Molecular Recognition in Helix-Loop-Helix and Helix-Loop-Helix-Leucine Zipper Domains

DESIGN OF REPERTOIRES AND SELECTION OF HIGH AFFINITY LIGANDS FOR NATURAL PROTEINS*

Received for publication, November 25, 2002, and in revised form, December 27, 2002
Published, JBC Papers in Press, January 3, 2003, DOI 10.1074/jbc.M211991200

Robert A. Pescitelli‡, Jessica Rosati§, Gianni Cesareni¶, and Sergio Nasi‖

From the ‡Istituto di Biologia e Patologia Molecolari Consiglio Nazionale delle Ricerche, Università La Sapienza, 00185 Roma and §Dipartimento di Biologia, Università Tor Vergata, 00133 Roma, Italy

Helix-loop-helix (HLH) and helix-loop-helix-leucine zipper (HLHZip) are dimerization domains that mediate selective pairing among members of a large transcription factor family involved in cell fate determination. To investigate the molecular rules underlying recognition specificity and to isolate molecules interfering with cell proliferation and differentiation control, we assembled two molecular repertoires obtained by directed randomization of the binding surface in these two domains. For this strategy we selected the Heb HLH and Max Zip regions as molecular scaffolds for the randomization process and displayed the two resulting molecular repertoires on λ phage capsids. By affinity selection, many domains were isolated that bound to the proteins Mad, Rox, MyoD, and Id2 with different levels of affinity. Although several residues along an extended surface within each domain appeared to contribute to dimerization, some key residues critically involved in molecular recognition could be identified. Furthermore, a number of charged residues appeared to act as switch points facilitating partner exchange. By successfully selecting ligands for four of four HLH or HLHZip proteins, we have shown that the repertoires assembled are rather general and possibly contain elements that bind with sufficient affinity to any natural HLH or HLHZip molecule. Thus they represent a valuable source of ligands that could be used as reagents for molecular dissection of functional regulatory pathways.

The JOURNAL OF BIOLOGICAL CHEMISTRY Vol. 278, No. 14, Issue of April 4, pp. 12182–12190, 2003

© 2003 by The American Society for Biochemistry and Molecular Biology, Inc.
Printed in U.S.A.
HLH Repertoires

HLHZip domains were PCR amplified and inserted into the Δ4 vector DNA, between SpeI and NotI restriction sites at the 3' end of a second copy of the D-gene (8). pGEX-2T (Amersham Biosciences) expression plasmids containing GST fusions to human ID2, mouse MyoD, human Max, baboon Mad (amino acids 36–221) and mouse Rox (amino acids 197–346) were introduced into BL21 E. coli cells. Cells were grown at 37 °C to an A600 = 0.5 and induced with 0.1 mm isopropyl-β-D-thiogalactopyranoside for 3 h at 37 °C (MyoD, ID2) or at room temperature (Max, Mad, Rox). After lysis in the presence of 1% Triton X-100, fusion proteins were affinity-purified on glutathione-Sepharose beads (Amersham Biosciences) and analyzed by PAGE.

Construction of HLH and bHLHZip Libraries—A HLH domain repertoire was obtained by PCR amplification of the hev gene HLH domain sequences. A degenerate primer set containing ATG and SpeI restriction sites at the 3' and 5' ends of a second copy of the D-gene (8), 5'-CCGTACATCTGGTGAGGATGTTAATGSMWGCTATTASWGGCATSSRMSCCTTRRCGAGTSDBHCAG-3'; HLH-NotI, 5-GTTCCTGCGCCGCTCTGCTGCTTAGCTACTAAGGATGGTACCC-CTTGTTAGTACGAGABITTTGCDDGGGG-3’ (sequence symbols for degenerate oligonucleotides are: V = AC/G, G = GC, W = AT, M = AC, R = AG, D = AT/G, B = C/T, K = TG, Y = CT). The reactions, containing 100 ng of template DNA, 2 μM oligonucleotide primers, and 4.5 PFU polymerase units, were cyclized 35 times at two different annealing temperatures (45 and 52 °C). The resulting products were used to guarantee the highest level of variability.

A bHLHZip repertoire was generated by two successive PCR amplifications of the HLHZip template. A leucine zipper (Zip) repertoire was obtained in the first reaction with the two degenerate primers: Lz, 5’-ACAGGATATACCTGCGATATGSRARAGVMRASCACACACACWC-MDAAAAMVRMWDAGAGACG-3’ and Lz-NotI: 5’-CCTAGTATATCCGGCGCCGCTCTGCTGCTTAGCTACTAAGGATGGTACCC-CTTGTTAGTACGAGABITTTGCDDGGGG-3’ (sequence symbols for degenerate Zip regions linked to Max bHLHZip domains are: V = AC/G, G = GC, W = AT, M = AC, R = AG, D = AT/G, B = C/T, K = TG, Y = CT). The reactions, containing 100 ng of template DNA, 2 μM oligonucleotide primers, and 4.5 PFU polymerase units, were cyclized 35 times at two different annealing temperatures (45 and 52 °C). The resulting products were used to guarantee the highest level of variability.

A bHLHZip repertoire was generated by two successive PCR amplifications of the HLHZip template. A leucine zipper (Zip) repertoire was obtained in the first reaction with the two degenerate primers: Lz, 5’-ACAGGATATACCTGCGATATGSRARAGVMRASCACACACACWC-MDAAAAMVRMWDAGAGACG-3’ and Lz-NotI: 5’-CCTAGTATATCCGGCGCCGCTCTGCTGCTTAGCTACTAAGGATGGTACCC-CTTGTTAGTACGAGABITTTGCDDGGGG-3’ (sequence symbols for degenerate Zip regions linked to Max bHLHZip domains are: V = AC/G, G = GC, W = AT, M = AC, R = AG, D = AT/G, B = C/T, K = TG, Y = CT). The reactions, containing 100 ng of template DNA, 2 μM oligonucleotide primers, and 4.5 PFU polymerase units, were cyclized 35 times at two different annealing temperatures (45 and 52 °C). The resulting products were used to guarantee the highest level of variability.

A bHLHZip repertoire was generated by two successive PCR amplifications of the HLHZip template. A leucine zipper (Zip) repertoire was obtained in the first reaction with the two degenerate primers: Lz, 5’-ACAGGATATACCTGCGATATGSRARAGVMRASCACACACACWC-MDAAAAMVRMWDAGAGACG-3’ and Lz-NotI: 5’-CCTAGTATATCCGGCGCCGCTCTGCTGCTTAGCTACTAAGGATGGTACCC-CTTGTTAGTACGAGABITTTGCDDGGGG-3’ (sequence symbols for degenerate Zip regions linked to Max bHLHZip domains are: V = AC/G, G = GC, W = AT, M = AC, R = AG, D = AT/G, B = C/T, K = TG, Y = CT). The reactions, containing 100 ng of template DNA, 2 μM oligonucleotide primers, and 4.5 PFU polymerase units, were cyclized 35 times at two different annealing temperatures (45 and 52 °C). The resulting products were used to guarantee the highest level of variability.

A bHLHZip repertoire was generated by two successive PCR amplifications of the HLHZip template. A leucine zipper (Zip) repertoire was obtained in the first reaction with the two degenerate primers: Lz, 5’-ACAGGATATACCTGCGATATGSRARAGVMRASCACACACACWC-MDAAAAMVRMWDAGAGACG-3’ and Lz-NotI: 5’-CCTAGTATATCCGGCGCCGCTCTGCTGCTTAGCTACTAAGGATGGTACCC-CTTGTTAGTACGAGABITTTGCDDGGGG-3’ (sequence symbols for degenerate Zip regions linked to Max bHLHZip domains are: V = AC/G, G = GC, W = AT, M = AC, R = AG, D = AT/G, B = C/T, K = TG, Y = CT). The reactions, containing 100 ng of template DNA, 2 μM oligonucleotide primers, and 4.5 PFU polymerase units, were cyclized 35 times at two different annealing temperatures (45 and 52 °C). The resulting products were used to guarantee the highest level of variability.

A bHLHZip repertoire was generated by two successive PCR amplifications of the HLHZip template. A leucine zipper (Zip) repertoire was obtained in the first reaction with the two degenerate primers: Lz, 5’-ACAGGATATACCTGCGATATGSRARAGVMRASCACACACACWC-MDAAAAMVRMWDAGAGACG-3’ and Lz-NotI: 5’-CCTAGTATATCCGGCGCCGCTCTGCTGCTTAGCTACTAAGGATGGTACCC-CTTGTTAGTACGAGABITTTGCDDGGGG-3’ (sequence symbols for degenerate Zip regions linked to Max bHLHZip domains are: V = AC/G, G = GC, W = AT, M = AC, R = AG, D = AT/G, B = C/T, K = TG, Y = CT). The reactions, containing 100 ng of template DNA, 2 μM oligonucleotide primers, and 4.5 PFU polymerase units, were cyclized 35 times at two different annealing temperatures (45 and 52 °C). The resulting products were used to guarantee the highest level of variability.
In more detail, in the bHLHZip repertoire the degeneration was restricted to the 29-amino acid-long Zip region, which previously had been shown to dictate recognition specificity among bHLHZip domains (6, 24–26). We introduced variations at 13 amino acids occupying the a, d, e, and g positions of the helical wheel (Fig. 3B). These residues represent the interface between the two Zip monomers, whereas the b, c, and f positions are solvent-exposed and were therefore kept invariant (17, 20, 25, 27).

The 44-amino acid-long HLH domain has a more complex structure (Fig. 2A). The helix-loop-helix dimerization motif is a compact four-helix bundle, where the two α-helices package in a coiled-coil only near the carboxyl terminus of the dimer (19). In this case, also residues at b, c, and f positions significantly contribute to the four-helix bundle. Moreover, loop residues, such as Gln22 and Thr23 in the E-proteins, are involved in intermolecular bonds (19). On the basis of these observations, the 15 positions illustrated in Fig. 2B were degenerated in the designed repertoire. Among the residues that were left unchanged there are those at positions 8, 24, 28, 35, 38 in which mutation had previously been shown to impair dimerization (28).

Degenerate DNA sequences encoding the designed HLH and bHLHZip domain repertoires were synthesized by PCR and cloned in the display vector λD4 as fusions to the D capsid protein C terminus (8). Following in vitro packaging, ~2 x 10^6 and ~1 x 10^6 pfu were obtained for the HLH and bHLHZip libraries, respectively. By PCR amplification and sequencing of DNA inserts from randomly chosen phage plaques, we found that ~80% of the phages in each library were recombinant, and that each one contained an insert incorporating from 5 to 10 amino acid changes when compared with the natural scaffold sequence (data not shown).

Affinity Selection with GST-tagged HLH and HLHZip Domains—GST fusions to MyoD and Id2, or to Mad and Rox, were used as baits for panning the HLH and the HLHZip libraries, respectively. For each experiment, after three rounds of selection, ~100 phage clones were amplified, and the interactions with the protein baits were tested by a filter assay. Approximately 10% of the isolated phage clones could be proved to display protein domains that consistently bound the bait. Binding was specific because the clones did not bind GST alone or GST fusions to unrelated protein domains, such as p75 neuro-
trophin receptor and amyloid precursor protein cytoplasmic regions. We quantified the interaction to MyoD, Id2, HEB, Rox, Mad, and Max by ELISA, revealing a number of phage clones with high binding affinity (Figs. 4 and 5). The amino acid sequences of HLH(Zip) inserts were deduced from the DNA sequences and aligned to pinpoint the residues responsible for dimerization specificity and affinity. A number of differences were evident in the sequence alignment (Figs. 4B and 5B). The amino acid frequency profiles of the domains with the highest and the lowest affinity for Id2, MyoD, Mad, and Rox are shown in Tables I and II.

The protein domains isolated from the HLH repertoire were shown in ELISA experiments to bind MyoD, Id2, HEB, Rox, Mad, and Max by ELISA, revealing a number of phage clones with high binding affinity (Figs. 4 and 5). The amino acid sequences of HLH(Zip) inserts were deduced from the DNA sequences and aligned to pinpoint the residues responsible for dimerization specificity and affinity. A number of differences were evident in the sequence alignment (Figs. 4B and 5B). The amino acid frequency profiles of the domains with the highest and the lowest affinity for Id2, MyoD, Mad, and Rox are shown in Tables I and II.

The protein domains isolated from the HLH repertoire were shown in ELISA experiments to bind MyoD, Id2, and Hep with different intensities, ranging from 1 to 8 on an arbitrary scale (Fig. 4B and Table I). Id2 was invariably bound more strongly than MyoD, reflecting the different interaction strength between natural E-proteins and the two baits (1, 28). Amino acid alignment showed a preference for many residues of the E-protein consensus sequence, suggesting that these residues increase dimer stability (Fig. 4B and Table I). They include Ile9, Gly9, Met11, and Cys12 in helix 1, Gln22 and Thr23 in the loop, and Leu26 and Val34 in helix 2. The sequence glycine, methionine, and cysteine at positions 9, 11, and 12 is a specific motif of E-proteins, which precedes their extra helical turn at the helix 1 C terminus (Fig. 2B (19)). At positions 11 and 12 only a few of the residues present in the repertoire were found in the selected domains; the preference for Cys12 was stronger than for Met11 (76 versus 53%). All possible amino acids were found at position 9, where glycine occurred with a 65% frequency, and it was strongly preferred by high affinity binders (domains 43M, 72I, 42I, 13I, 98M, 27M, 18I, and 43I). Gly9 was present whenever Ile27 was found (domains 13I, 53I, 98M, 27M), an observation that suggests a possible interaction between residues 9 and 27, two positions involved in intrachain interactions according to HLH modeling studies (30). The positive correlation between a Gly9 residue and dimerization strength can be explained by structural similarity to the E47 dimer (19), which shows an intrachain hydrogen bond between Gly9 and Gln22. It is interesting to remark the E39Q and V34Y substitutions in the 72I domain, a high affinity binder to Id2 and MyoD, because Gln and Tyr are found at the corresponding helix 2 positions in MyoD and Id2 and in the yeast bHLH, Pho4. In the Pho4 dimer,
in particular, the two residues form an interhelical hydrogen bond, which is not possible in the E47 dimer (22). Because of the presence of the same Gln39 and Tyr34 residues, the hydrogen bond is possible instead in heterodimers between Id2 or MyoD and the 72I domain. Thus, these two residues contribute in specifying the dimerization partner. Valine was also present at position 34 of the high affinity binders. Hydrophobic residues (Ile or Val) were more frequent at position 32 in the high affinity binders, whereas Lys occurred with similar frequency in low and high affinity binding domains. Usually, charged residues were found predominantly in low affinity domains at specific HLH positions (Asp6, Asp7, Glu7, Lys7, Glu9, Arg9, Asp34, Phe34), indicating that their presence weakens heterodimeric associations (Fig. 4B and Table I). The consensus sequences for high affinity binding to MyoD and Id2 did not show substantial differences, making it hard to identify the criteria for dimerization selectivity. The pattern LKAG at positions 5, 6, 7, and 9 was present in two clones (42I and 18I) with higher than average relative affinity for Id2.

Mad and Rox binding affinities to the protein domains isolated from the bHLHZip repertoire ranged from 1 to 5, Mad consistently being a stronger interactors than Rox. Rox and Mad at positions 2, 8, 11, 12, 16, 23, 25, and 26 favored the same amino acids. Surprisingly, Max residues occurred at low frequency in the clones showing the highest binding affinity for Mad and Rox (Table II), with the only exceptions being Lys8 (46%) and Asn7 (53%), as if the Max Zip amino acid sequence was tuned to guarantee dimerization flexibility rather than strength (Fig. 5B and Table II). In the Max dimer, the Asn5 residue is located in front of Asn5 and destabilizes the complex (19, 31). Consistent with the presence of negatively charged residues at position 5 in Mad and Rox (Asp and Glu, respectively), Glu5, which occurred with a 18% frequency, was correlated to low affinity binding of the phage clones (m19, r10, y71, y25). The role of residues 8, 18, 19, and 23 in molecular recognition, suggested by the Max bHLHZip dimer crystallographic structure and by the Myc/Max heterodimeric leucine zipper solution structure (17, 26), was consistent with the amino acid frequency profiles of Table II. Histidine at position 8 was present mainly in clones with low binding affinity, whereas the hydrophobic leucine was strongly preferred by domains with high affinity to Mad and Rox. Position 8 is His in Max, Ala in Mad and Tyr in Rox. Max His8 plays a role in Myc/Max recognition via specific interactions with Myc Glu5 and Glu12 residues (26). Only one of the two salt bridges observed in Myc/Max would be possible in heterodimers with Mad and Rox, which have a negatively charged residue at position 5 only (Asp and Glu, respectively). In the Max Zip dimer, histidine 8 is close to residues 8 and 9 (histidine and glutamine, respectively) of the other monomer. Glutamine 9, although present in the repertoire (Fig. 5B and Table II), never occurred in the selected domains, where ILR substituted it. The binding affinity to Mad and Rox was similar in the presence of a hydrophobic residue (Ile or Leu) at position 9 (clones r45, m50, r15, r32). Position 18 (Gln) is closest to 19 (Asn) in the Max dimer; the Gln18–Asn19 tetrad is involved in stabilization of the dimer (32). Residue 18 is a Gln in both Mad and Rox, whereas residue 19 is Gln in Mad and Lys in Rox. Amino acids 18 and 23 (Glu in Max, Lys in Mad, and Gln in Rox; Fig. 3B) are in the g and e positions of the coiled-coil, flanking the dimer interface, and have the possibility of forming favorable...

Fig. 3. Design of an HLHZip domain repertoire. A, overview of Max bHLHZip dimers complexed with DNA (17). The first and last residues of Max bHLHZip domain (A22 and L104) are indicated. The subdomains are highlighted with different colors in one monomer; the bHLH has the same color code as described in the legend for Fig. 2, and the leucine zipper is red. The positions mutated in the repertoires are in lighter tones. B, outline of the Zip region repertoire. Zip region sequence alignments of the most representative bHLHZip proteins, grouped in subfamilies, are shown underneath the Max sequence, which is used as scaffold. The most conserved residues are highlighted. Degenerate position numbers are shown above these sequences. Nucleotide composition and encoded amino acids for each degenerate position are shown at the top; the classical a-b-c-d-e-f-g heptad repeat of helical structures is indicated.
electrostatic or hydrophobic interactions (24, 26). Positively charged residues (Arg, Lys) were prevalent at position 18 in the domains with lowest affinity, whereas Glu18, which has the potential to establish a salt bridge with Mad Lys 23, occurred frequently in the Mad high affinity binders (domains r45, r27, r10). No preference at position 18 was instead apparent for Rox binding. At position 19 all residues allowed by the repertoire design were accepted. A glutamic acid at position 23, as in Max, was correlated to low binding affinity to Mad and Rox. This is consistent with the presence of a glutamic acid residue at position 18 in Mad and Rox, which would lead to a repulsive electrostatic interaction. Accordingly, high affinity binders preferred a hydrophobic leucine or a basic lysine at position 23.

**DISCUSSION**

In this work, we have shown that it is possible to display HLH and bHLHZip domain repertoires as fusion to the C terminus of protein D on λ phage head, a system that in our hands proved to be better suited than filamentous phage. The repertoires contained different combinations of amino acids found in naturally occurring proteins, grafted into a limited number of positions involved in partner recognition by Heb HLH and Max Zip. Using this approach, it was possible to assemble in an artificial repertoire a large fraction of the binding surfaces of HLH and HLHZip domains explored by natural evolution. To identify patterns of recognition specificity, domains that bind to some natural proteins (MyoD, Id2, Mad1, Rox) with different affinities were isolated by in vitro screening. Overall, it proved difficult to explain the changes in binding affinity by single amino acid substitutions. It appears that the complexity due to multiple amino acid changes produced many alternative combinations of similar binding strength. This is compatible with a view of dimerization as a distributed property of the amino acids in the domain and is consistent with the E47 dimer structure, in which conserved hydrophobic residues at the interior of the HLH form an extensive van der Waals surface that provides most of the favorable dimer interactions (19). However, several correlations were uncovered in our experiments. The presence of hydrophobic residues correlated to stronger interaction of HLH domains, confirming the importance of a hydrophobic core at the dimerization interface for the helix-loop-helix dimerization affinity (29). The presence of a number of residues that were found at high frequency in the HLH domains (Gln22 and Thr23; Ile1, Leu5, Met11/Val11, and Cys12) did not correlate to either greater affinity or specificity to any of the targets, suggesting that these residues have a role in proper folding of the domain and its display on phage coat. The strong bias for the two loop residues Gln22 and Thr23 is in agreement with previous work describing the loop as a key determinant of bHLH stability (33). This role is particularly evident for Gln22, which occurred in all domains; its structural role is visible in the E47 dimer structure, where it participates, together with Gln13 and Gln30, in a hydrogen bond network that connects the loop with helices 1 and 2, stabilizing the four helix bundle (19).

In the HLH domain as well as in the Zip region, several charged residues at the dimer interface appear to represent discontinuity points that are critical for molecular recognition. In the domains isolated from the HLH repertoire, hydrophobic or neutral amino acids were preferred to the charged glutamic acid residues (Arg, Lys) that were prevalent at position 18 in the domains with lowest affinity, whereas Glu18, which has the potential to establish a salt bridge with Mad Lys 23, occurred frequently in the Mad high affinity binders (domains r45, r27, r10). No preference at position 18 was instead apparent for Rox binding. At position 19 all residues allowed by the repertoire design were accepted. A glutamic acid at position 23, as in Max, was correlated to low binding affinity to Mad and Rox. This is consistent with the presence of a glutamic acid residue at position 18 in Mad and Rox, which would lead to a repulsive electrostatic interaction. Accordingly, high affinity binders preferred a hydrophobic leucine or a basic lysine at position 23.

**DISCUSSION**

In this work, we have shown that it is possible to display HLH and bHLHZip domain repertoires as fusion to the C terminus of protein D on λ phage head, a system that in our hands proved to be better suited than filamentous phage. The repertoires contained different combinations of amino acids found in naturally occurring proteins, grafted into a limited number of positions involved in partner recognition by Heb HLH and Max Zip. Using this approach, it was possible to assemble in an artificial repertoire a large fraction of the binding surfaces of HLH and HLHZip domains explored by natural evolution. To identify patterns of recognition specificity, domains that bind to some natural proteins (MyoD, Id2, Mad1, Rox) with different affinities were isolated by in vitro screening. Overall, it proved difficult to explain the changes in binding affinity by single amino acid substitutions. It appears that the complexity due to multiple amino acid changes produced many alternative combinations of similar binding strength. This is compatible with a view of dimerization as a distributed property of the amino acids in the domain and is consistent with the E47 dimer structure, in which conserved hydrophobic residues at the interior of the HLH form an extensive van der Waals surface that provides most of the favorable dimer interactions (19). However, several correlations were uncovered in our experiments. The presence of hydrophobic residues correlated to stronger interaction of HLH domains, confirming the importance of a hydrophobic core at the dimerization interface for the helix-loop-helix dimerization affinity (29). The presence of a number of residues that were found at high frequency in the HLH domains (Gln22 and Thr23; Ile1, Leu5, Met11/Val11, and Cys12) did not correlate to either greater affinity or specificity to any of the targets, suggesting that these residues have a role in proper folding of the domain and its display on phage coat. The strong bias for the two loop residues Gln22 and Thr23 is in agreement with previous work describing the loop as a key determinant of bHLH stability (33). This role is particularly evident for Gln22, which occurred in all domains; its structural role is visible in the E47 dimer structure, where it participates, together with Gln13 and Gln30, in a hydrogen bond network that connects the loop with helices 1 and 2, stabilizing the four helix bundle (19).

In the HLH domain as well as in the Zip region, several charged residues at the dimer interface appear to represent discontinuity points that are critical for molecular recognition. In the domains isolated from the HLH repertoire, hydrophobic or neutral amino acids were preferred to the charged glutamic acid residues (Arg, Lys) that were prevalent at position 18 in the domains with lowest affinity, whereas Glu18, which has the potential to establish a salt bridge with Mad Lys 23, occurred frequently in the Mad high affinity binders (domains r45, r27, r10). No preference at position 18 was instead apparent for Rox binding. At position 19 all residues allowed by the repertoire design were accepted. A glutamic acid at position 23, as in Max, was correlated to low binding affinity to Mad and Rox. This is consistent with the presence of a glutamic acid residue at position 18 in Mad and Rox, which would lead to a repulsive electrostatic interaction. Accordingly, high affinity binders preferred a hydrophobic leucine or a basic lysine at position 23.
acid residues occurring at positions 3, 7, and 39, allowing the formation of stable heterodimers with MyoD and Id2 in the absence of all three Glu residues. Thus, they appear to destabilize the dimers. Previous work suggested that heterodimers of MyoD with the E12 E-protein are stabilized by attractive pairs formed by Glu3, Glu7, and Glu39 residues of E12 with MyoD residues Arg29, Arg33, and Gln39, respectively (34). Because more stable dimers can be obtained with noncharged amino acids, it seems that the role of the charged Glu residues in the E-protein is to prevent an excessively strong interaction with MyoD or Id2, allowing the physiological partner exchange. Similarly, the presence of histidine at Zip position 8 appears to destabilize dimers and promote partner exchange, because this residue was counter-selected in the high affinity binders to Mad and Rox (Fig. 5B, Table II). Consistent with our findings, Max homodimers were strongly stabilized by the replacement of His8 with a leucine and to a lower extent by alanine and tyrosine (31). Leu8 is also present in the bHLHZip protein USF, which forms homodimers that are topologically indistinguishable from Max but does not form heterodimers (17).

The consensus sequences for high affinity binding to MyoD and Id2 were quite similar. Likewise, the amino acids in many Zip region positions (2, 8, 11, 12, 16, 23, 25, and 26) showed the same preference for Rox or Mad binding, indicating that these positions per se are unable to determine specificity. Actually, it was shown previously that it is necessary to mutate four residues (residues 5, 12, 18, and 19) in the Myc Zip to overcome its inability to dimerize (6), that Id1 dimerization specificity can be conferred to E47 by replacing four amino acids at the helix 1/loop junction (36), and that a 6-fold increase in MyoD bHLH dimer stability is obtained by substituting 18 amino acids from the loop and the adjacent regions of E47 (33). Most of the mutants identified as binders show affinity for more than one protein. Thus, a domain recognition code, if it exists, must be rather tolerant. A strategy to increase specific binding to a particular partner would be to assemble and screen second-ary libraries containing a larger number of mutations at a more restricted set of sites, such as those that we found most critical for molecular recognition. Altogether, these findings indicate that natural selection did not operate to maximize specific recognition between E-proteins and tissue-specific HLH, or between Max and the other bHLHZip of the network, but rather to guarantee that these proteins have a broad recognition spectrum to ensure effective binding to their HLH or HLHZip partners. Unnecessarily high affinity for a partner may represent an undesirable property, from an ev-
olutionary standpoint, since it may diminish the reversibility of HLH(Zip) complex formation essential for cellular and developmental plasticity. The charged residues (e.g., the three Glu residues in the HLH and His8 in the Zip) may be critical for providing such function.

On the other hand, a mutant domain with a higher affinity

| Repertoire | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 |
|------------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| E proteins | I E S F K R E G M C V Q T L I V V E |
| strong binders | I V L L F D K T N G S M V C Q T L T I K \_ I V Y Q P |
| Id2 weak binders | I V L L D N E D R E M V C Q T S L T I K \_ I V Y D Q P |
| strong binders | I V L L F N K T N G S M V C Q T L T I K \_ I V Y Q P |
| MyoD weak binders | I V L D D E G R M V C Q T S L T I K \_ I V Y Q P |
| Random clones | L E L D E G M C Q T L I E V Y P |
| 2e | 4g | 5a | 8d | 9e | 11g | 12a | 16e | 18g | 19a | 23e | 25g | 26a |
| Repertoire | R K N H W Q D I K Q N E Q V I |
| Max proteins | Q E K Q N D L H R L V N V K E Q R H L K R K R L |
| strong binders | Q E K N H L R L V D V E Q R K R N E K R K I |
| weak binders | Q E K N H L R L V D V E Q R K R N E K R K I |
| strong binders | Q E K N H L R L V D V E Q R K R N E K R K I |
| Rox weak binders | Q E K N H L R L V D V E Q R K R N E K R K I |
| Random clones | Q O D D H H Q L V Q R Q E K I |

Table II: Zip region amino acid frequency profile of affinity-selected bHLHZip domains

Table I: Amino acid frequency profile of affinity-selected HLH domains

Table III: HLH Repertoires
HLH Repertoires

for a partner can be exploited for functional interference (6, 7). Therefore the phage libraries described in this work represent a valuable collection of reagents and can be used for the selection of HLH and bHLHZip domains with novel recognition properties, to be employed for molecular dissection of the pathways involving HLH transcriptional regulators. This possibility is made more appealing by recent findings that implicate HLH and HLHZip domains in direct interaction not only with proteins of the HLH family but also with other transcriptional regulators such as Miz-1 and JLP, which interact with Myc and Max, or GRIP and Pip, which interacts with the E-proteins (37–40). Such interactions are biologically relevant and enrich the functional plasticity of HLH proteins. Furthermore, mutant domains may be valuable for designing therapeutic approaches to diseases in which cell differentiation or proliferation is perturbed as a consequence of a deregulated HLH protein function. In this context, the HLH domain may represent a target for antiangiogenic drug design, because the naturally occurring HLH proteins Id1 and Id3, as well as Myc, appear to be required for tumor-induced angiogenesis (41, 42). The domains that showed increased affinity for Id2 versus MyoD, such as Id1 and others, are intriguing in view of the role of Id2 as an antagonist of multiple tumor suppressor proteins (43). More particularly, Id2 and Myc were shown to collaborate in overriding the tumor suppressor function of Rb in neuroblastomas, and it was suggested that it might be possible to restore Rb control on cell proliferation in tumor cells, by sequestering Id2 (44). As the 13I domain is able to bind intracellular Id2 (data not shown), it would be tempting to investigate its in vivo function or that of other domains with altered binding properties.

Acknowledgments—We thank Nicola Rizzo for technical assistance, Robert Eisenman, Germana Meroni, and Armando Felsani for GST fusion plasmids, Simona Panni and Giovanna Vacarello for help with λ phage display technology, Barbara Brannetti and Richard Jucker for thoughtful advice, and Laura Soucek for discussions and for critical reading of the manuscript.

REFERENCES

1. Massari, M. E., and Murre, C. (2000) Mol. Cell. Biol. 20, 429–440
2. Grandori, C., Cowley, S. M., James, L. P., and Eisenman, R. N. (2000) Annu. Rev. Cell Dev. Biol. 16, 653–699
3. Baudino, T. A., and Cleveland, J. L. (2001) Mol. Cell. Biol. 21, 691–702
4. Noah, C. E., Tersak, J. M., and Prochownik, E. V. (1999) Oncogene 18, 3004–3016
5. Nasi, S., Carapica, R., Jucker, R., Rosati, J., and Soucek, L. (2001) FEBS Lett. 490, 153–162
6. Soucek, L., Helmer-Citterich, M., Sacco, A., Jucker, R., Cesareni, G., and Nasi, S. (1998) Oncogene 17, 2463–2472
7. Soucek, L., Jucker, R., Fanacchia, L., Ricordy, R., Tato, F., and Nasi, S. (2002) Cancer Res. 62, 3507–3519
8. Panni, S., Dente, L., and Cesareni, G. (2002) J. Biol. Chem. 277, 21666–21674
9. Castagnoli, L., Zacconi, A., Quondam, M., Rossi, M., Vaccaro, P., Panni, S., Paoluzi, S., Santonico, E., Dente, L., and Cesareni, G. (2001) Comb. Chem. High Throughput Screen 4, 121–133
10. O’Connell, D., Becerril, B., Roy-Burnam, A., Daws, M., and Marks, J. D. (2002) J. Mol. Biol. 321, 49–56
11. Santini, C., Brennan, D., Mennuni, C., Hoess, R. H., Nicosia, A., Cortese, R., and Luzzago, A. (1998) J. Mol. Biol. 282, 125–135
12. Santi, E., Capone, S., Mennuni, C., Lahn, A., Tramontano, A., Luzzago, A., and Nicosia, A. (2000) J. Mol. Biol. 296, 487–508
13. Hoess, R. H. (2002) Curr. Pharm. Biotechnol. 3, 23–28
14. Felli, F., Castagnoli, L., Musacchio, A., Jappelli, R., and Cesareni, G. (1991) J. Mol. Biol. 223, 301–310
15. Saggio, I., Glaquen, I., and Lauffer, R. (1995) Gene 152, 35–39
16. Cichin, C., Anusini, H., Amicone, L., Alonzi, T., Nicosia, A., Cortese, R., Tripodi, M., and Luzzago, A. (2002) J. Mol. Biol. 322, 697
17. Ferre-D’Amare, A. R., Prendergast, G. C., Ziff, E. B., and Burley, S. K. (1993) Nature 363, 38–45
18. Brownlie, P., Ceska, T., Lamers, M., Romier, C., Stier, G., Teo, H., and Suck, W. (1997) Structure 5, 509–520
19. Elenderberger, T., Fass, D., Arnaud, M., and Harrison, S. C. (1994) Genes Dev. 8, 970–980
20. Ferre-D’Amare, A. R., Pognonec, P., Roeder, R. G., and Burley, S. K. (1994) EMBO J. 13, 180–189
21. Ma, P. C., Roald, M. A., Weintraub, H., and Pabo, C. O. (1994) Cell 77, 451–459
22. Shimizu, T., Tsumoto, A., Ibara, K., Shimizu, M., Kyogoku, Y., Ogawa, N., Oshima, Y., and Hakoshima, T. (1997) EMBO J. 16, 4899–4907
23. Paggi, L., Bellosillo, L., Ferre-D’Amare, A. R., and Burley, S. K. (1998) Structure 6, 661–672
24. Amati, B., Brooks, M. W., Levy, N., Littlewood, T. D., Evan, G. I., and Land, H. (1993) Cell 72, 233–245
25. Muhele-Goll, C., Gibson, T., Schuck, P., Schubert, D., Nalis, D., Niles, M., and Pastore, A. (1994) Biochemistry 33, 11296–11306
26. Kuo, H., Thoma, R., Kishimoto, N., and Pastore, C. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 4722–4726
27. Nagulapalli, S., Goheer, A., Pitt, L., McIntosh, L. P., and Atchison, M. L. (2002) J. Biol. Chem. 277, 2683–2688
28. Chavali, G. B., Vijayakalmeshi, C., and Salunke, D. M. (2001) Proteins 42, 471–480
29. Tchan, M. C., Weisz, A. S. (2001) FEBS Lett. 509, 177–180
30. Tchan, M. C., Choy, K. J., Mackay, J. P., Lyons, J. T., Bains, N. W., and Weiss, A. S. (2000) J. Biol. Chem. 275, 37454–37461
31. Wendt, H., Thomas, R. M., and Ellenberger, T. (1998) J. Biol. Chem. 273, 5735–5743
32. Shirakata, M., Friedman, F. K., Wei, Q., and Paterson, B. M. (1993) FEBS Lett. 318, 277–282
33. Pastore, A. (1994) J. Biol. Chem. 269, 14189–14194
34. Nagalapalli, S., Gobeer, A., Pitt, L., McIntosh, L. P., and Atchison, M. L. (2002) Mol. Cell. Biol. 22, 7337–7350
35. Heng, J. I., and Tan, S. S. (2002) J. Biol. Chem. 277, 41352–41359
36. Pastore, A. (1994) J. Biol. Chem. 269, 14189–14194
37. Schneider, A., Poukert, K., Elbers, M., and Hanel, F. (1997) Curr. Top. Microbiol. Immunol. 224, 137–146
38. Lee, C. M., Onesime, D., Reddy, C. D., Dhansasekar, N., and Reddy, E. P. (2002) Proc. Natl. Acad. Sci. U. S. A. 99, 14189–14194
39. Nagalapalli, S., Gobeer, A., Pitt, L., McIntosh, L. P., and Atchison, M. L. (2002) Mol. Cell. Biol. 22, 7337–7350
40. Lymen, D. Young, A. Z., Zagag, D., Yan, W., Gerald, W., O’Reilly, R., Bader, B. L., Hynes, R. O., Zaugu, Y., Manova, K., and Benzero, R. (1999) Nature 401, 670–677
41. Baudino, T. A., McKay, C., Pendeville-Samain, H., Nilsson, J. A., Maclean, K. H., White, E. L., Davis, A. C., Igle, J. N., and Cleveland, J. L. (2002) Genes Dev. 16, 2530–2543
42. Lasorella, A., Iavarone, A., and Israel, M. A. (1996) Mol. Cell. Biol. 16, 2570–2578
43. Lasorella, A., Noseda, M., Beyna, M., Yokota, Y., and Iavarone, A. (2000) Nature 407, 592–596