mutations at position 31 [22]. Positions highlighted in bold signify results of shotgun scanning that match the human HD consensus sequence at positions associated with human disease.

of the protein, forming $\pi$-cation and van der Waals interactions to neighboring side chains. Interestingly, the S25Y mutation in the yeast HD, **a1**, improves DNA binding affinity through an analogous structural effect [21].

R31 also contributes critical hydrogen bonds to En-HD functionality, as demonstrated by a strong preference for the guanidine functionality of arginine (R:A = 8.6). Substitution of arginine and lysine in the position by homolog shotgun scanning (Table 2) demonstrated a slight preference for the bidentate hydrogen bonding capabilities of the guanidine functionality of arginine (R:K = 2.5:1). Mutagenesis of such key residues in human HDs is associated with disease (Table 3). For example, R31, termed a "hot spot" of disease, causing mutations in human HDs, contributes substantially to En-HD functionality; mutations in this position are associated with five different developmental diseases [22] (Table 3). Disabling direct contacts to DNA can disrupt En-HD function, but contributions made by the noncontacting residues explored by shotgun scanning also dictate protein function.

### Indirect Contributions to DNA Binding

Shotgun scanning preferences can also uncover indirect influences on En-HD•DNA binding specificity. Results

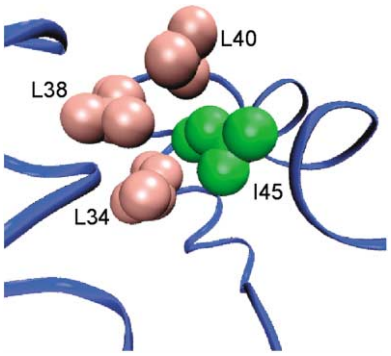


Figure 3. A Turn-Forming Hydrophobic Splint between Helices Two and Three

Labeled residues were highly resistant to alanine substitution.

from En-HD shotgun scanning, for example, identify a crucial hydrophobic core and rank the relative contributions made by leucine and isoleucine residues to form the turn between helices two and three. First, shotgun scanning confirmed a strong preference for L34, L38, L40, and I45 in a network of turn-stabilizing residues (Figure 3). This hydrophobic splint is anchored by L34, an alanine-intolerant position, at one end and by I45 (I:A = 31) at the C-terminal end of the network. These two strongly conserved residues bracket the supporting residues L38 and L40. Interestingly, the smaller valine residue is slightly preferred over isoleucine in position 45, a preference also observed for the consensus sequence of human homeodomains. A subtle packing effect may dictate such preferences. Truncating F20 to alanine was also untolerated, because F20 brackets the hydrophobic interior of En-HD [23]. This position strongly resists accepting a nonhydrogen-bonded hydroxyl, as evidenced by the strong preference for phenylalanine in this position (F:Y = 55). The additional hydroxyl group could disrupt interaction with R53 or I56. Residues F20 and Y25 have been previously shown to be highly conserved for the naturally occurring HDs, further corroborating our data [22].

Demonstrating that shotgun scanning can identify residues important for protein folding, the importance of R30 to En-HD function was demonstrated by its nearly 9:1 preference over alanine. R30 was previously shown to be necessary to folding by virtue of its hydrogen-bonding ability to the En-HD backbone at residues 19, 23, and 25 [24]. L26, which is part of the conserved hydrophobic core network described above, also influences the conformation of R30 [24].

Shotgun scanning revealed some surprising substitutions. First, a strong preference for the K17E mutation was observed (E:K = 12.5:1). This mutation substitutes negative for positive charge in a position that, in the wild-type protein, places two lysine residues close to each other. Substitution of a glutamic acid for K17 abrogates this repulsion and could provide an E17-K52 salt bridge. The converse salt bridge, K17/E52, has been previously observed and, although the substitution substantially increased the stability of En-HD, the affinity for DNA dropped significantly (1000-fold) [24]. The K17E variant of En-HD exhibited decreased affinity by phage

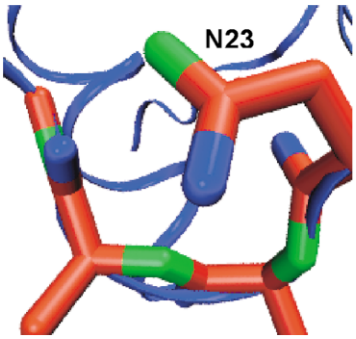


Figure 4. En-HD N23 Interacts with the Protein Backbone

The asparagine carboxamide moiety, conserved during both alanine and homolog shotgun scanning, is within hydrogen bonding distance (3 Å) of both a backbone NH and carbonyl group.

ELISA, but not to the extent of the K17/E52 mutant (Figure 2).

The alanine shotgun scanning data also uncovered the importance of some previously unexamined En-HD residues. E22 was preferred nearly 6-fold over alanine, nearly 5-fold over aspartic acid, and an astonishing 45-fold over glutamine (Table 2). These results demonstrate the utility of both the carboxylic acid moiety and side chain length in this position. Similarly, the importance of N23 is suggested by preference for asparigine over both alanine and aspartic acid (N:A = 7, N:D = 5). The carboxamide of N23 could provide two hydrogen bonds to assist formation of the first turn between En-HD helices one and two (Figure 4).

**Residues Preferring Alanine or Homolog Substitution**

Shotgun scanning also identified a few positions with strong preference for alanine or a homolog in place of the wild-type residue. R18, for example, appeared to hinder binding to DNA when compared to alanine, which was favored 5-fold over arginine. This substitution patterns appears counterintuitive, because one might expect arginine to stabilize DNA binding as opposed to obstructing it. Although asparigine in position 21 was strongly disfavored compared to alanine (N:A = 0.2, Table 1), homolog shotgun scanning revealed little preference for carboxylate or carboxamide in the position (N:D = 0.8). E28, a residue near the phosphodiester backbone, strongly preferred an alanine substitution (E:A = 0.2), since negatively charged glutamic acid could repel the DNA backbone; perhaps not surprisingly, a preference for the unprogrammed mutation to arginine in position 28 was found (R:E = 3). Position 28 has been shown to be highly variable among a variety of HD missense mutations, yet is not strongly associated with human disease [22].

A slight modification, I45V, to the leucine/isoleucine splint (Figure 3) can improve the En-HD affinity for DNA (Figure 2). This mutation illustrates the subtle contributions to protein function made by En-HD residues not in direct contact with DNA. In position 26, both leucine and valine were strongly preferred to alanine to an approximately equal degree (Table 1). Thus, the mutation L26V resulted in a slight gain to En-HD DNA binding specificity without improvement to binding affinity (Figure 2). Structure determination and further characterization of these mutant En-HDs are ongoing in our laboratory.

Previous comprehensive mutagenesis efforts have focused largely on surface residues of proteins in contact with binding partners. These studies, in general, have highlighted the tremendous plasticity of protein surfaces. Often, for example, binding surfaces can tolerate alanine substitutions in a majority of positions; just a few side chains, sometimes clustered into a hot spot, are intolerant of alanine substitutions and contribute the bulk of binding energy to the interaction [19]. Analogous studies focused on protein cores uncover a more finely tuned set of interactions. The relatively small, nominally stable En-HD provides a case in point. With the exception of L26 (wt:A = 4:1), the mainly core-forming hydrophobic residues scanned here vigorously reject alanine substitution. Alanine could introduce unfilled space in the hydrophobic core, thus forcing energetically disfavorable inclusion of water.

En-HD core residue substitutions less drastic than alanine do not provoke the same strong response. In fact, homologous substitutions sampled by the En-HD libraries were sometimes selected as the preferred substitution for a given protein core position. Thus, though not nearly as plastic as surface residues, the En-HD core exhibits a surprising degree of malleability. Residues composing the hydrophobic core often tolerate homologous substitutions, which can even improve upon wild-type En-HD (Figure 2). Our data also demonstrate that En-HD residues are not optimal for binding; through evolution, En-HD has obtained adequate, though not maximized, binding affinity and specificity for DNA. We suggest that it may be advantageous for En-HD to have lower binding affinity to DNA in order to adjust response dynamics at the promoter, thereby ensuring proper transcriptional regulation.

## Significance

**This study expands our knowledge of protein-DNA interactions, particularly for En-HD residues contributing indirectly to DNA specificity and affinity. Such residues precisely set up the En-HD recognition helix. For example, many of the En-HD residues shown to be important by alanine shotgun scanning were in the two En-HD loops. E22 and N23 help form the first loop, while hydrophobic residues L34, L38, L40, and I45 shape the second En-HD loop.**

**With missense mutations in the human HD consensus sequence known to result in human disease [22], our data clarify how En-HD residues not in direct contact with the DNA target can alter protein function. The En-HD core tolerated substitutions with homologous residues to a surprising extent, but in general failed to accept alanine substitutions in place of core hydrophobic residues. Significantly, both natural and in vitro evolution converged on the same consensus sequence. However, improvements to En-HD affinity were uncovered by shotgun scanning, which suggests that natural evolution stopped affinity optimization after reaching an adequate affinity.**

## Experimental Procedures

### Materials

Streptavidin-coupled magnetic beads were purchased from Dynal. Salmon sperm DNA was obtained from Sigma. DNase I was from Amersham Biosciences. Neutravidin was purchased from Pierce. Maxisorp immunoplates were from NUNC. Reagents for dideoxynucleotide sequencing were from ABI/PE Biosciences. All enzymes were from New England Biolabs, excluding Taq polymerase, which was from Continental Lab Products. Anti-M13/HRP conjugate was purchased from Amersham Life Science.

### Oligonucleotides

DNA degeneracies are represented by IUB code (K = G/T, M = A/C, N = A/C/G/T, R = A/G, S = G/C,W = A/T, Y = C/T). The following oligonucleotides were used (mutation-encoding substitutions are shown in italics): En-stop1, 5′-CCAGCGAGCAGTTGGCC*TAATAA*AAGCGGGAGTTCAACGAG-3′; En-stop2, 5′-GCTATCTGACCGAGCGG*TAATAA*CAGCAGCTGAGCAGCGA-3′; En-sg1, 5′-CCAGCGAGCAGTTGGCCCGCCTC*RMASSTGMAKYTRMCGMARMC-SSTKMTSYTRCTGMASSTSSTSSTC*AGCAGCTGAGCAGCGA-3′; En-sg2, 5′-AACGAGAATCGCTATCTGACCGAGCGGAGACGC*SMASMASYTKCC-KCCGMASYTGSTSYTRMCGMAGSTSMARYTRMA*ATCTGGTTCCAGAACAAGCGG-GCCAAG-3′.

### Phage Selection

Phage from the engrailed homeodomain libraries, constructed as previously described [18], were cycled through serial rounds of selection for binding to biotinylated target DNA (TAATTA), similar to [20]. Each library was subjected to separate binding reactions, followed by isolation with streptavidin-coupled magnetic beads. Salmon sperm DNA was used to prevent nonspecific binding to target DNA. To remove phage with nonspecific binding to magnetic beads, the naive phage library ($10^{13}$ phage/ml) was incubated at 4°C overnight with streptavidin-coupled magnetic beads in engrailed library buffer (20 mM HEPES [pH 7.6], 100 mM KCl, 5 mM $MgCl_2$, 5% glycerol, 0.1 mg/mL BSA or casein, which were alternated for each round). After negative selection, nonmagnetic bead binding phage were incubated with 30 to 100 nM biotinylated target DNA, 50 mg/ml salmon sperm DNA in engrailed binding buffer for 1.5 hr at room temperature. Streptavidin-coupled magnetic beads were added, and the solution was mixed for 30 min at room temperature. The beads were captured and the supernatant was removed. The magnetic beads were washed with engrailed wash buffer (20 mM HEPES [pH 7.6], 100 mM KCl, 5 mM $MgCl_2$, 5% glycerol, 0.5% Triton X, 0.1 mg/mL BSA or casein), and the washing was repeated 2–10 times. After capturing the beads, the bound phage were eluted by incubation with 1 μg/ml DNase I in engrailed wash buffer for 5 min at room temperature. After capturing the beads, the phage in the supernatant were used for library propagation.

### Library Propagation

Mid-log-phase XL1-Blue cells were infected with eluted phage for 20 min. After the cells were titered, infected cells were diluted into 2YT media with carbenicillin and M13-K07 helper phage. The culture was shaken at 37°C overnight. The phage were isolated by PEG-NaCl precipitation and used for the next round of selection. After 3 or 4 rounds, the phage-infected cells were grown on LB-carbenicillin plates. The resulting colonies were individually grown in a 96-well format (2YT supplemented with carbenicillin and M13-K07 helper phage) before screening by DNA binding phage ELISA (below).

### DNA Binding Phage ELISA

Phage ELISA experiments were carried out essentially as described [20]. Cultures of *E. coli* harboring individual phagemids were grown in 96-well format for 20 hr at 37°C in 2YT medium (1.2 ml), carbenicillin (50 μg/ml), and M13-K07 helper phage ($10^{10}$ phage/ml). Cells were removed by centrifugation (30 min, 24,000 × g), and the culture supernatant was used directly in the phage ELISA. Maxisorp immunoplates (96 well) were coated with neutravidin overnight at 4°C (100 μl, 5μg/ml neutravidin in 50 mM carbonate buffer [pH 9.6]). The plates were then blocked for 1 hr with BSA (1%) in engrailed binding buffer (25 mM HEPES, 5 mM $MgCl_2$, 100 mM KCl) and washed two times with engrailed wash buffer (25 mM HEPES, 5 mM $MgCl_2$, 100 mM KCl, 0.05% Tween 20). Biotinylated target DNA (30 nM TAATTA) or the biotinylated scrambled DNA (TATATA) were added to separate wells of a Maxisorp immunoplate, respectively. The plate was shaken for 1 hr at room temperature and washed two times with engrailed wash buffer. Cultures of individual selectants from the phage library diluted in engrailed wash buffer were transferred to the wells. The plate was shaken for 2 hr at room temperature and washed two times with engrailed washing buffer. After washing, plates were incubated with anti-M13/horseradish peroxidase conjugate (100 μl, 1:5000 dilution) in engrailed binding buffer, BSA (1%), and Tween 20 (0.025%) for 30 min, and washed one time with engrailed wash buffer and twice with engrailed binding buffer. Plates were developed with an *o*-phenylenediamine dihydrochloride/$H_2O_2$ solution (100 μl, 1 mg/mL OPD, and 0.02% $H_2O_2$) and read spectrophotometrically at 492 nm. Culture supernatant from positive DNA binding selectants was subjected to DNA sequencing. Phage-displayed En-HD mutants assayed for DNA binding in Figure 2 were also subjected to an ELISA to verify that all En-HD mutants displayed equally well on the surface of the phage; this second ELISA quantified protein display through a N-terminal-fused $His_6$ tag binding to anti-$His_6$ antibody (data not shown).

### Supplemental Data

Sequences of selectants from alanine shotgun scanning are available as Supplemental Data at http://www.chembiol.com/cgi/content/full/11/7/1017/DC1.

### Acknowledgments

We thank Joseph Hsu for preliminary studies. This research was supported by a Young Investigator Award from the Arnold and Mabel Beckman Foundation (to G.A.W.), a NSF graduate fellowship (to M.D.S.), and a predoctoral NIH training grant (to A.M.L.) (GMT3207311).

Received: April 3, 2004
Revised: May 7, 2004
Accepted: May 12, 2004
Published: July 23, 2004

### References

1. Kissinger, C.R., Liu, B.S., Martin-Blanco, 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.
2. Wolberger, C., Vershon, A.K., Liu, B., Johnson, A.D., and Pabo, C.O. (1991). Crystal structure of a MAT alpha 2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions. Cell *67*, 517–528.
3. Billeter, M., Qian, Y.Q., Otting, G., Muller, M., Gehring, W., and Wuthrich, K. (1993). Determination of the nuclear magnetic resonance solution structure of an Antennapedia homeodomain-DNA complex. J. Mol. Biol. *234*, 1084–1093.
4. Hirsch, J.A., and Aggarwal, A.K. (1995). Structure of the even-skipped homeodomain complexed to AT-rich DNA: new perspectives on homeodomain specificity. EMBO J. *14*, 6280–6291.
5. 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.
6. Wilson, D.S., Guenther, B., Desplan, C., and Kuriyan, J. (1995). High resolution crystal structure of a paired (Pax) class cooperative homeodomain dimer on DNA. Cell *82*, 709–719.
7. Fraenkel, E., Rould, M.A., Chambers, K.A., and Pabo, C.O. (1998). Engrailed homeodomain-DNA complex at 2.2 Å resolution: a detailed view of the interface and comparison with other engrailed structures. J. Mol. Biol. *284*, 351–361.
8. 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.
9. Mayor, U., Guydosh, N.R., Johnson, C.M., Grossmann, J.G., Sato, S., Jas, G.S., Freund, S.M., Alonso, D.O., Daggett, V., and

Fersht, A.R. (2003). The complete folding pathway of a protein from nanoseconds to microseconds. Nature *421*, 863–867.

10. Tucker-Kellogg, 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 A resolution: structural basis for enhanced affinity and altered specificity. Structure *5*, 1047–1054.
11. 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.
12. Marshall, S.A., Morgan, C.S., and Mayo, S.L. (2002). Electrostatics significantly affect the stability of designed homeodomain variants. J. Mol. Biol. *316*, 189–199.
13. Hanes, S.D., and Brent, R. (1991). A genetic model for interaction of the homeodomain recognition helix with DNA. Science *251*, 426–430.
14. 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.
15. Ekker, S.C., von Kessler, D.P., and Beachy, P.A. (1992). Differential DNA sequence recognition is a determinant of specificity in homeotic gene action. EMBO J. *11*, 4059–4072.
16. Ekker, S.C., Young, K.E., von Kessler, D.P., and Beachy, P.A. (1991). Optimal DNA sequence recognition by the Ultrabithorax homeodomain of Drosophila. EMBO J. *10*, 1179–1186.
17. 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.
18. Sidhu, S.S., and Weiss, G.A. (2004). Oligonucleotide-directed construction of phage display libraries. In Phage Display: A Practical Approach, T. Clackson and H.B. Lowman, eds. (Oxford: Oxford University Press).
19. 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.
20. Simon, M.D., Sato, K., Weiss, G.A., and Shokat, K.M. (2004). A phage display selection of engrailed homeodomain mutants and the importance of residue. Nucleic Acids Res. *32*, 3623–3631.
21. Hart, B., Mathias, J.R., Ott, D., McNaughton, L., Anderson, J.S., Vershon, A.K., and Baxter, S.M. (2002). Engineered improvements in DNA-binding function of the MATa1 homeodomain reveal structural changes involved in combinatorial control. J. Mol. Biol. *316*, 247–256.
22. D'Elia, A.V., Tell, G., Paron, I., Pellizzari, L., Lonigro, R., and Damante, G. (2001). Missense mutations of human homeoboxes: a review. Hum. Mutat. *18*, 361–374.
23. Qian, Y.Q., Billeter, M., Otting, G., Muller, M., Gehring, W.J., and Wuthrich, K. (1989). The structure of the Antennapedia homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors. Cell *59*, 573–580.
24. Stollar, E.J., Mayor, U., Lovell, S.C., Federici, L., Freund, S.M., Fersht, A.R., and Luisi, B.F. (2003). Crystal structures of engrailed homeodomain mutants: implications for stability and dynamics. J. Biol. Chem. *278*, 43699–43708.