Rapid identification of key amino-acid–DNA contacts through combinatorial peptide synthesis

Rachel L Winston* and Joel M Gottesfeld

Background: Basic helix–loop–helix (bHLH) transcription factors are characterized by a conserved four-helix bundle that recognizes a specific hexanucleotide DNA sequence in the major groove. Previous studies have shown that amino acids in the basic region make base-specific contacts, whereas the HLH region is responsible for dimerization. Structural data suggest that portions of the loop region may be proximal to the DNA; however, the role of the loop in DNA-binding affinity and specificity has not been investigated.

Results: Protein–DNA recognition by the *Drosophila* bHLH transcription factor Deadpan was probed using combinatorial solid-phase peptide synthesis methods. A series of bHLH peptide libraries that modulate amino acid content and length in the loop region was screened with DNA and peptide affinity columns, and analyzed using matrix-assisted laser desorption ionization mass spectrometry (MALDI–MS). A functional bHLH peptide with reduced loop length was found, and Lys80 was unambiguously identified as the sole loop residue critical for DNA binding. Unnatural amino acids were substituted at this position to assess contributions of the terminal amino group and the alkyl chain length to DNA-binding affinity and specificity.

Conclusions: Using combinatorial solid-phase peptide synthesis methods and MALDI–MS, we were able to rapidly identify a key amino acid involved in DNA binding by a bHLH protein. Our approach provides a powerful alternative to current recombinant DNA methods to identify and probe the energetics of protein–DNA interactions.

Address: Department of Molecular Biology, MB-27, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA.

*Present address: Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA.

Correspondence: Joel M Gottesfeld
E-mail: joelg@scripps.edu

Key words: bHLH proteins, depsipeptide, MALDI–MS, solid-phase peptide synthesis

Received: **21 December 1999**
Revisions requested: **31 January 2000**
Revisions received: **4 February 2000**
Accepted: **9 February 2000**

Published: **14 March 2000**

Chemistry & Biology 2000, **7**:245–251

1074-5521/00/$ – see front matter
© 2000 Elsevier Science Ltd. All rights reserved.

Introduction

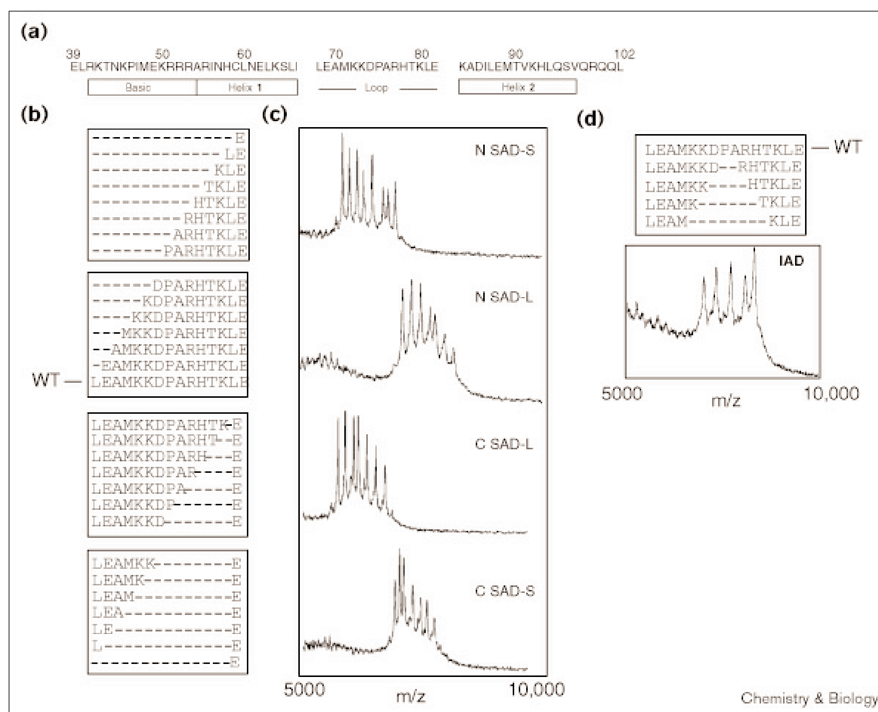
Basic helix–loop–helix (bHLH) transcription factors [1] are characterized by a conserved, parallel four-helix bundle that recognizes a specific hexanucleotide DNA sequence in the major groove [2]. The least characterized region of these proteins is the loop region, which ranges in length from 5 to23 amino acids [1] and varies in amino acid content, especially between proteins of different subfamilies. The structures of six different bHLH domains show that the loop regions display a large degree of structural variation, whereas the helical and basic regions are nearly superimposable [3–6]. It was proposed that a minimum loop of five amino acids is necessary to position correctly helices 1 and 2 in the bHLH fold [3]. Longer loop regions may play more than a structural role, however, by contributing to DNA-binding affinity and/or specificity through phosphate-backbone [3,4,6] or base-specific interactions [4]. Identification of bHLH loop residues that interact with DNA, and the energetic significance of these contacts have yet to be investigated.

Here we present a novel strategy to identify amino acids involved in energetically significant DNA interactions. In this approach, a depsipeptide unit (a peptide that contains an amide to ester substitution) is scanned through the loop region of a chemically synthesized library, corresponding to the bHLH domain of the *Drosophila* transcription factor Deadpan. High-affinity-binding bHLH peptides are selected using DNA- and peptide-affinity columns, and subsequently identified by base-cleavage and matrix-assisted laser desorption ionization mass spectrometry (MALDI–MS).

Results and discussion

The predicted loop region of the *Drosophila* bHLH protein Deadpan (Dpn) [7] is 12–18 amino acids in length [1,7–9]. Although the location of helix 2 is defined in all bHLH domains by a strictly conserved lysine residue (Lys83 in the Dpn sequence, Figure 1a), the precise end of helix 1 is not obvious for bHLH proteins that lack a semiconserved proline residue [1]. Based on a recent systematic classification of bHLH proteins [8], the predicted location of the helices and loop region of Dpn are shown in Figure 1a. In order to define the boundary between helix 1 and the loop, and to determine what role, if any, amino acid sidechains in the loop region play in DNA binding, a series of four combinatorial bHLH libraries of Dpn was generated, in which the length of the loop region

Figure 1



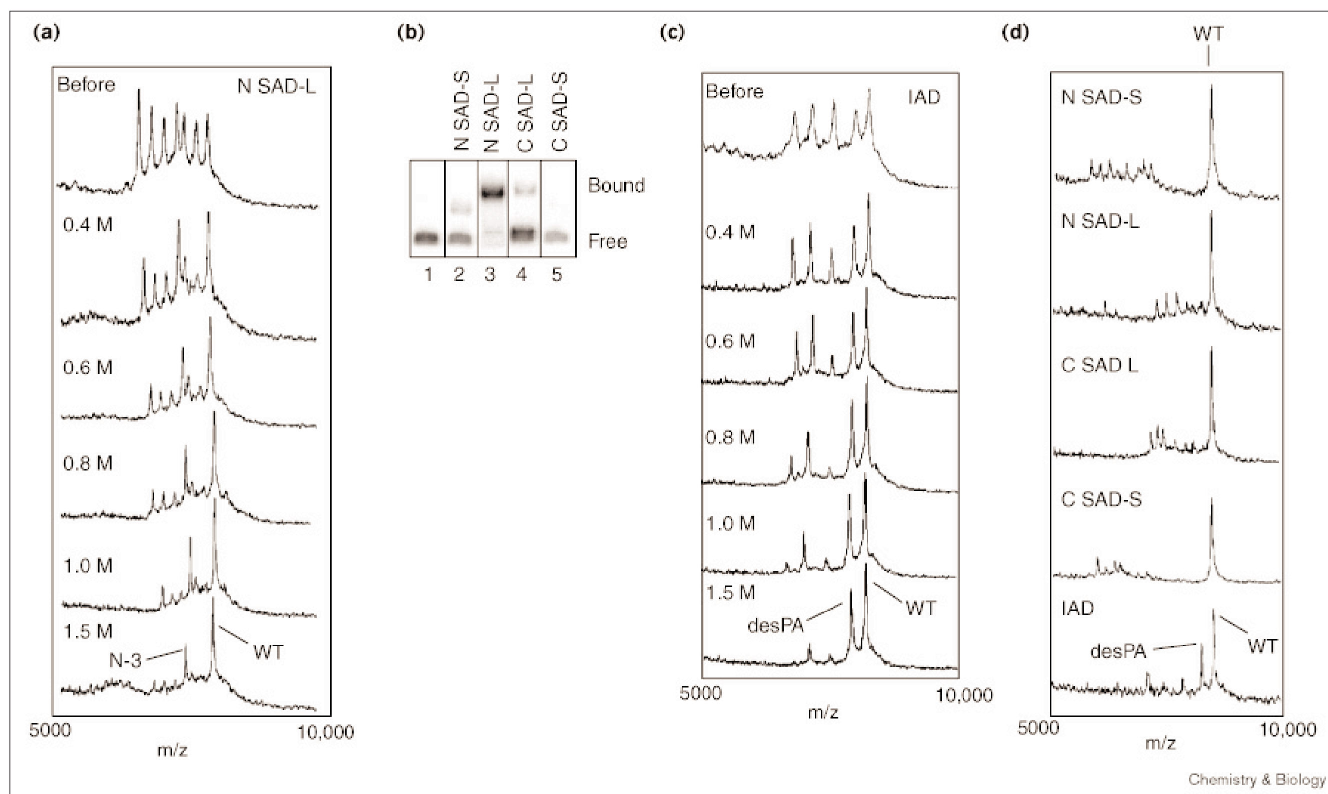
Systematic truncation of the Deadpan (Dpn) loop region. **(a)** Amino acid sequence of the bHLH domain of Dpn (residues 39–102) [7]. **(b)** Schematic of four libraries containing successive, single amino acid deletions (SADs) in the amino-terminal or carboxy-terminal loop region. Dashes represent deleted amino acids. Although only the loop region sequence is shown, for clarity, these modifications exist within the context of the full length bHLH sequence shown in (a). From top to bottom: N SAD-S, N SAD-L, C SAD-S, C SAD-L (L, long; S, short). Note that the wild-type (WT) Dpn loop sequence is present in the N SAD-L library. **(c)** MALDI mass spectrum of each library as indicated. Observed masses were within experimental uncertainty to calculated masses (± 0.1% Da). **(d)** Schematic and MALDI mass spectrum of internal amino acid deletion (IAD) library. Two amino acids were deleted successively from the center of the loop region to generate this five-component peptide library.

was systematically reduced (Figure 1b). Manual, stepwise solid-phase peptide synthesis (SPPS) methods [10] were employed to prepare the bHLH portion of Dpn (residues 39–102 in [7]) and a split-resin method was used to introduce successive, single amino acid deletions (SADs) from both the amino- and carboxy-terminal ends of the loop region (see the Materials and methods section). This chemical approach obviates recombinant DNA techniques, such as plasmid construction, optimization of protein expression and purification, and characterization of individual mutants. Using this strategy, 26 bHLH domain variants were generated in a few days (Figure 1b). Each component in each library has a unique mass corresponding to a particular mutant bHLH peptide that can be resolved using MALDI–MS (Figure 1c).

In order to determine which mutant peptides retain DNA-binding activity, each library was passed over a DNA affinity column containing a known Dpn recognition sequence [11]. A large excess of protein to DNA (80-fold) was used to ensure competition between peptides, and a gradient of increasing ionic strength buffer was used to select for high-affinity binding peptides. Fractions from each step were collected, and subjected to concentration and desalting for MALDI–MS analysis [12]. Within each peptide library, it is likely that a complex mixture of bHLH heterodimers exists. Heterodimers that form unproductive complexes will be selected against during DNA affinity chromatography. High protein concentrations ensure that all possible heterodimer combinations are represented. Figure 2a shows MALDI mass spectra of the elution profile from the functional selection of the N SAD-L library. Before selection, all components in the mixture display roughly equal ion intensities (top spectrum); during the course of DNA affinity selection, however, only ion signals corresponding to wild-type Dpn and a mutant peptide missing three amino acids from its amino-terminal loop (N-3) remain. This result suggests that these three amino acids (residues 68–70) represent the final α-helical turn of helix 1, and deletion of all three (but not one or two) amino acids restores the proper helix–loop geometry. Dpn and members of the Dpn family probably share a similar structure to E47 [5], which contains an extra helical turn at the end of helix 1, as compared with other bHLH proteins such as Max [3].

The other libraries (C SAD-S, C SAD-L and N SAD-S) were also assayed for DNA-binding affinity, where exogenous wild-type Dpn was added as a positive control. Results from these studies indicated that no one mutant could compete effectively against wild-type Dpn for DNA binding (data not shown). To corroborate these findings, peptides eluted from the DNA column were also assayed using an electrophoretic mobility shift assay (EMSA; Figure 2b). These results are consistent with the MALDI–MS analyses, showing that only the N SAD-L library contains significantly active material. Indeed, with the exception of the N-3 peptide, these results suggest

Figure 2



DNA affinity and bHLH affinity selection of Dpn libraries. **(a)** MALDI mass spectra of the N SAD-L library before (top) and after DNA affinity selection, eluted with the indicated KCl concentrations. Ion signals corresponding to wild-type Dpn and a mutant missing three amino acids from its amino-terminal loop (N-3) are indicated. **(b)** EMSA of selected elution fractions from the DNA affinity column. Samples were equilibrated with a specific DNA probe (see the Material and methods section), subjected to EMSA, and visualized by phosphorimage analysis. Lane 1 is DNA alone, and lanes 2–5 correspond to DNA equilibrated with a 1 μl aliquot from the 0.6 M KCl fraction for each library as indicated. Each lane contains a similar amount of total protein. Note that only N SAD-L contains wild-type Dpn. **(c)** DNA affinity selection of the IAD peptide library. Ion signals corresponding to wild-type Dpn and a mutant missing two amino acids (desPA) are indicated. **(d)** MALDI mass spectra of libraries after bHLH affinity selection. An approximately equimolar concentration of soluble wild-type Dpn was added to each library (except N SAD-L and IAD, which contain wild-type Dpn). These peptide mixtures were incubated with the bHLH column, and the identity of bound peptides was determined by MALDI–MS analysis of desalted and concentrated elution fractions [12].

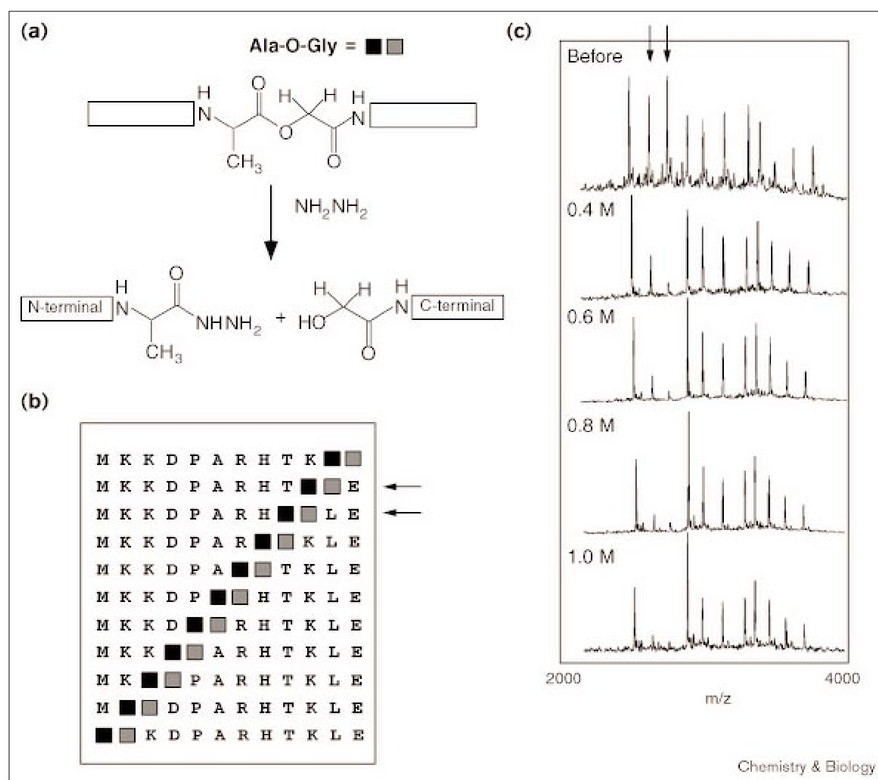
the possibility that absolute length of the loop region is critical for DNA binding. To further assess loop length, we generated an internal amino acid deletion (IAD) Dpn library where successive, two amino acid deletions were introduced in the center of the loop (Figure 1d). Monitoring DNA affinity selection of this library by MALDI–MS revealed that only one mutant peptide missing two amino acids (desPA) from the center of the loop has activity comparable to wild-type Dpn (Figure 2c). Loop length *per se* is not critical for function; however, residues at the loop termini are important for DNA binding.

Because bHLH proteins bind DNA as dimers, we constructed a wild-type Dpn peptide affinity column to determine whether deletions to the loop region interfered with dimerization. Solution studies using circular dichroism spectroscopy as a measure of α-helical content show that Dpn is completely unfolded in the presence of 2 M guanidine hydrochloride (GuHCl; data not shown). We evaluated the binding activity of the bHLH column, and found that a linear gradient of 0–2 M GuHCl was sufficient to elute a soluble wild-type Dpn standard. These conditions were used to assay each library (spiked with wild-type Dpn as a positive control), and MALDI–MS analysis of these selections shows that there is significant loss of ion signal from N SAD and C SAD libraries compared with wild-type Dpn (Figure 2d). Although MALDI–MS is not quantitative, we estimate that ~90% of the protein eluted in these fractions is wild type. Interestingly, the N-3 peptide that is active in DNA binding (Figure 2a) is unable to heterodimerize with wild-type protein (Figure 2d). Based on the DNA-binding activity of the N-3 peptide, we presume that this peptide is capable of homodimerization, but that the shortened helix

1 interferes with heterodimerization with the full-length bHLH peptide. In contrast, analysis of the IAD library reveals that the desPA mutant is capable of dimerizing with immobilized wild-type Dpn, indicating that deletions originating from the ends of the loop are more deleterious to dimerization than a small deletion in the center of the loop. Controls were performed to confirm that binding and elution from the bHLH column reflected the specificity of bHLH dimerization: increasing concentrations of soluble wild-type Dpn added to the libraries resulted in MALDI mass spectra in which only signals corresponding to wild-type Dpn were detected, indicating effective competition of wild-type Dpn with the mutant peptides. Additionally, libraries incubated with a nonrelated bovine serine albumin (BSA)-linked column showed no selection for wild-type Dpn (data not shown).

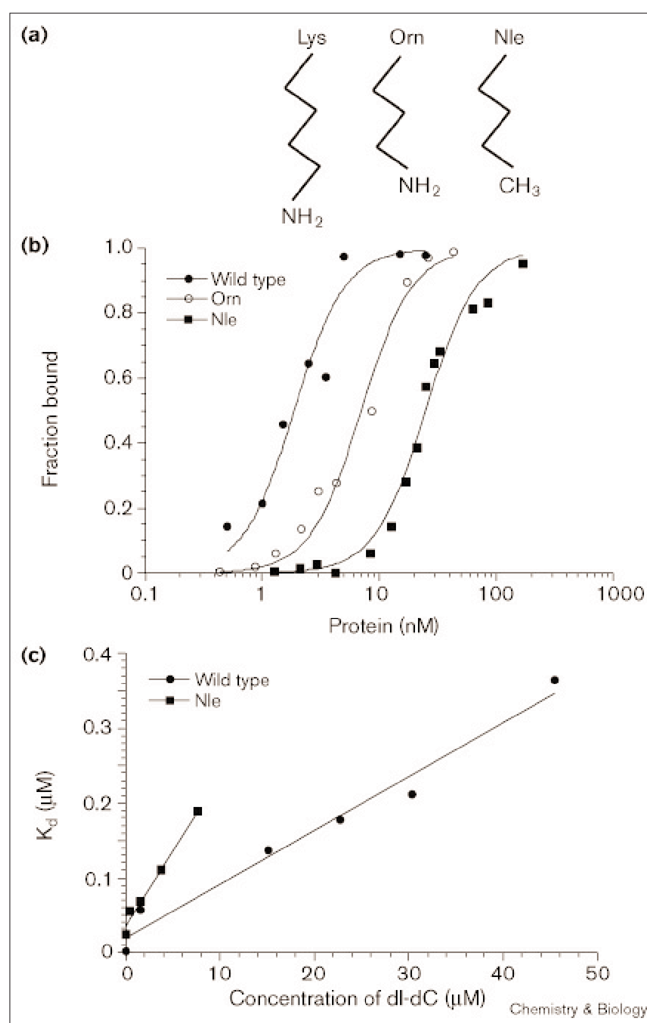
In order to probe amino acid content without modulating the length of the loop region, another library was prepared in which the depsipeptide Ala-O-Gly was systematically scanned through eleven positions in the loop (Figure 3a,b). This depsipeptide serves two purposes: it removes the sidechains of two adjacent amino acids and it allows selective cleavage of the peptide at the ester backbone linkage, so no external tagging scheme is required. The utility of this approach was demonstrated previously with a similar peptide analog unit containing a thioester backbone [13]. Chemical synthesis of this library was accomplished using a resin shuffling procedure [14] in which the Ala-O-Gly unit is incorporated once per peptide at a unique position within the loop region. This library was passed over the DNA affinity column and bound peptides were eluted with increasing concentrations of KCl as before. To 'decode' components in the eluted fractions, the Ala-O-Gly library was cleaved with 1 M hydrazine and then immediately concentrated and desalted for MALDI–MS analysis. This step breaks each bHLH domain into two fragments, yielding amino- and carboxy-terminal ladders that reveal the exact location of the depsipeptide. Figure 3c (top) shows a MALDI mass spectrum of the carboxy-terminal ladder generated after decoding a sample that had not been subjected to DNA affinity selection. After DNA selection, MALDI–MS analysis reveals that two mutant peptides, each missing the sidechain of Lys80, could not compete for DNA-binding in the presence of the other nine loop mutants. Because the position of the ester linkage differs in these two peptides, the possibility of backbone amide contributions to DNA-binding affinity is eliminated. It is conceivable that other basic residues from the loop contribute to DNA-binding activity; peptides lacking these amino acid sidechains (Lys72, Lys73, Arg77) were not selected against, however, suggesting that Lys80 makes a significant and specific DNA contact. The Ala-O-Gly library

Figure 3



DNA affinity selection of the depsipeptide library. **(a)** Structure of the Ala-O-Gly linker incorporated into the loop region of Dpn (top). Cleavage of the linker with hydrazine (bottom). **(b)** Schematic of the position of the Ala-O-Gly linker in the loop region sequence. Only the sequence corresponding to the modified loop region is shown. Cleavage of the library results in the generation of two peptide fragments (between the linker) for each of the eleven bHLH constructs. **(c)** MALDI mass spectra of the library before and after application to the DNA affinity column. Only ion signals from the carboxy-terminal fragments are shown. Ion signals corresponding to bHLH domains with mutated Lys80 (indicated with arrows) disappear in the fractions eluting from the DNA affinity column, indicating that these peptides were unable to compete effectively for DNA binding in the presence of the other nine peptides.

Figure 4



Binding affinity and specificity of Lys80 Dpn mutant peptides. **(a)** Chemical representation of the wild-type Dpn sidechain and the two unnatural amino acid substitutions (Nle and Orn). **(b)** Graphical representation of EMSA peptide titrations for wild-type Dpn, Nle80 Dpn, and Orn80 Dpn. For each bHLH peptide titration, the Hill coefficient was 2, indicating that each protein binds DNA as a dimer [11]. The K_d values for wild-type Dpn, Orn80 Dpn, and Nle80 Dpn are 2.6 nM, 7 nM, and 25 nM, respectively. **(c)** The DNA-binding specificity of Nle80 Dpn is shown in comparison to wild-type Dpn. Each point in the graph represents an apparent K_d at the indicated concentration of the nonspecific competitor poly dI-dC (expressed in concentration of base pairs). The ratio of specific to nonspecific binding was calculated to be 140-fold for wild-type Dpn and 50-fold for Nle80 Dpn from the inverse slope of the fitted line.

was also assayed for dimerization with the bHLH peptide affinity column. MALDI–MS analysis shows that none of the depsipeptide mutations affect dimerization activity (data not shown), indicating that decreased DNA-binding activity for Lys80 mutants is a direct consequence of weakened peptide–DNA interactions, as opposed to diminished bHLH dimerization activity.

To investigate further the nature of this contact, two peptides each containing an unnatural amino acid substitution at position 80 were individually synthesized and characterized [10]. One contained norleucine in place of Lys80 (Nle80), which leaves the alkyl sidechain of lysine intact but deletes the epsilon amino group (Figure 4a). The second peptide contained ornithine in place of Lys80 (Orn80), which maintains the terminal amine, but shortens the alkyl sidechain by one methylene group (Figure 4a). Crude peptides were purified by reverse-phase high performance liquid chromatography (HPLC) and characterized by analytical reverse-phase HPLC and electrospray ionization mass spectrometry (ESI–MS) as described previously [11]. Purified peptides were individually assayed for DNA binding by EMSA using a specific DNA probe, and apparent dissociation constants (K_ds) were determined for a 24 base-pair double-stranded oligonucleotide containing a known Dpn-binding site (Figure 4b). The observed K_ds were 25 nM and 7 nM for the Nle80 and Orn80 Dpn mutants, respectively, compared to 2.6 nM for wild-type Dpn [11]. The epsilon amino group of Lys80 therefore contributes –1.3 kcal/mol to DNA-binding affinity, consistent with the energy gained through a phosphate contact [15]. Adding back the terminal amino group, but shortening the sidechain by one methylene group partially restores binding activity (–0.6 kcal/mol). We also performed a competition assay to measure the binding specificity of wild-type Dpn and Nle80 Dpn. In this experiment, we monitored the extent of binding of a constant amount of protein and radiolabeled oligonucleotide as a function of increasing concentrations of poly dI-dC (a double-stranded DNA mimic) [11]. A threefold loss in specificity was observed for Nle80 Dpn compared with wild-type Dpn (Figure 4c). Taken together, these results demonstrate that the epsilon amine of Lys80 makes significant contributions to both DNA affinity and specificity.

Significance

In this report we present a combinatorial strategy that provides information about residues critical for protein–DNA and protein–protein interactions within the *Drosophila* deadpan (Dpn) basic helix–loop–helix (bHLH) domain. This method provides a rapid alternative to standard recombinant techniques for the generation and assay of mutant proteins. The boundary between a 12 amino acid loop and the adjacent helix was determined, and despite a wide range of loop lengths found throughout the bHLH protein family, only a small deletion to the center of the loop is tolerated in Dpn. Moreover, we demonstrate that the loop region of Dpn is directly involved in DNA binding, providing significant affinity and specificity to Dpn activity. This finding is important because prior investigations into bHLH–DNA interactions have neglected the contributions of amino acids outside the basic region. Additionally, the lysine residue identified here is conserved in the

loop regions of over 40 bHLH subfamilies, suggesting its general importance for maintaining DNA-binding affinity and specificity throughout bHLH proteins. In support of this hypothesis, the corresponding lysine (Lys57) in the crystal structure of the Max homodimer was found to contact the phosphodiester backbone across the minor groove adjacent to the consensus binding site [3]. Using the power of synthetic chemistry, novel functional groups rationally incorporated at this lysine position in Dpn allowed us to assess directly the energetics of this loop contact in bHLH–DNA interactions.

The ability to replace key residues involved in protein–protein or protein–DNA recognition with unnatural amino acids provides a powerful tool with which to dissect and probe energetic contributions to molecular recognition. Because most DNA-binding domains are within the accessible range of total chemical synthesis (< 100 amino acids), the strategy presented here can be readily be adapted to other structural motifs. Another advantage of this method is that chemical synthesis, selection, and matrix-assisted laser desorption ionization mass spectrometry (MALDI–MS) analysis steps are all amenable to automation. Rapid characterization of a vast number of synthetic protein domains is therefore feasible. We envision variations of this strategy in which novel DNA-binding modules could be generated through repeated rounds of synthesis, binding and selection. Alternatively, a minimal protein domain that interacts with a desired target DNA site could be found by incorporating multiple peptide analogs into a protein scaffold. Our approach could also be extended to study full-length proteins by incorporating synthetic peptide libraries into recombinant proteins using the expressed-protein-ligation strategy [16].

Materials and methods

Solid-phase peptide synthesis

The bHLH domain of Dpn (amino acids 39–102 from the wild-type sequence, [7]) and single amino acid mutant peptides (Nle and Orn80) were synthesized manually using stepwise SPPS methods, according to published *in situ* neutralization Boc chemistry protocols [10]. The observed masses were within experimental uncertainty to the calculated masses (WT: calc'd, 7682.2 Da and observed, 7682.3 ± 0.5 Da; Nle80: calc'd, 7666.1 Da and observed, 7666.4 ± 0.8 Da; Orn80: calc'd, 7667.1 Da and observed, 7667.5 ± 0.5 Da). For combinatorial synthesis, 4-methylbenzhydrylamine polystyrene resin was functionalized with residues comprising helix 2 (83–102), and then split in half for the generation of amino- and carboxy-terminal libraries. Amino-terminal deletions in the loop region sequence were easily introduced to one half of helix 2 resin, by transferring equimolar portions of resin after each amino acid coupling step to a separate vessel where no amino acid coupling took place. To facilitate subsequent mass spectral analysis, resin containing shorter (N SAD-S) and longer (N SAD-L) loop sequences were transferred to separate vessels. Introducing deletions from the carboxy-terminal end required a different resin shuffling strategy, because peptide synthesis proceeds in the carboxy→amino direction. In this case, equimolar portions of helix 2 resin were added to the main reaction vessel after every amino acid coupling. By repeating this process, a mixture of peptides was generated with systematically deleted loop regions originating from the carboxy-terminal end of the loop. Again, resin containing shorter (C SAD-S) and longer (C SAD-L) loops were kept in separate vessels. To complete the synthesis of the bHLH domain, amino acids from helix 1 and the basic region were assembled, in a parallel fashion, on the four existing resin pools.

DNA affinity chromatography

A Dpn-specific DNA affinity column was prepared as described previously [17] using complimentary oligonucleotides containing a Dpn recognition sequence: 5′-CGTACGCCGG**CACGCG**ACAGGTCC-3′ (top strand shown, where the bold sequence is the Dpn-binding site [18]). The loading capacity of the column was determined to be 2 nmole/100 μl of resin using a wild-type Dpn standard. The following buffer was used in DNA affinity selection experiments: 20 mM Hepes, pH 7.6, 1 mM EDTA, 5% glycerol. Initial binding was carried out using buffer containing 100 mM KCl, and elution steps contained increasing KCl concentrations, as indicated in the text and figure captions. Controls were performed to validate that increasing ionic strength competes away weakly bound peptides and selects for high affinity peptides. Equimolar amounts of three Dpn bHLH peptides (wild-type Dpn, Dpn(desPA 75, 76), and Dpn(desDPAR 74–77)) with a range of binding affinities (K_ds of 2.6 nM, 4.4 nM, and 44 nM, respectively, for the Dpn site oligonucleotide as determined by EMSA) were pooled and subjected to DNA affinity selection. MALDI–MS analysis of eluted fractions reflected the individual activity of each peptide, i.e. weaker binding peptides eluted at lower ionic strength.

MALDI mass spectrometry

Each crude synthetic library was dissolved in 50% acetonitrile, 0.1% trifluoroacetic acid (TFA) to a concentration of 1–5 μM. A 2 μl aliquot was mixed with an equal volume of saturated matrix solution (α-cyano-4-hydroxycinnamic acid in 50% acetonitrile, 0.1% TFA in water), and 1 μl of the resulting mixture was placed on a MALDI plate and quickly dried with a heat gun. MALDI mass spectra were collected using a Thermo BioAnalysis DYNAMO mass analyzer with delayed extraction and calibrated with an external standard. Typically, the ion signals generated from 50 laser pulses were summed to give a single mass spectrum. Only signals for the singly charged molecules of bHLH mutants are detected.

bHLH affinity chromatography

The polypeptide H–Cys–Ahx-[WT-Dpn (39–102)] (where Ahx is amino hexanoic acid) was synthesized and purified using the general procedures described previously [11], and reacted with pre-swollen Sulfolink resin (Pierce) under conditions suggested by the manufacturer. Tris-2-carboxyethyl phosphine/HCl (10 mM, pH 8.3) was added to the coupling reaction to prevent peptide disulfide formation. The functional substitution of the column was determined by Bradford assay to be ~200 μM. WT-Dpn (1 μM in the following assay buffer: 20 mM Hepes, pH 7.6, 100 mM KCl, 1 mM EDTA, 5% glycerol) and 83 μg/ml BSA were incubated with 200 μl of packed bHLH column resin for 30 min with gentle agitation. After washing with 40 column volumes of assay buffer, bound peptide was eluted from the column at approximately 1 M guanidine hydrochloride (GuHCl) with a 4 ml gradient of 0–2 M GuHCl in assay buffer. Fractions were concentrated and desalted [12]. MALDI–MS analysis of individual fractions was used to monitor and characterize peptide elution.

Chemical synthesis of Boc-Ala-O-Gly

To prepare the depsipeptide, the succinimide ester of Boc-Ala-OH (Boc-Ala-OSu) was reacted under argon with a 2.5 molar excess of glycolic acid in the presence of diisopropylethylamine and methylene chloride. After 12 h, the reaction was neutralized with 1 M HCl, and extracted with ethyl acetate. The desired product was isolated by flash chromatography. The purity and identity of the depsipeptide were established by ^{1}H-NMR and ESI–MS.

Incorporation of Boc-Ala-O-Gly in SPPS

For incorporation into the Dpn polypeptide chain, Boc-Ala-O-Gly (0.25 mmol, 250 μl of 1 M oil in dimethylformamide, DMF) was preacti-

vated for 1 h with 1,3-diisopropylcarbodiimide (0.25 mmol, 39 µl) and N-hydroxybenzotriazole (HOBt; 0.25 mmol, 34 mg) in DMF (311 µl) and used for five consecutive cycles. 125 µl of the preactivated depsipeptide was then coupled to preneutralized resin for 30 min.

Cleavage of Ala-O-Gly libraries

Hydrazine hydrate was added to eluted protein fractions from the DNA or protein affinity columns (final concentration of 1M hydrazine) and immediately diluted with 1 ml of water. Fractions were desalted and concentrated for MALDI-MS as described previously [12].

Electrophoretic mobility shift assays

Dpn mutant peptides were assayed using a double stranded specific oligonucleotide (top strand: 5′-CGTACGCCGG**CACGCG**ACAGG-GC-3′, where the bold sequence is the Dpn-binding site), in the assay buffer with 100 nM KCl and 83 µg/ml BSA as described above for bHLH affinity selection. Samples were electrophoresed on a 10% non-denaturing polyacrylamide gel, and the data were analyzed as described previously [11].

Acknowledgements

We thank J.W. Kelly and P.E. Dawson for generously providing lab space and equipment; L.A. Dickinson, J.W. Kelly, V.B. Oza, P. Raman, and R. Kaul for helpful discussions and advice; G.F. Joyce, J.R. Williamson, and D.K. Treiber for critical comments on our manuscript. This work was supported by a grant from the NIH (GM 47530 to JMG).

References

1. Littlewood, T. & Evan, G.I. (1998). *Helix–Loop–Helix Transcription Factors*. Oxford University Press, New York.
2. Anthony-Cahill, S.J., *et al*., & DeGrado, W.F. (1992). Molecular characterization of helix–loop–helix peptides. *Science* **255**, 979-983.
3. Ferré-D'Amaré, A.R., Prendergast, G.C., Ziff, E.B. & Burley, S.K. (1993). Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. *Nature* **363**, 38-45.
4. Ferré-D'Amaré, A.R., Pognonec, P., Roeder, R.G. & Burley, S.K. (1994). Structure and function of the b/HLH/Z domain of USF. *EMBO J.* **13**, 180-189.
5. Ellenberger, T., Fass, D., Arnaud, M. & Harrison, S.C. (1994). Crystal structure of transcription factor E47: E-box recognition by a basic region helix–loop–helix dimer. *Genes Dev.* **8**, 970-980.
6. Ma, P.C.M., Rould, M.A., Weintraub, H. & Pabo, C.O. (1994). Crystal structure of MyoD bHLH domain–DNA complex: perspectives on DNA recognition and implications for transcriptional activation. *Cell* **77**, 451-459.
7. Bier, E., Vaessin, H., Younger-Shepherd, S., Jan, L.Y. & Jan, Y.N. (1992). Deadpan, an essential pan-neural gene in Drosophila, encodes a helix–loop–helix protein similar to the hairy gene product. *Genes Dev.* **6**, 2137-2151.
8. Atchley, W.R. & Fitch, W.M. (1997). A natural classification of the basic helix–loop–helix class of transcription factors. *Proc. Natl Acad. Sci. USA* **10**, 5172-5176.
9. Dawson, S.R., Turner, D.L., Weintraub, H. & Parkhurst, S.M. (1995). Specificity for the hairy/enhancer of split basic helix–loop–helix (bHLH) proteins maps outside the bHLH domain and suggests two separable modes of transcriptional repression. *Mol. Cell. Biol.* **15**, 6923-6931.
10. Schnölzer, M., Alewood, P., Jones, A., Alewood, D. & Kent, S.B.H. (1992). *In situ* neutralization in Boc-chemistry solid phase peptide synthesis. Rapid, high yield assembly of difficult sequences. *Int. J. Pept. Protein Res.* **40**, 180-193.
11. Winston, R.L., Millar, D.P., Gottesfeld, J.M. & Kent, S.B.H. (1999). Characterization of the DNA binding properties of the bHLH domain of Deadpan to single and tandem sites. *Biochemistry* **38**, 5138-5146.
12. Winston, R.L. & Fitzgerald, M.C. (1998). Concentration and desalting of protein samples for mass spectrometry analysis. *Anal. Biochem.* **262**, 83-85.
13. Muir, T.W., Dawson, P.E., Fitzgerald, M.C. & Kent, S.B.H. (1996). Probing the chemical basis of binding activity in an SH3 domain by protein signature analysis. *Chem. Biol.* **3**, 817-825.
14. Dawson, P.E., Fitzgerald, M.C., Muir, T.W. & Kent, S.B.H. (1997). Methods for the chemical synthesis and readout of self-encoded arrays of polypeptide analogs. *J. Am. Chem. Soc.* **119**, 7917-7927.
15. Lesser, D.R., Kurpiewski, M.R. & Jen-Jacobson, L. (1990). The energetic basis of specificity in the *Eco*RI endonuclease–DNA interaction. *Science* **250**, 776-786.
16. Muir, T.W., Sondhi, D. & Cole, P.A. (1998). Expressed protein ligation: a general method for protein engineering. *Proc. Natl Acad. Sci. USA* **95**, 6705-6710.
17. Kadonaga, J.T. & Tjian, R. (1986). Affinity purification of sequence-specific DNA binding proteins. *Proc. Natl Acad. Sci. USA* **83**, 5889-5893.
18. Villares, R. & Cabrera, C.V. (1987). The achaete-scute gene complex of *D. melanogaster*: conserved domains in a subset of genes required for neurogenesis and their homology to myc. *Cell* **50**, 415-424.