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*

Roberta Ciarapica‡§, Jessica Rosati‡§, Gianni Cesareni¶, and Sergio Nasi‡‖

From the ‡Istituto di Biologia and 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 helix-loop-helix (HLH)[1] proteins, with over 250 representatives in organisms ranging from yeast to man, are one of the most important and versatile families of eukaryotic transcription factors and are involved in diverse processes such as lineage commitment and differentiation, angiogenesis, cell cycle, growth control, and apoptosis (1–3). They are characterized by a highly conserved structural motif organized in a DNA binding sequence, the basic region, and a dimerization domain, either HLH (helix-loop-helix) or HLHZip (helix-loop-helix-leucine zipper). They associate in homo- and heterodimeric complexes that recognize E-box sequences (CANNTG) on DNA, recruit cofactors, and activate or repress transcription of many genes (1–3). Selective dimerization is a regulatory mechanism that allows the expansion of their functional repertoire and also a fine tuning of gene expression by competition of different complexes able to bind the same DNA target sequences. The bHLHZip protein Max, constitutively expressed, is able to homodimerize as well as to heterodimerize with the other bHLHZip factors of the Max network (Myc, Mad1–4, Mnt/Rox), in which expression is regulated and which work only in association with Max (2, 3). *Myc*, one of the most frequently altered genes in human cancer, induces proliferation, growth, and apoptosis but inhibits differentiation (2–5). Mad and Mnt proteins, although possessing DNA binding specificities quite similar to Myc, have only partially overlapping, and frequently opposite, biological functions such as the ability to promote cell survival and differentiation. Similar to Max, among the factors lacking the Zip region, the omnipresent E-proteins (Heb, E47, E12, E2-2) also bind DNA as homodimers (1). The numerous tissue-specific bHLH proteins (MyoD, SCL/Tal, Mash, and many others) poorly homodimerize but require the association with E-proteins to bind DNA and exert their biological functions. HLH proteins lacking a basic region, such as the mammalian Id1–Id4, impose another level of regulation by sequestering E-proteins in dimers that are unable to bind to DNA (1). Understanding molecular recognition is a step toward a rational design of molecules that interfere with HLH protein function. In this regard, we showed that it is possible to inhibit Myc tumorigenic capacity by means of Omomyc, a mutant bHLHZip domain, obtained by changing four residues in the Myc Zip region (6). Omomyc sequesters Myc in complexes unable to bind DNA, preventing transcriptional activation, enhancing repression, potentiating apoptosis (7), and suppressing Myc-induced papillomatosis.[2]

To gain insight into the rules of protein-protein recognition and to isolate mutant domains capable of functional interference, repertoires of HLH and HLHZip domains were designed, exposed on λ phage head, and screened by *in vitro* panning. Several domains that bound with different affinity to MyoD, Id2, Mad-1, and Rox were isolated; their comparison allowed us to elucidate the contribution of different amino acid residues to the stability and specificity of monomer-monomer interactions. These repertoires are a source of potential competitive inhibitors, useful for functional dissection and for drug design.

EXPERIMENTAL PROCEDURES

Phage, Plasmids, and GST Fusion Proteins—DNA sequences encoding Max bHLHZip (Ala²² to Leu¹⁰²) and repertoires of HLH and bHL-

---

§ These two authors contributed equally to the work.

‖ To whom correspondence should be addressed: Università La Sapienza, Istituto di Biologia e Patologia Molecolari CNR, Piazzale Aldo Moro 5, 00185 Roma, Italy. Tel.: 39-0649912241; Fax: 39-0649912500; E-mail: sergio.nasi@uniroma1.it.

[1] The abbreviations used are: HLH, helix-loop-helix; bHLH, basic helix-loop-helix region; Zip, leucine zipper; GST, glutathione *S*-transferase; HRP, horseradish peroxidase; ELISA, enzyme-linked immunosorbent assay; PBS, phosphate-buffered saline.

[2] L. Soucek, S. Nasi, and G. Evan, submitted for publication.

HZip domains were PCR amplified and inserted into the λD4 vector DNA, between *Spe*I and *Not*I 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 $A_{600} \sim 0.5$ 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 *heb* gene HLH domain sequence with two degenerate primers that contained *Spe*I and *Not*I sites: HLH-*Spe*I, 5′-GAACGC**ACTAGT**GTGCGGGAT**V**TTAAT**SWM**G-CATT**SRAMRMS**CTT**RR**GCGA**DTSDB**TCAG-3′; HLH-*Not*I, 5′GTTC-CT**GCGGCCGC**CTTGCTGT**KS**TAGACTAAGGATG**WM**TGCT**WY**GG-CTTGATGAAGA**R**TGAGGA**B**TTTTG**DTW**GGGG-3′ (sequence symbols for degenerate oligonucleotides are: **V** = ACG, **S** = GC, **W** = AT, **M** = AC, **R** = AG, **D** = ATG, **B** = GCT, **K** = TG, **Y** = CT). The reactions, containing 100 ng of template DNA, 2 μM oligonucleotide primers, and 4.5 *Pfu* polymerase units, were cycled 35 times at two different annealing temperatures (45 and 52 °C). The resulting products were mixed to guarantee the highest level of variability.

A bHLHZip repertoire was generated by two successive PCR amplifications on a *max* bHLHZip template. A leucine zipper (Zip) repertoire was obtained in the first reaction with the two degenerate primers: Lz, 5′-ACAGAGTATATCCAGTATATG**SR**AAGG**VAMRASC**ACACAC**W**C-**MD**ACAA**VWMRW**AGACGAC-3′; and Lz-*Not*I: 5′-CAGTGAATTCCC-GGGG**CGGCCGC**GCCCAGTGCACGAA**BTYK**CTGC**WB**CAGAAGAGC-**SYKCYB**CCGTY**K**GAG-3′. The Zip repertoire was used as 3′-primer for the second PCR reaction, whereas an oligonucleotide matching the *max* basic region (Max-*Spe*I, 5′-TGGGT**ACTAGT**GCTGACAAACGGG-CT-3′) served as 5′-primer, creating a bHLHZip repertoire with degenerate Zip regions linked to Max bHLH. Following hot start with *Taq* polymerase (Sigma), the reaction was cycled 35 times (1 min at 95 °C, 1 min at 55 °C, 1 min at 72 °C) followed by a 7-min elongation step.

DNA of both repertoires was digested with *Spe*I and *Not*I restriction enzymes and gel-purified. 20–30 ng of purified insert was ligated to 2 μg of *Spe*I/*Not*I-digested λD4 vector DNA, purified by isopropanol precipitation. The ligation products were phenol/chloroform-extracted, isopropanol-precipitated, and *in vitro* packaged with a Gigapack III Gold kit (Stratagene). The libraries were amplified once by infection of *Escherichia coli* BB4 cells, plated onto LB-agarose plates, and grown for 6–8 h at 37 °C. Phage was eluted overnight at 4 °C with SM buffer (100 mM NaCl, 10 mM MgSO₄, 35 mM Tris-HCl, pH 7.5), precipitated with polyethylene glycol, and suspended at 1×10^{10} pfu/ml.

Panning with GST Fusions to Target HLH Proteins—Affinity selection of phage libraries was performed with GST fusion Id2, MyoD, Mad, and Rox proteins. Phage particles (1×10^9 pfu) were incubated for 1 h at 4 °C with 10 μg of purified GST fusion protein, immobilized on glutathione-Sepharose beads, and preincubated for 2 h in PBS, 3% bovine serum albumin. The beads were washed repeatedly in 50 mM Tris-HCl, pH 7.5, 150 mM NaCl, and 0.5% Tween 20 and suspended in 100 μl of SM buffer. Bound phage was recovered by infection of BB4 cells and plated onto 143-mm dishes. Phage was eluted with SM, titered, and subjected to two more biopanning rounds.

Filter Immunoscreening of Phage Clones—Lysates were prepared from single phage plaques, concentrated by polyethylene glycol precipitation, and titered. 1×10^7 pfu from each phage stock were spotted onto nitrocellulose membrane (Nitroplus, Micron Separation Inc.), which was incubated at room temperature for 2 h in blocking buffer (PBS, 5% milk, 0.1% Nonidet P-40) and again for 2 h with 1 μg/ml GST target protein in the same buffer. After washing in PBS, 0.1% Triton, membranes were incubated for 1 h at room temperature with anti-GST goat serum (Amersham Biosciences, 1:1000) and preadsorbed on bacterial lysate, followed by horseradish peroxidase (HRP)-conjugated anti-goat IgG (1:10000), washed, and developed with an enhanced chemiluminescence kit (ECL, from Amersham Biosciences).

ELISA—Multiwell plates (Nunc) were coated overnight at 4 °C with 100 μl of anti-GST goat serum (5 μg/ml in PBS), washed in PBS, 0.05% Tween, and incubated in PBS, 0.05% Tween, 5% milk for 1 h at 37 °C. 0.5 μg of GST fusion protein was added to each well, for 1 h at room temperature. After washing, phage (10^8 pfu/well) was added and incubated for 1 h at room temperature. The plates were washed with PBS, 0.05% Tween, incubated for 1 h at room temperature with anti-λ phage rabbit IgG (1:1000, courtesy of R. Cortese, Istituto di Richerche di Biologia Moleculare, Pomezia (Rome)), and then incubated with HRP-

conjugated protein A (1:10000, Sigma). Reactions were revealed by adding 100 μl/well tetramethylbenzidine solution (Promega), and the absorbance (*A*) values were recorded by an automated ELISA reader set at 450 nm. All assays were repeated at least three times. The reported values are in arbitrary units, calculated by normalization to the background interaction with GST and to the interaction of empty vector phage to GST, according to the following formula:

$$[A_{\text{phage clone-GST fusion}} - (A_{\text{vector-GST fusion}} - A_{\text{vector-GST}})]/A_{\text{phage clone-GST}}.$$

DNA Sequencing—Phage DNA inserts were PCR-amplified from 1 μl of phage lysate with two primers flanking the *Spe*I and *Not*I cloning sites: 5′-CACGTTCCGTTATGAGGATGT-3′ and 5′-ATGTATCAGTGC-CTAGC-3′. The PCR products were purified from agarose gel using the Concert™ Rapid PCR Purification system (Invitrogen), and their sequences were determined with an ABI-3700 automated sequencer.

Western Blotting—Phage was lysed by boiling for 5 min in 2× SDS-gel sample buffer; proteins were separated by SDS-PAGE and transferred to polyvinylidene difluoride membranes (Amersham Biosciences). Blots were incubated for 1 h at room temperature with anti-D-protein (1:1500, courtesy of R. Cortese) or anti-Max (Santa Cruz C-124; 1:5000) antibodies followed by HRP-protein A (1:10000) and developed with an Amersham Biosciences ECL kit.

RESULTS

Display of Max bHLHZip Domain on λ Phage—To identify the most appropriate vector for the display of HLH and HL-HZip domain repertoires, we tested both filamentous phage vectors, successfully exploited for the construction of peptide or antibody repertoires (9, 10), and λ phage, reported to be generally more suitable for exposing large polypeptides (11–13). The DNA sequence encoding Max bHLHZip was cloned into the three filamentous phage vectors pC89, pC178, and pHENΔ, to obtain N-terminal fusions to pVIII or pIII coat proteins (14, 15) and into the λ display vector 4 (λD4) to display fusions to the D-protein C terminus (8, 16). We asked which vector would efficiently display Max bHLHZip and allow its binding to a natural dimerization partner, the GST fusion protein Mad (2). We found that only the λ vector particles were able to incorporate the D-Max chimeric capsid protein in an amount sufficient for immunological detection in Western blots (Fig. 1*A*). Furthermore, in a simulated panning experiment, we were able to selectively enrich λ phages displaying Max by 1000-fold after three cycles of affinity purification over glutathione resin containing GST-Mad (Fig. 1*B*). Thus, λD4 was selected for the display of domain repertoires.

Design of HLH and HLHZip Repertoires—Repertoires were constructed by mutating only selected amino acids within the scaffold domain sequences, because the library size necessary to fully represent the diversity obtainable by random variations would rapidly saturate the possibilities of phage display libraries. The sequences of Max HLHZip and Heb E-protein HLH were taken as scaffolds for the two domain families (Figs. 2 and 3) because of their dimerization versatility and because of the availability of either their high resolution crystallographic structure (Max (17, 18)) or that of a close relative (E47, an E-protein that shares a high degree of homology with Heb (19)). The amino acid sequences of a large number of HLH and HLHZip domains from different organisms were aligned and the occurrence of different amino acids in each position determined. Strictly conserved residues, likely to be essential for domain stability, were maintained constant in the repertoire design, whereas the artificial repertoire variation was directed at residues that presented natural variability or were shown to be involved in contacts between subunits in the dimeric structures of Max, E47, MyoD, USF, PHO4, and SREBP (17–23). Because a complete randomization of these residues could not be represented fully in a phage display library, only the amino acids found in natural proteins were included in the design. In this way, diversity was reduced to about 7×10^8 combinations, representing a large fraction of the variability observed in natural domains (Figs. 2*B* and 3*B*).
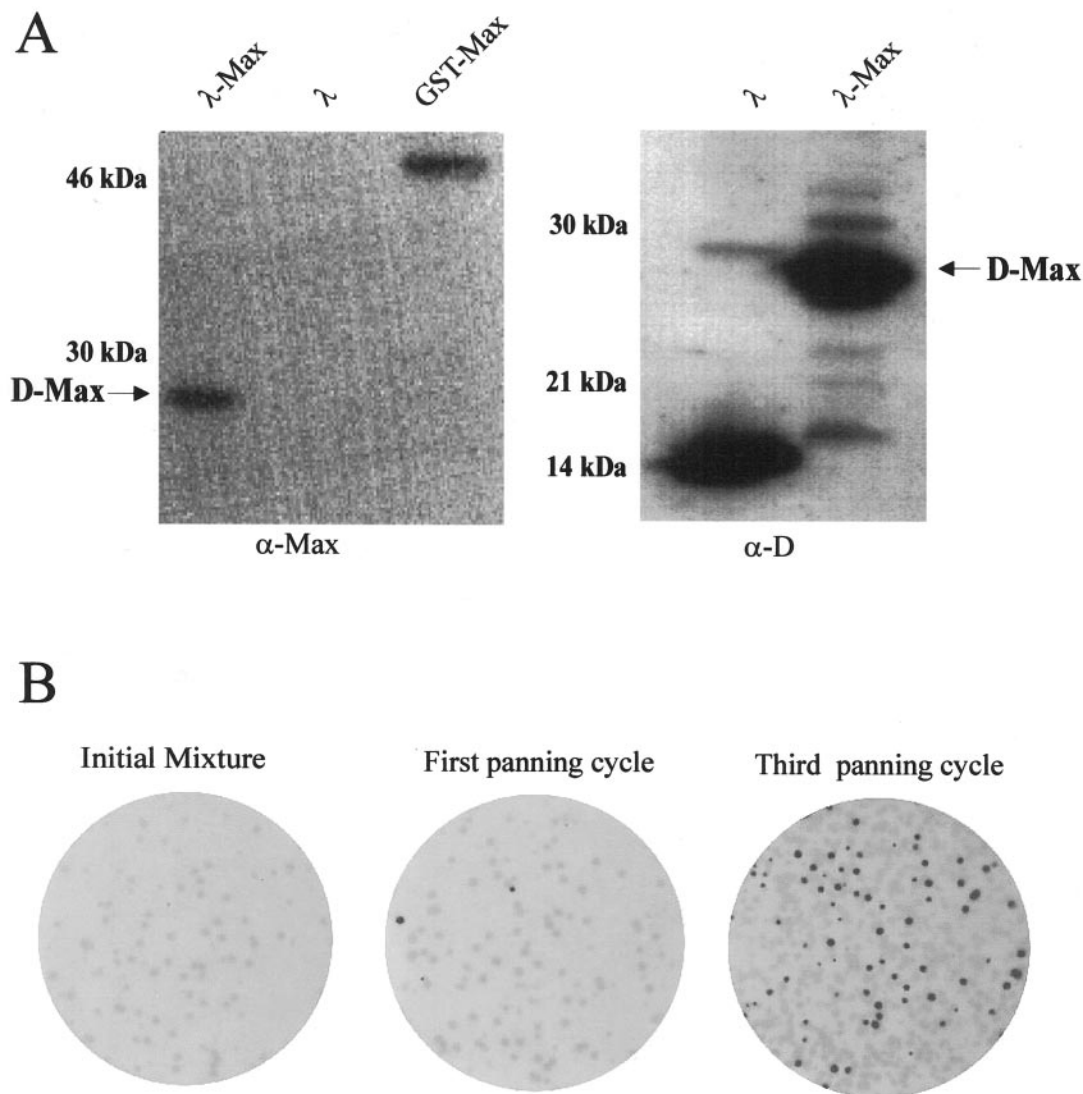
FIG. 1. **Display on λ phage coat and affinity selection of Max bHLHZip domain.** *A*, Western blots. Phage (10^5 pfu/lane) proteins were probed with Max (*left*) and phage head D-protein (*right*) antisera. λ, empty D4 vector phage; λ-*Max*, phage containing a D coat protein-Max bHLHZip domain fusion (27 kDa); *GST-Max*, GST-Max fusion protein (47 kDa). *B*, plaque immunoscreening. λ-Max, at a 1:10000 ratio with empty vector phage, was subjected to three panning cycles with immobilized GST-Mad. Plaques were screened with Max antibodies. The amount of phage exposing Max bHLHZip was enriched by ~1000-fold after the third cycle.

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. 3*B*). 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. 2*A*). 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 Gln[22] and Thr[23] in the E-proteins, are involved in intermolecular bonds (19). On the basis of these observations, the 15 positions illustrated in Fig. 2*B* 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, ~2×10^6 and ~1×10^6 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-
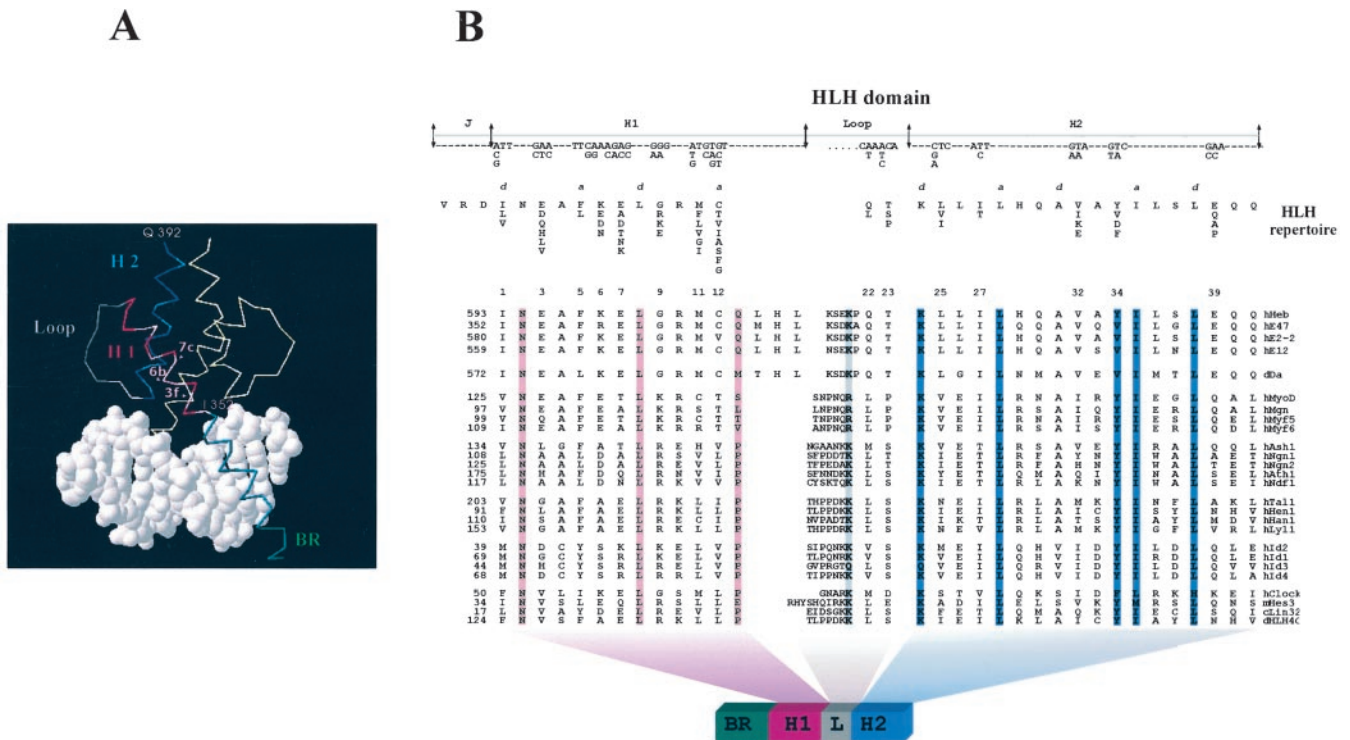
FIG. 2. **Design of an HLH domain repertoire.** *A*, structural overview of E47 bHLH dimers complexed with DNA (19). The first and last residues of the E47 HLH region (Ile[352] and Gln[392]) are indicated. The subdomains of one of the two monomers are *highlighted* in different colors: basic region (*BR*) in *green*, helix 1 (*H1*) in *fuchsia*, loop in *gray*, and helix 2 (*H2*) in blue. The amino acid residues mutated in the repertoire are in lighter tones. The *arrows* denote three mutated helix 1 residues at positions *f*, *b*, and *c* of the helical wheel. They correspond, respectively, to residues Glu[354], Arg[357], and Glu[358] of the E47 sequence, which are on the surface of helix 1, nearest to helix 2′ (Glu[356], Glu[358]) or helix 2 (Arg[357]) (19). *B*, outline of the HLH repertoire. Sequence alignments of the most representative HLH domains, grouped in subfamilies, are shown below the Heb scaffold domain. The most conserved residues are *highlighted* with the same color scheme that was used for the subdomains. Positions degenerated in the repertoire were numbered as shown above the sequence alignment. 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.

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. 4*B* and 5*B*). 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 Heb with different intensities, ranging from 1 to 8 on an arbitrary scale (Fig. 4*B* 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. 4*B* and Table I). They include Ile[1], Gly[9], Met[11], and Cys[12] in helix 1, Gln[22] and Thr[23] in the loop, and Leu[25] and Val[34] 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. 2*B* (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 Cys[12] was stronger than for Met[11] (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). Gly[9] was present

whenever Ile[27] 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 Gly[9] residue and dimerization strength can be explained by structural similarity to the E47 dimer (19), which shows an intrachain hydrogen bond between Gly[9] and Gln[22], a loop residue present in all selected clones. The four-helix bundle must be stabilized if this interaction is preserved in the mutant domains. A similar argument can also explain the preference for Thr[23], which, in the E47 dimer, interacts with Leu[26], a residue not mutated in the repertoire. Thr[23] was found in all domains but two (71I and 37M) that have a Ser residue and are not very strong binders, whereas Pro was never selected. Unlike the majority of the residues, the three negatively charged glutamates found in E-proteins at positions 3, 7, and 39 were either totally absent (Glu[3], Glu[39]) or present (Glu[7]) only in domains that did not strongly interact with MyoD and Id2 (14I, 24I, 30I, 92M; Fig. 4*B*), whereas hydrophobic or neutral amino acids (Leu, Val, Ala, Pro, Asn, Gln, Thr) were preferred in the domains isolated by panning. This was not because of under-representation, because the glutamates were present at the expected frequency in the HLH repertoire, as indicated by sequencing of random clones (Table I). The three glutamates are involved in E47 dimerization; Glu[3] and Glu[7] are on the surface of helix 1, nearest to helix 2′, whereas Glu[39], on helix 2, interacts with His[15′], on helix 1′ (19). 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,

HLH Repertoires

A

B

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.

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 Gln[39] and Tyr[34] 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 (Asp[6]; Asp[7]; Glu[7], Lys[7]; Glu[9], Arg[9]; Glu[32]; Asp[34], Phe[34]), indicating that their presence weakens heterodimeric associations (Fig. 4*B* 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 interactor 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 Lys[4] (46%) and Asn[5] (53%), as if the Max Zip amino acid sequence was tuned to guarantee dimerization flexibility rather than strength (Fig. 5*B* and Table II). In the Max dimer, the Asn[5] residue is located in front of Asn[5′] 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), Glu[5], 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 His[8] plays a role in Myc/Max recognition via specific interactions with Myc Glu[5] and Glu[12] 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. 5*B* 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 Gln[18]–Asn[19] tetrad is involved in stabilization of the dimer (32). Residue 18 is a Glu 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. 3*B*) are in the *g* and *e* positions of the coiled-coil, flanking the dimer interface, and have the possibility of forming favorable

A

B

FIG. 4. **Sequence and binding affinity of selected HLH domains.** *A*, ribbon representation of the E47 HLH (19) depicting the residues that were mutated in the repertoire. E47 residues, in the same color code as described in the legend for Fig. 2, are connected to the amino acid substitutions introduced in the repertoires (*yellow*). *B*, amino acid sequences and relative binding strengths. Phage clones were affinity selected from the HLH repertoire using GST-Id2 and GST-MyoD as baits. Dimerization with Id2, MyoD, and Heb was measured by ELISA. Relative binding strengths, normalized and expressed in arbitrary units (average values ± S.D. from five independent experiments), are indicated at the *left* of each clone. The Heb HLH amino acid sequence, used as scaffold in the repertoire design, is *underlined*. The residues introduced in the repertoire at each degenerate position are indicated *above* the Heb sequence, and the sequences of each selected clone are indicated *below* the E47 sequence.

electrostatic or hydrophobic interactions (24, 26). Positively charged residues (Arg, Lys) were prevalent at position 18 in the domains with lowest affinity, whereas Glu[18], 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 (Gln[22] and Thr[23]; Ile[1], Leu[5], Met[11]/Val[11], and Cys[12]) 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 Gln[22] and Thr[23] is in agreement with previous work describing the loop as a key determinant of bHLH stability (33). This role is particularly evident for Gln[22], which occurred in all domains; its structural role is visible in the E47 dimer structure, where it participates, together with Gln[13] and Gln[30], 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
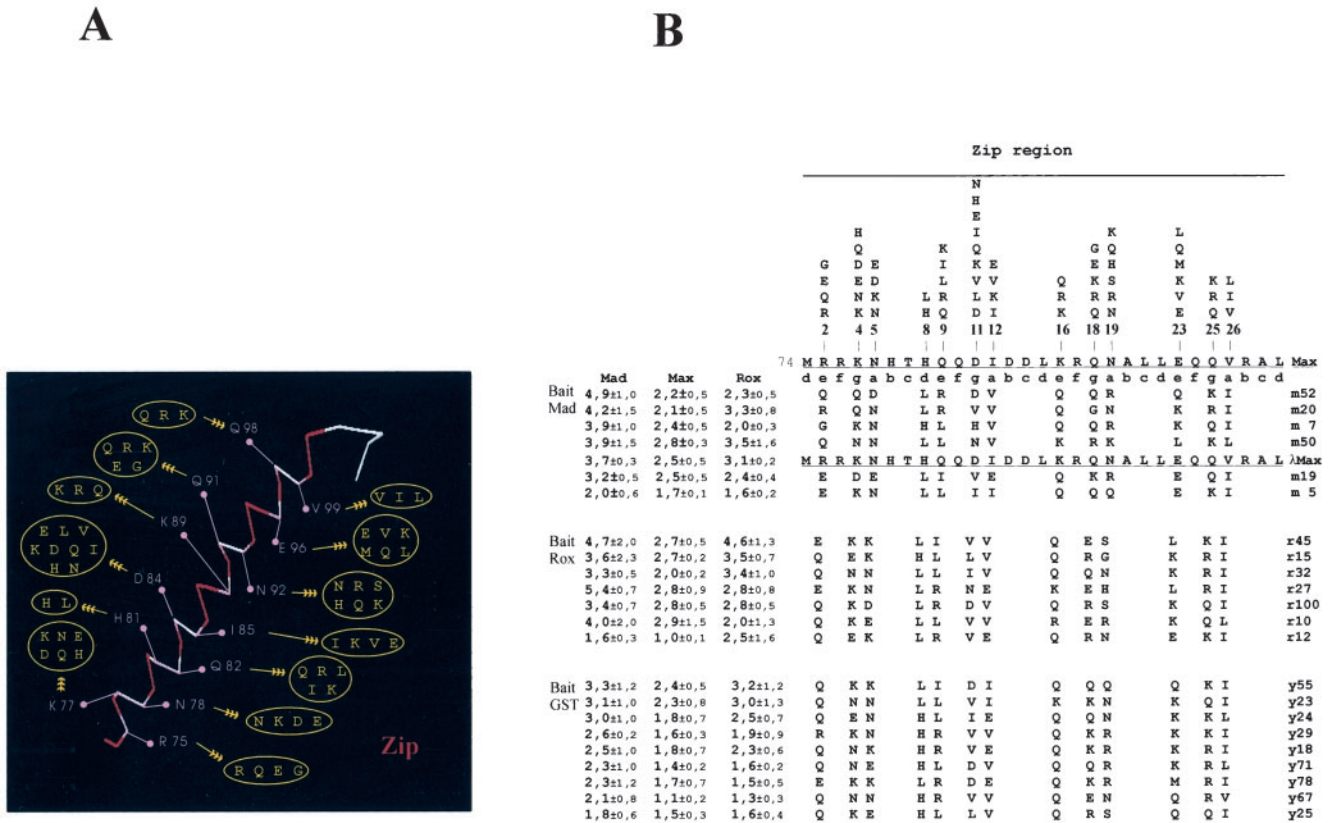
FIG. 5. **Sequence and binding affinity of selected bHLHZip domains.** *A*, *ribbon* representation of Max Zip region (17) depicting the residues that were mutated in the repertoire. Residues, in the same color code as described in Fig. 3 legend, are connected to the amino acid substitutions introduced in the repertoires (*yellow*). *B*, amino acid sequences and relative binding strengths. Phage clones were affinity selected from the bHLHZip repertoire using GST-Mad and GST-Rox as baits. Dimerization of phage clones and a λ-Max control with Max, Mad, and Rox bHLHZip domains was measured by ELISA. Relative binding strengths, normalized and expressed in arbitrary units (average values ± S.D. from five independent experiments), are indicated on the *left* of each clone. The amino acid sequence of Max Zip region, used as scaffold in the repertoire design, is *underlined*; the residues introduced in each degenerate position are indicated *above* the Max sequence.

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 Glu[3], Glu[7], and Glu[39] residues of E12 with MyoD residues Arg[29], Arg[33], and Gln[39], 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. 5*B*, Table II). Consistent with our findings, Max homodimers were strongly stabilized by the replacement of His[8] with a leucine and to a lower extent by alanine and tyrosine (31). Leu[8] is also present in the bHLHZip protein USF, which forms homodimers that are topologically indistinguishable from Max but does not form heterodimers (17). The two *e-g* salt bridges, Myc Glu[11]-Max Lys[16] and Myc Arg[18]-Max Glu[23], contribute to Myc/Max heterodimerization (24, 26). The residues found at positions 16 and 23 in the highest affinity binders to Mad and Rox (*e.g.* domains r27, m52, r45, m20) make either one or both of these electrostatic interactions impossible. Thus they are dispensable for heterodimerization with Mad and Rox, which is consistent with findings on bZip proteins showing that interhelical salt bridges in heterodimers do not necessarily contribute favor-

ably to dimerization specificity and may indeed be unfavorable, when compared with alternative neutral charge interactions (35).

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 secondary 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-

TABLE I
Amino acid frequency profile of affinity-selected HLH domains

	1	3	5	6	7	9	11	12	22	23	25	27	32	34	39
Repertoire	I V L	V L H Q E D	L F	D K E N	A T N E K D	G R E K	M V L F G I	C G S F A I V T	Q L	T S P	L I V	T I	K I V E	V F Y D	Q P A E
E proteins	I	E s	F	K R	E	G	M	C v	Q	T	L	I	V	V	E
Id2 strong binders	I v	V L Q	L F	D K E N	T N A	G K	M V	C G S	Q	T	L I	T I	K I V	V Y	Q P
Id2 weak binders	I V L	V L	L	D N K	E D A K	R E G	M V L	C G	Q	T S	L I V	T I	K I	V Y D E	Q P F
MyoD strong binders	I v	V L Q	L F	N K E D	T N A	G K	M V	C G S	Q	T	L I	T I	K I V	V Y	Q P
MyoD weak binders	I v L	V L H	L F	D K E	N E D A K	G R E	M V L	C	Q	T S	L I V	T I	K I	V D v E	Q P F
Random clones	L L I	E D H	L L F	D K D	E N E	G G G	M V M	C G C	Q Q Q	T S T	L V V	I I T	E V K	V V Y	P E E

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

	2e	4g	5a	8d	9e	11g	12a	16e	18g	19a	23e	25g	26a
Repertoire	R E Q G	K Q N E D H	E D K N	H L	I L R Q K	D I N V H L E Q K	E V I K	Q R K	Q E K R G	H S R N Q K	L Q M K E V	K R Q	L I V
Max proteins	R	K	N I	H N	Q	D	I	K	Q	N	E A	Q	V I
Mad strong binders	Q E R G	K Q N	N D K E	L H	R L I	V N H D	V E	Q K R	E Q R G	R H S N	L K Q	K R Q	I L
Mad weak binders	Q E	K N E	K N E	H L	R L	V D I L	V E I	Q	R K Q E	R N Q S	E K Q M	R K Q	I L v
Rox strong binders	Q E R	N K Q E	N K	L H	L I R	V N	V I	Q K	R Q G E	N K Q S G	K L Q	R K	I L
Rox weak binders	Q E R	K N	N E K	H L	L R	D V L I	V E I	Q	K Q E R	R Q S N	K Q E M	R K Q	I L v
Random clones	Q Q Q	D E D	D D D	H H H	Q R Q	L Q V	V K V	Q K Q	R E R	Q E R	E M V	K Q Q	I I I

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 His[8] in the Zip) may be critical for providing such function.

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

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 GRIPE 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 13I 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.

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. Nesbit, C. E., Tersak, J. M., and Prochownik, E. V. (1999) *Oncogene* **18,** 3004–3016
5. Nasi, S., Ciarapica, 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., Panacchia, L., Ricordy, R., Tato, F., and Nasi, S. (2002) *Cancer Res.* **62,** 3507–3510
8. Panni, S., Dente, L., and Cesareni, G. (2002) *J. Biol. Chem.* **277,** 21666–21674
9. Castagnoli, L., Zucconi, 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-Burman, 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., Lahm, A., Tramontano, A., Luzzago, A., and Nicosia, A. (2000) *J. Mol. Biol.* **296,** 497–508
13. Hoess, R. H. (2002) *Curr. Pharm. Biotechnol.* **3,** 23–28
14. Felici, F., Castagnoli, L., Musacchio, A., Jappelli, R., and Cesareni, G. (1991) *J. Mol. Biol.* **222,** 301–310
15. Saggio, I., Gloaguen, I., and Laufer, R. (1995) *Gene* **152,** 35–39
16. Cicchini, C., Ansuini, 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, D. (1997) *Structure* **5,** 509–520
19. Ellenberger, 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., Rould, M. A., Weintraub, H., and Pabo, C. O. (1994) *Cell* **77,** 451–459
22. Shimizu, T., Toumoto, A., Ihara, K., Shimizu, M., Kyogoku, Y., Ogawa, N., Oshima, Y., and Hakoshima, T. (1997) *EMBO J.* **16,** 4689–4697
23. Parraga, A., Bellsolell, 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. Muhle-Goll, C., Gibson, T., Schuck, P., Schubert, D., Nalis, D., Nilges, M., and Pastore, A. (1994) *Biochemistry* **33,** 11296–11306
26. Lavigne, P., Crump, M. P., Gagne, S. M., Hodges, R. S., Kay, C. M., and Sykes, B. D. (1998) *J. Mol. Biol.* **281,** 165–181
27. Hu, Y. F., Luscher, B., Admon, A., Mermod, N., and Tjian, R. (1990) *Genes Dev.* **4,** 1741–1752
28. Voronova, A., and Baltimore, D. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87,** 4722–4726
29. Goldfarb, A. N., Lewandowska, K., and Shoham, M. (1996) *J. Biol. Chem.* **271,** 2683–2688
30. Chavali, G. B., Vijayalakshmi, C., and Salunke, D. M. (2001) *Proteins* **42,** 471–480
31. Tchan, M. C., and Weiss, A. S. (2001) *FEBS Lett.* **509,** 177–180
32. Tchan, M. C., Choy, K. J., Mackay, J. P., Lyons, A. T., Bains, N. P., and Weiss, A. S. (2000) *J. Biol. Chem.* **275,** 37454–37461
33. Wendt, H., Thomas, R. M., and Ellenberger, T. (1998) *J. Biol. Chem.* **273,** 5735–5743
34. Shirakata, M., Friedman, F. K., Wei, Q., and Paterson, B. M. (1993) *Genes Dev.* **7,** 2456–2470
35. Lumb, K. J., and Kim, P. S. (1995) *Science* **268,** 436–439
36. Pesce, S., and Benezra, R. (1993) *Mol. Cell. Biol.* **13,** 7874–7880
37. Schneider, A., Peukert, K., Eilers, M., and Hanel, F. (1997) *Curr. Top. Microbiol. Immunol.* **224,** 137–146
38. Lee, C. M., Onesime, D., Reddy, C. D., Dhanasekaran, N., and Reddy, E. P. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99,** 14189–14194
39. Nagulapalli, S., Goheer, A., Pitt, L., McIntosh, L. P., and Atchison, M. L. (2002) *Mol. Cell. Biol.* **22,** 7337–7350
40. Heng, J. I., and Tan, S. S. (2002) *J. Biol. Chem.* **277,** 43152–43159
41. Lyden, D., Young, A. Z., Zagzag, D., Yan, W., Gerald, W., O'Reilly, R., Bader, B. L., Hynes, R. O., Zhuang, Y., Manova, K., and Benezra, R. (1999) *Nature* **401,** 670–677
42. Baudino, T. A., McKay, C., Pendeville-Samain, H., Nilsson, J. A., Maclean, K. H., White, E. L., Davis, A. C., Ihle, J. N., and Cleveland, J. L. (2002) *Genes Dev.* **16,** 2530–2543
43. Lasorella, A., Iavarone, A., and Israel, M. A. (1996) *Mol. Cell. Biol.* **16,** 2570–2578
44. Lasorella, A., Noseda, M., Beyna, M., Yokota, Y., and Iavarone, A. (2000) *Nature* **407,** 592–598

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

Roberta Ciarapica, Jessica Rosati, Gianni Cesareni and Sergio Nasi

Access the most updated version of this article at doi: 10.1074/jbc.M211991200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 44 references, 19 of which can be accessed free at
http://www.jbc.org/content/278/14/12182.full.html#ref-list-1