An in vivo library-versus-library selection of optimized protein–protein interactions

Joelle N. Pelletier[1,2], Katja M. Arndt[2], Andreas Plückthun[2], and Stephen W. Michnick[1]*

*[1]Département de biochimie, Université de Montréal, 2900 Edouard Montpetit, Montréal, Québec, Canada. [2]Biochemisches Institut, Universität Zürich, Winterthurerstr. 190, CH-8057 Zürich, Switzerland. *Corresponding author (e-mail: stephen.michnick@umontreal.ca).*

Received 29 March 1999; accepted 26 April 1999

We describe a rapid and efficient in vivo library-versus-library screening strategy for identifying optimally interacting pairs of heterodimerizing polypeptides. Two leucine zipper libraries, semi-randomized at the positions adjacent to the hydrophobic core, were genetically fused to either one of two designed fragments of the enzyme murine dihydrofolate reductase (mDHFR), and cotransformed into *Escherichia coli*. Interaction between the library polypeptides reconstituted enzymatic activity of mDHFR, allowing bacterial growth. Analysis of the resulting colonies revealed important biases in the zipper sequences relative to the original libraries, which are consistent with selection for stable, heterodimerizing pairs. Using more weakly associating mDHFR fragments, we increased the stringency of selection. We enriched the best-performing leucine zipper pairs by multiple passaging of the pooled, selected colonies in liquid culture, as the best pairs allowed for better bacterial propagation. This competitive growth allowed small differences among the pairs to be amplified, and different sequence positions were enriched at different rates. We applied these selection processes to a library-versus-library sample of 2.0×10^6 combinations and selected a novel leucine zipper pair that may be appropriate for use in further in vivo heterodimerization strategies.

Keywords: library screening, library-versus-library, protein-fragment complementation assay, selection strategy, protein–protein interaction, leucine zipper, coiled coil, dihydrofolate reductase, trinucleotide

Many of the problems currently studied in molecular biology and biochemistry are governed by protein–protein interactions. Important examples are the identification and functional characterization of novel gene products, the dissection of proteins into structural or functional motifs, and the study of the physical basis of protein–protein complementarity, whether in naturally occurring proteins or in designed products. The ability to address these problems has been transformed by the development of peptide- and protein-library screening techniques such as the yeast two-hybrid strategy[1,2] and phage display[3], in which a library of proteins is screened for interaction with a "bait" protein. However, as the study of interacting partners is a two-dimensional problem influenced by variations in either partner, it would be advantageous to screen a library of proteins not against a single bait protein, but against a second library of proteins. To date, no large-scale library-versus-library selection of protein–protein interactions has been reported.

We present a strategy for library-versus-library screening in intact cells based on the folding of murine enzyme dihydrofolate reductase (mDHFR) from complementary fragments[4–7]. The mDHFR was genetically dissected into two rationally designed fragments, each of which was fused to a library of proteins or peptides (Fig. 1A). Members of one library that heterodimerized with a member of the other library drove folding of mDHFR from the fragments, resulting in reconstitution of enzymatic activity (Fig. 1B). Activity was detected in vivo using an *Escherichia coli*–based selection assay, in which the bacterial DHFR was specifically inhibited with trimethoprim, preventing biosynthesis of purines, thymidylate, methionine, and pantothenate, and therefore cell division. The reconstituted mDHFR, which was insensitive to the low trimethoprim concentration present in selection, restored the biosynthetic reactions required for bacterial propagation. As a result, the interaction between library partners was directly linked to cell survival and detected by colony formation. We have previously demonstrated the utility of this strategy with GCN4 leucine zipper-forming peptides, as well as with larger heterodimerizing partner proteins[5–7] with dissociation constants ranging between 3 and 160 nM (refs. 6–9), although the affinity limits have not been determined.

In this study we demonstrate a large-scale library-versus-library selection based on the mDHFR fragment complementation assay: We screened two designed libraries of complementary heterodimeric coiled-coil–forming sequences against each other. Our goal was to determine whether the strategy would select interacting peptide pairs in which the amino acids at the semi-randomized positions were similar to those observed in naturally occurring or successfully designed coiled coils, which form stable heterodimers[10–12]. Furthermore, it is not currently possible to predict sequences of coiled-coil–forming peptides that will simultaneously have high stability and heterospecificity as well as advantageous in vivo properties, such as resistance to proteases. In this approach, the heterodimerizing peptides have such characteristics by the nature of their selection. This is crucial to practical applications of optimal interacting heterodimers for in vivo studies of protein oligomerization (*e.g.*, the design of bispecific miniantibodies[13]).

Three selection strategies were tested, each with a different level of stringency. In the lowest-stringency selection, we screened two expression libraries against each other in a single-step selection (Fig. 1B), thereby identifying all interacting polypeptide partners. In the second strategy, we increased the selection stringency by using a mutant mDHFR fragment (I114A) that prevents stable reassembly of mDHFR (ref. 5) and, thus, should require more efficiently heterodimerizing, but not homodimerizing, partners to

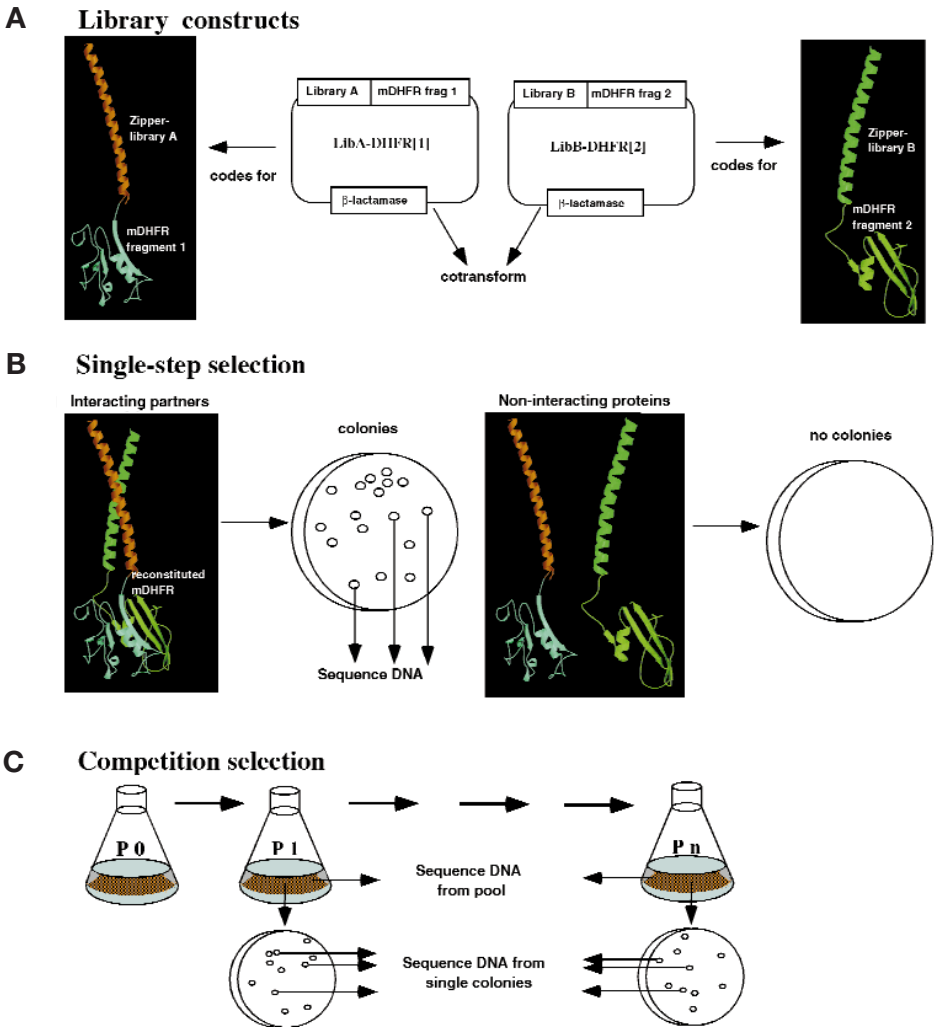

Figure 1. (A) DNA constructs code for fusions between library proteins (shown as α-helical leucine zippers) and either fragment of murine DHFR (mDHFR). Fusions were created using either the wild-type or the mutant mDHFR fragment 2 (I114A), yielding LibA-DHFR[1] and LibB-DHFR[2] or LibB-DHFR[2:I114A], respectively. (B) Principle of the mDHFR fragment complementation assay: *E. coli* cells are cotransformed with both fusion libraries in minimal medium, in the presence of IPTG (for induction of expression) and trimethoprim (for inhibition of the bacterial DHFR). If the library proteins heterodimerize, mDHFR can fold from the individual fragments resulting in active enzyme and bacterial growth. Both mDHFR fragments must be present, and dimerization of the fused proteins is essential, in order for cell propagation to be possible. No growth is observed if any of these conditions is not fulfilled[5]. The surviving colonies are the result of single-step selection and can be directly analyzed by DNA sequencing. (C) Competition selection is undertaken by pooling colonies from (B) in selective, liquid culture (passage 0 or P0), propagating the cells and diluting into fresh selective medium for further passages. An aliquot can be plated and the resulting colonies analyzed by DNA sequencing.

drive enzyme reconstitution. Finally, we introduced competitive metabolic selection, in which clones obtained with the second strategy were pooled and passaged through several rounds of competition selection, in order to enrich for the optimally heterodimerizing partners (Fig. 1C).

By simultaneously screening two libraries against each other, we illustrated the advantages of screening a large, combinatorial sequence space for identifying stably heterodimerizing pairs. We partially sampled a sequence space of 1.72×10^{10} combinations to select novel leucine zipper pairs with characteristics consistent with stable and specific heterodimerization. We directly demonstrated that the bias toward stability and specificity increased with increasing stringency of selection and observed the rate at which different sequence positions reach a consensus. In addition, the in vivo selection process ensured that solubility and stability toward proteolysis were essential for selection, yielding products suited to in vivo applications.

Results

We investigated a large-scale selection of dimerizing leucine zipper pairs from two designed semi-randomized libraries. These libraries are a hybrid between GCN4 and c-Jun/c-Fos (Fig. 2A), where the

Table 1. Stringency of the selection steps: selection factors

Single-step selection	**Selection factor[a]**		
Wild-type mDHFR fragments (5 or 20 ng)	2.8 ±1.5		
I114A mDHFR fragments (5 or 20 ng)	$(1.4 \pm 0.45) \times 10^2$		
Competition selection	**Selection factor[c]**	**Initial diversity**	**Frequency of dominant pair at P12[b]**
Competition (I114A)	3.2×10^6	3.9×10^6	WinZipA1-B1: 18/22 (82)
Shuffling: WinZip-A1 + LibB-DHFR[2:I114A]	8.7×10^4	1.3×10^5	WinZipA1-B2: 4/6 (67%)
Shuffling: WinZip-B1 + LibA-DHFR[1]	$\geq 1.3 \times 10^5$	1.3×10^5	WinZipA2-B1: 4/4 (100%)

[a]The selection factor in single-step selection is defined as the number of cotransformed cells plated (considering only the 50% that give combinations with no mutations or frameshifts), divided by the number of colonies surviving under selective conditions (see Results); average of two independent experiments, given with the standard deviation. This value must be calculated at low DNA concentrations (≤20 ng of each DNA), since the multiple cotransformations occurring at high DNA concentrations mask the actual selection factor.
[b]P12 is the twelfth round of serial cell passaging and competitive growth.
[c]The selection factor in competition selection is defined as the proportion of the dominant pair multiplied by the sequence diversity it was selected from, and is the result of a single experiment.

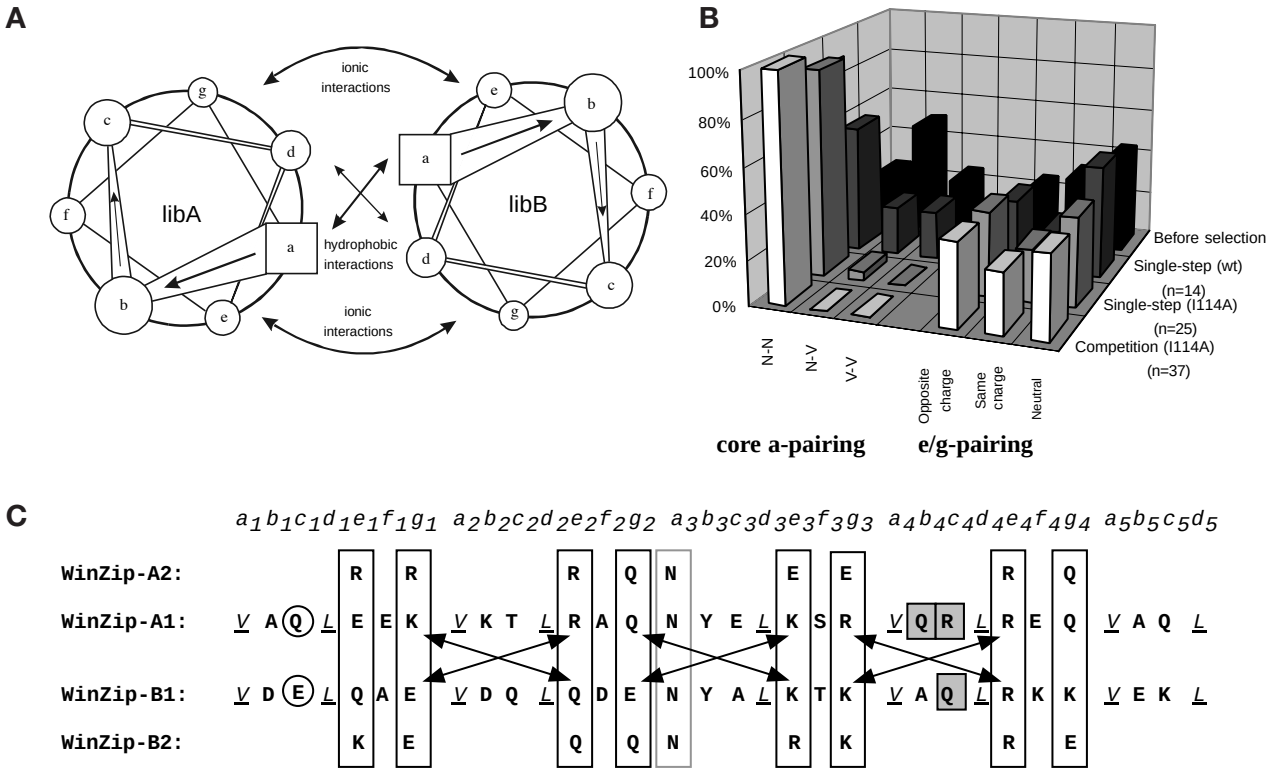

Figure 2. (A) Schematic representation of a leucine zipper pair visualized from the N-terminus illustrating e/g-interactions and the hydrophobic core formed by the a- and d-positions. (B) Distribution of residues at the semi-randomized positions throughout selection. The number of zipper pairs sequenced is given in parentheses, except "Before selection," where the theoretical distribution is reported. Each pair carries one core a-pair and six e/g-pairs. Neutral e/g-pairs have one or both residues as Gln. In "Competition (I114A)," only clones from P6 to P12 (not from earlier passages) were considered for analysis. Thus, 37 individual clones were identified, giving rise to 10 unique sequences due to multiple occurrences of the enriched clones. The distributions were calculated according to the frequency of sequence occurrence (*n* = 37). (C) Leucine zipper sequences obtained after competition selection and chain shuffling. The heptad positions (*a–g*) are followed by the heptad number (*1–5*). Invariant residues from GCN4 are underlined. Clear boxes indicate the semi-randomized e- and g-positions (black outline) and core a-position (a3) (gray outline). Circled residues were designed to contribute to helix capping. Shaded residues were designed for the introduction of restriction sites. Other residues are from c-Jun (LibA) or c-Fos (LibB). Arrows indicate putative e/g-interactions.

central, core a-position (a_3) was randomized to either Asn or Val, with equal probability, and the recurring e- and g-positions were randomized to Gln (neutral), Glu (acidic), or Arg or Lys (basic), each with 25% probability. This was achieved by synthesizing oligonucleotides containing synthetic codon building blocks[14]; the details of the design will be published elsewhere (K.M. Arndt, J.N. Pelletier, K.M. Müller, T. Alber, S.W. Michnick, and A. Plückthun, submitted). This library design allowed a number of complex optimization problems to be solved simultaneously by biological selection. At the core a-position the choice of Val-Val pairing, which confers higher thermodynamic stability to helix pairs, competes with Asn-Asn pairing, which confers specificity of parallel dimerization with a defined packing register and disfavors formation of antiparallel dimers and higher order oligomers[15,16]. In addition, the importance of charged residues at the e- and g-positions was investigated. Formation of salt bridges between these positions of opposite monomers has been observed via X-ray crystallography[17] and has been proposed to contribute to the stability of dimer formation[18–20]. Moreover, unfavorable electrostatic interactions between same-charged residues may be more important in driving stable, specific interactions and avoiding the formation of homodimers[21]. Furthermore, the energy of charged-neutral interactions has been shown to be similar to that of charged e/g-pairs in several cases[18,21]. Other factors, such as contribution of e/g-residues to helix propensity and helix dipole stabilization add to the difficulty of predicting the optimal e/g-pairs in dimerization even in simple model systems. Although a restricted

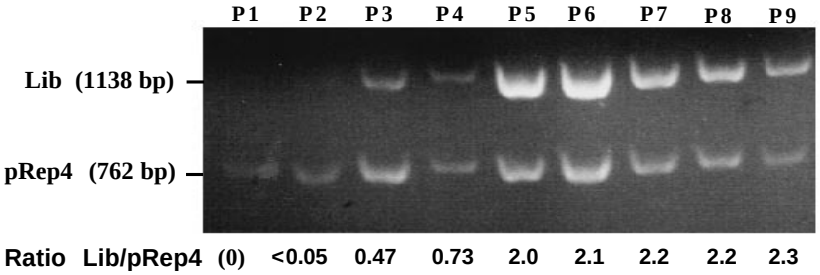

Figure 3. Efficiency of competition in a model selection. The selection was set up by mixing known numbers of cells expressing either GCN4-DHFR[1]/GCN4-DHFR[2:I114A] fusions or one of seven LibA-DHFR[1]/LibB-DHFR[2:I114A] pairs previously selected by single-step selection. The starting ratio was 2.9×10^4 : 1 (GCN4 to Lib). Competition selection was undertaken as described in Figure 1C and in the Experimental Protocol. The appearance of the library pairs in the pool was monitored by restriction analysis. A *Pvu*II fragment (1,138 bp) is unique to the LibB sequence of the LibB-DHFR[2] plasmid, while another (762 bp) is from pRep4 (repressor plasmid) and remains approximately constant. The bands were quantitated using the NIH Image gel analysis function to calculate the ratio of LibB/pRep4 (indicated below each lane).

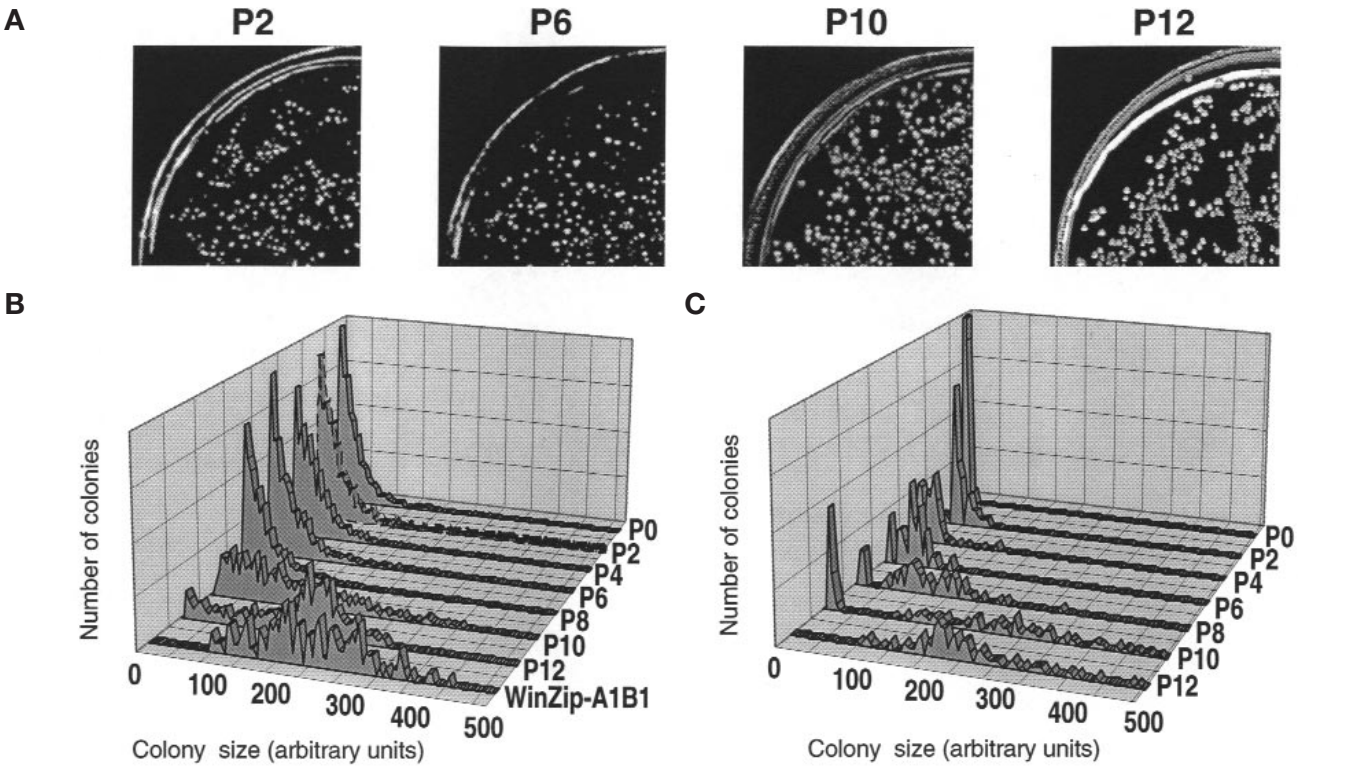

Figure 4. Competition selection and chain shuffling. (A) Approximately 1.42×10^4 clones resulting from single-step, I114A mutant selection were pooled (P0), and competition selection was undertaken as described in Figure 1C and in the Experimental Protocol. At each passage, some cells were plated and colony sizes were quantified. (B) Quantification of the colony sizes from (A). For comparative purposes, quantification of colony sizes of cells transformed with DNA of WinZip-A1B1 (but not passaged in liquid culture) is shown. (C) Quantification of the colony sizes from passages of the chain shuffling experiment: WinZip-B1-DHFR[2:I114A] + LibA-DHFR[1]. In (B) and (C) the numbers of colonies were normalized such that passages could be directly compared.

number of positions were semi-randomized here (four residue types at eight positions and two residue types at one position, resulting in 1.31×10^5 variants per library, and 1.7×10^{10} library-versus-library combinations), a problem of extraordinary complexity was generated, making predictions of the outcome very challenging. Resolution of this problem required a powerful selection strategy that could be performed and analyzed rapidly; to our knowledge the mDHFR fragment complementation system is currently the only strategy amenable to this.

Single-step selection. The semi-randomized designed leucine zipper libraries were subcloned into the appropriate vector harboring either mDHFR fragment (Fig. 1A; see also Experimental Protocol). As a first step in selection of heterodimerizing leucine zippers, a single-step selection was undertaken using the wild-type mDHFR fragments, by cotransforming the libraries LibA-DHFR[1] and LibB-DHFR[2] and plating on selective media (Fig. 1B). This strategy applies only a low stringency of selection to the potential pairs; thus, many library combinations were expected to be selected. Approximately 1.7% of the resulting ampicillin-resistant cells were doubly transformed, harboring at least one plasmid from each library, when we used 5 ng of each DNA, or 8% were doubly transformed when we used 20 ng of each DNA, as seen from control transformations (calculated as described in the Experimental protocol; data not shown). Of the doubly transformed cells that harbor no mutations or frameshifts, approximately 35% formed colonies under selective conditions (Table 1). This result demonstrates that even with a relatively low stringency of selection, only a fraction of the possible combinations of the two libraries allows zipper heterodimerization and efficient mDHFR reassembly.

Fourteen colonies resulting from two independent cotransformations were picked, and the sequences encoding the zippers were determined. Even under these low-stringency conditions there exist important sequence biases in these sequences relative to the unselected ones (Fig. 2B). A reduction in same-charged e/g-pairs from 31.3% (unselected) to 19% (selected) and an increase in opposite-charged pairs from 25% (unselected) to 31% (selected) were seen. In addition, a strong enrichment of Asn-Asn pairing at the core a-position (25% unselected vs. 57% selected) was observed. The characteristics that have been enriched are consistent with the selection of stable leucine zipper heterodimers.

Use of the mDHFR I114A mutation. In order to increase stringency, we repeated the single-step selection, using the I114A mutant of mDHFR[4,5]. We reasoned that only library partners that form the most stable heterodimers would compensate for the reduced ability of the mDHFR(I114A) fragments to fold into active enzyme. When bacteria were cotransformed with LibA-DHFR[1] and LibB-DHFR[2:I114A], we observed a 50-fold decrease in the number of colonies upon selective plating compared with the wild-type DHFR fragments (Table 1). Twenty-five colonies were picked from three independent cotransformations, and the DNA sequences were analyzed. The increase in selectivity correlated with an extremely strong selection for Asn-Asn pairing at the core a-position (92%; Fig. 2B), illustrating that the specificity of in-register parallel alignment provided by Asn-Asn pairing is more highly favored under these in vivo selection conditions than the higher stability afforded by Val-Val pairing. Reassembly of mDHFR from its fragments requires that the two fragment N-termini be brought close enough together to allow native-like refolding of DHFR (Fig. 1)[5,22]. The peptide linkers that connect the library sequences to the DHFR fragments must be sufficiently flexible to allow DHFR to fold, but not so long that any C-terminal to N-terminal orientation of the final folded leucine zipper would be allowed. As a result of this structural requirement, parallel

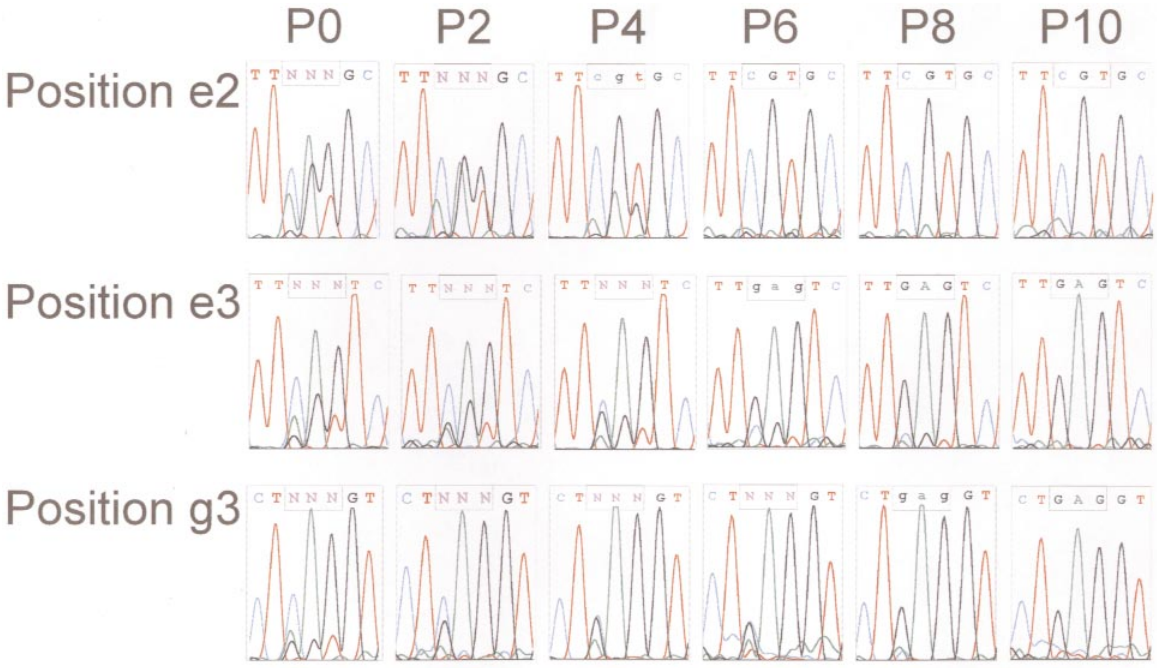

Figure 5. Sequencing profile of pools from passages of the chain shuffling WinZip-B1-DHFR[2:I114A] + LibA-DHFR[1]. Representative semi-randomized positions (see Fig. 2) were taken from a single competition experiment, such that the selection rates can be directly compared. The ratio of the individual triplet codons (central three nucleotides of each frame) was visually estimated (CAG = Gln; GAG = Glu; AAG = Lys; CGT = Arg; the equimolar random mix of the four codons results in the predominance of C at the first position, A at the second, and G at the third). Mixed positions are marked by (*N*), positions where a single codon is dominant (>50%) are marked in lowercase, and those where the codon is clear (>90%) are marked in uppercase. For passages 0, 2, and 8, two independent sequencing reactions were performed, which yielded identical results.

in-register heterodimerization of the library peptides is the only configuration possible. Other biases in these sequences were also more pronounced than with the wild-type DHFR fragments (Fig. 2B). In particular, an additional increase in opposite-charged e/g-pairs from 31% to 37% was seen. In one case, a point mutation resulted in a single clone (1/25) with a Val-Thr pair at the core a-position.

Competition selection: efficiency of selection. To further increase the selection pressure, we applied the principle of competition selection (Fig. 1C). We reasoned that, among selected zipper pairs, those that result in more stable heterodimerization will allow the most efficient enzyme reconstitution, leading to higher DHFR activity. If DHFR activity is limiting for growth, the higher activity should result in more rapid bacterial propagation; hence, these cells would become enriched in a pool. Thus, after sequential rounds of growth-competition, subtle differences in growth rate can be amplified, increasing the stringency of selection relative to the single-step selection.

To determine the rate at which competition can enrich for particular partner pairs, we first set up a model competition with a limited number of clones (Fig. 1C). The initial cell mixture (P0) contained known amounts of viable cells expressing either GCN4-DHFR[1]/GCN4-DHFR[2:I114A] or one of seven LibA-DHFR[1]/LibB-DHFR[2:I114A] pairs previously obtained in a single-step selection of those libraries, mixed at a ratio of $2.9 \times 10^4 : 1$ (GCN4 to library clones). Productive association of the homodimeric GCN4 pairs should occur only 50% of the time versus up to 100% for heterodimerizing library clones, which represents a disadvantage. Within three passages, the library pairs were already visibly enriched (Fig. 3), and after five passages the measured ratio between a restriction fragment indicative of the library and a constant fragment from the repressor plasmid had reached its maximium, showing that enrichment was maximal. Colonies resulting from passage 9 (P9) were sequenced. No GCN4 leucine zippers were present among 24 sequences analyzed. Therefore, enrichment of the library pairs over GCN4 by a factor of at least $24 \times (2.9 \times 10^4) = 7 \times 10^5$ was achieved. Four out of the seven library clones initially present survived until P9, with varying distributions (data not shown). The experiment was also repeated at a lower starting ratio of GCN4 to the 7 library pairs, and the same library clones were enriched, consistent with their enrichment being truly the result of selection (and not of unrepresentative sampling). This indicated that selection among the preselected clones was not as rapid as that seen between preselected and GCN4 zippers, but that the smaller differences between the preselected ones can still be amplified in selection. These results demonstrate that there is a direct link between reconstitution of mDHFR and growth rate.

Competition selection for optimal pairs. Our ultimate goal was to select for the "best" among the zipper pairs obtained by single-step selection. We obtained a large initial number of clones by cotransforming bacteria with 0.5 μg of DNA each from LibA-DHFR[1] and LibB-DHFR[2:I114A]. Approximately 50% of cells were at least doubly transformed (52% ± 10%, average of two independent control experiments, calculated as described in the Experimental Protocol). We obtained approximately 1.42×10^4 clones on selective medium, which arose from a 1.4×10^2-fold selection factor (Table 1), and were thus selected from $(1.42 \times 10^4) \times (1.4 \times 10^2) = 2.0 \times 10^6$ library-versus-library cotransformants. These were pooled and passaged. A clear increase in colony sizes was observed with subsequent passages, indicating that faster-growing clones were predominant (Fig. 4A and B). At P12, the colonies were homogeneously large, showing similar growth rates among the clones. We picked and sequenced 22 individual colonies from P12, as well as 11 from P10 and two from each previous second passage. We use the term WinZip to represent dominant zipper pairs obtained from competition selection. A single pair (WinZip-A1B1, composed of WinZip-A1-DHFR[1] and WinZip-B1-DHFR[2:I114A]) was identified 18/22 times (82%) in P12, 4/11 (33%) in P10, but not in previ-

ous passages (Fig. 2C). Other sequences were found in early and late passages, but none was as enriched as WinZip-A1B1. To verify that the growth rate recorded after competition (P12) was independent of bacteria-specific factors resulting from passaging, we cotransformed DNA from a pure clone of WinZip-A1B1 into fresh bacteria. The colony size distribution is similar for P12 and for the transformants (Fig. 4B), illustrating that the growth rate is a direct product of mDHFR reconstitution directed by the WinZip-A1B1 pair.

The sequence bias observed at the core a-position was even stronger here, with only the Asn-Asn pairing recorded at this position. When the biases at the e/g-positions were calculated according to the occurrence of each sequence (n = 37), there was no significant change in opposite-charged pairing (37%), whereas a small increase in same-charged pairing was observed (from 23% to 26%) as a result of the two same-charged pairs that occur in the predominant WinZip-A1B1 (Fig. 2B and C). However, when each unique sequence was considered only once (n = 10), a further increase of opposite-charged e/g-pairing was observed.

Chain shuffling of the WinZip-A1B1 sequences. In the above experiment, WinZip-A1B1 was selected from a sample representing 2.0×10^6 library-versus-library cotransformants. As the theoretical library-versus-library diversity was $(1.31 \times 10^5)^2 = 1.72 \times 10^{10}$, approximately 0.01% of the library-versus-library space was sampled. However, we obtained a very high coverage of either single library (theoretical complexity of 1.31×10^5), where the probability that all members are present at least once is p = 0.973. Thus, each polypeptide sampled only a small portion of the opposite library ($2.0 \times 10^6/1.31 \times 10^5 = 15.4$ polypeptides of the other library with p = 0.999, assuming equal transformation rates for both libraries) and it is likely that better combinations for the WinZip-A1B1 peptides may be found. Using WinZip-A1B1 as a partially optimized starting point, we combined each of the two WinZip-A1B1 polypeptides with the opposite library (WinZip-A1-DHFR[1] + LibB-DHFR[2:I114A] and WinZip-B1-DHFR[2:I114A] + LibA-DHFR[1]). Single-step selection yielded preselected pools for either competition. In both cases, the library (1.3×10^5) was overrepresented by a factor of 24 and 14, respectively, and the probability that all members were present at least once as partners of the "constant" peptide is $p > 0.999$ and 0.882, respectively. With passages of selection competition, a clear increase in colony sizes was again observed, indicating that faster-growing clones were predominating (Fig. 4C).

At P0 and each second passage, DNA from the entire pool of cells was sequenced in order to follow the rate of evolution of each library against a constant partner. Figure 5 illustrates the results from representative semi-randomized positions. It is clear that the rate of selection is not constant at all positions: some positions showed a dominant residue (>50%) already at P4 and clear selection (>90%) at P6 (see position e2), whereas others remained mixed (<50%) until P6 and became clear only at P10 (see position g3). This was observed in both selections. The sequences from individual colonies were analyzed, and in both selections, a predominant clone was identified (Table 1 and Fig. 2C) that was similar, but not identical, to the originally selected WinZip-A1B1 pair. The selection of the predominant clone WinZip-A2B1 (comprising the original partner B1 and the new partner A2, from selection of LibA-DHFR[1] against WinZip-B1-DHFR[2:I114A]) was achieved before P10, as P10 (four clones analyzed) and P12 (four clones analyzed) revealed only this clone. The selection of the predominant clone WinZip-A1B2 (comprising the original partner A1 and the new partner B2, from selection of LibB-DHFR[2:I114A] against WinZip-A1-DHFR[1]) was clear but not complete after 12 passages, as it was identified four out of six times in P12 and three out of five times in P10.

During the multiple passages performed in competition selection, the spontaneous acquisition of trimethoprim resistance by the *E. coli* DHFR could, in principle, lead to a false-positive result, in which survival would be independent of the mDHFR fragment complementation. While this phenotype was observed on one occasion at a frequency of approximately one resistant clone per 2×10^8 bacteria in single-step selection, it was never observed in clones resulting from competition selection, although up to 10^{12} cells were used during each competition. Thus, this phenotype does not interfere with the selection process.

We sequenced the regions N- and C-terminal to all zipper pairs obtained, including the promoter region and part of the mDHFR fragment–coding sequence (including residue 114). As well, the entire mDHFR fragment–coding sequence was verified in all WinZip clones. In no case was a mutation, rearrangement, or recombination of any constant portion of the constructs observed. In addition, all clones were subjected to restriction analysis and showed normal restriction patterns (data not shown). As in all in vivo strategies based on fusion proteins, we cannot preclude that the selected zippers could induce folding of mDHFR from its fragments or stabilize mDHFR through interactions of the leucine zipper with either the peptide linkers or with one or both of the DHFR fragments. However, the strong selection biases we observe, particularly the perfect selection for Asn-Asn pairing under conditions of high stringency but also of complementary e/g pairings, support our hypothesis that selection is determined by heterodimerizing leucine zipper-forming peptides.

Discussion

In our library-versus-library screen for heterodimerizing leucine zippers, selection was successful in both single-step and competition assays. Many combinations of the two libraries were expected to form heterodimers, albeit of varying stability. The 2.8-fold selection factor observed in single-step selection using the wild-type mDHFR fragments is consistent with the expectation that many of the combinations should result in functional heterodimers, since nine of the 10 a- and d-positions that define the hydrophobic core were invariant. Use of the I114A mutant of mDHFR increased the stringency of selection 50-fold, and competition selection allowed amplification of the most successful pairs from this pool. The sequence biases observed indicate that selection favored Asn-Asn pairing very strongly over Val-Val pairing in the hydrophobic core, consistent with selection for specificity of parallel, in-register dimerization. This in-register alignment allows the direct comparison of the selected zippers as all helices are forced, by the Asn-Asn pair, to assume a parallel orientation, juxtaposing the same e and g residues in all selected library members. Opposite-charged e/g -pairs were generally, but not exclusively, favored, suggesting that building stable zippers with good in vivo performance is more complex than simply designing opposite-charged pairs. The increasing colony sizes observed during competition are consistent with selection based on higher levels of reconstituted mDHFR activity. Our results suggest that competition selection could be undertaken as a continuous culture in automated protein evolution schemes and should be robust, as we have observed no genetic instabilities. We efficiently isolated a predominant individual clone (WinZip-A1B1) from approximately 2×10^6 individual combinations, taken from a 10^{10} combinatorial space. The biophysical characterization of this novel leucine zipper confirms that it is stable and strongly heterodimerizing (K.M. Arndt *et al*., submitted). Taken with the observed sequence biases and success in growth competition, it appears that there is a direct link between stability of zipper interaction and success in the selection process.

We obtained the WinZip-A1B1 pair from a partial sampling of the combinatorial space. To determine if WinZip-A1B1 could be improved, we performed a "chain shuffling" experiment. As each library was well represented in this second selection, the best match within each library for the given partner should have been found. In

both shuffling competitions, the population gradually converged to a predominant clone (WinZip-A1/WinZip-B2 and WinZip-A2/WinZip-B1) that was similar, but not identical, to the originally selected WinZip-A1B1 pair. This indicates that the partial sampling of the 10^{10} sequence space yielded a good, though not optimal, product (WinZip-A1B1), which was easily further improved by shuffling. The residues selected at the semi-randomized positions of these novel leucine zipper pairs differ somewhat from known natural zippers or designed zippers; yet, they behaved best in this system. It is likely that other factors, such as helix propensity and interactions of the charges with the helix dipole, also contribute to the stability. This underscores the advantages of semi-rational design accompanied by selection in an appropriate in vivo setting.

An important insight that can be gained from varying a library against a constant partner is how the selection occurs. Two scenarios can be envisaged. In the first, the selective pressure is not equivalent at all randomized positions, such that the rate of selection is rapid at certain positions whereas other positions remain semi-randomized longer. In the second scenario, selective pressure is applied against each polypeptide as a whole, perhaps because the identity of a residue at one position constrains the choice at other positions within the same helix. This would result in a population in which the rate of selection is independent of positional pressures arising from the partner. The results we obtain are clearly consistent with the first scenario. In particular, the core a-position showed the fastest rate of selection. While an in-depth analysis of the positional rate of selection is beyond the scope of this paper and will be presented in conjunction with the biophysical characterization of the optimized WinZip peptides (K.M. Arndt, J.N. Pelletier, K.M. Müller, S.W. Michnick, T. Alber, and A. Plückthun, unpublished data), we believe that the rate of selection is a reflection of the contribution of a residue at a given position, to the global coiled-coil interaction.

The DHFR fragment complementation assay has important advantages over selection techniques such as phage display in that it is possible to perform library-versus-library screening. Furthermore, the interactions occur in vivo rather than in vitro, which is important when in vivo performance is a quality of interest. Among in vivo–based selection strategies, it has the principal advantages of being fast and simple to execute and of having a direct link between protein–protein interactions and cell propagation, thus allowing selection rather than screening. The λ-repressor dimerization strategy[23] has been applied to selection schemes similar to that presented here[15,24], but this system requires complex experimental analysis and interpretation in order to distinguish homo- from heterodimerization, and again from formation of higher-order oligomers, thus precluding thorough characterization of a large number of pairs. While library-versus-library screening should be possible using the selectively infective phage system[25,26], this has not yet been demonstrated. The yeast two-hybrid strategy has been very successfully applied, among other tasks, to the process of systematically mapping interactions in yeast by successive rounds against single baits[27,28]. Although there is no a priori reason for not performing library-versus-library screens with the two-hybrid strategy, with the possible exception of less efficient transformation than of *E. coli*, such a screening strategy has not been reported to date. In addition, we have already demonstrated[5] that the DHFR fragment complementation assay can be used as a three-hybrid assay[29], in which a third partner (e.g., a small ligand) is required to mediate the protein–protein interaction. It should be possible to extend the assay such that the third partner is either an expressed protein or a specific RNA. This library-versus-library strategy will be an invaluable tool in defining networks of interacting polypeptides in functional genomics. The strategy is currently limited by the transformation efficiency of the *E. coli* strain, which could possibly be improved by using a bacterial host with better transformation yields or a more efficient transformation strategy.

The possibility of now screening two libraries against each other will allow for a much deeper exploration of complementary surfaces than is possible using a single-library approach. By using a semi-rational approach for screening a large number of interacting partners in a two-dimensional fashion, many factors contributing to interaction specificity will be identifiable.

Experimental protocol

All reagents used were of the highest available purity. Sequencing was carried out either by cycle sequencing with fluorescence labeling (MWG-Biotech, Germany) using a LiCor detection system or by automated sequencing with an ABI sequencer. Restriction endonucleases and DNA-modifying enzymes were from Pharmacia Biotech (Piscataway, NJ) and New England Biolabs (Beverly, MA). *Escherichia coli* strain XL1-Blue (Stratagene, La Jolla, CA) was used for subcloning and propagation of the libraries. *Escherichia coli* strain BL21 harboring the lacIq plasmid pRep4 (Qiagen, Basel) was cotransformed with the appropriate DNA constructs for the survival assays.

Constructs for DHFR fragment complementation. The DNA constructs encoding the N-terminal (1–107) and C-terminal (108–186) mDHFR fragments have been previously described[5]. Briefly, each fragment was amplified by PCR with appropriate unique flanking restriction sites and subcloned into a bacterial expression vector (pQE-32 from Qiagen). Each plasmid encodes an N-terminal hexahistidine tag, followed by a designed flexible linker and the appropriate DHFR fragment. Unique restriction sites between the hexahistidine tag and the flexible linker allow subcloning of the desired library. After subcloning, the resulting linker between either library and DHFR fragment was: $A(SGTS)_2STSSGI$ for LibA and $SEA(SGTS)_2STS$ for LibB. The design of the semi-randomized libraries is illustrated in Figure 2 and will be described in detail elsewhere (K.M. Arndt *et al*., submitted). Both libraries were produced using triplet-encoding oligonucleotides[14] and amplified by PCR, using primers carrying the appropriate unique restriction sites at each terminus, and the digested gel-purified products were ligated to the appropriate vector (Fig. 1). To achieve maximal library representation, the ligation mixes were individually electroporated into XL1-Blue cells and selected with ampicillin on rich medium (Luria Broth). A two- to sevenfold overrepresentation of each library was obtained. The resulting colonies were pooled and the plasmid DNA purified such that supercoiled plasmid DNA was obtained for cotransformation. In order to verify that the library populations encode the designed amino acids with the expected frequency, single clones from each library were randomly picked and sequenced before selection. No statistically significant biases were detected. With 70–80% of each library showing no mutations or frameshifts, the library-versus-library combination yielded approximately 50% correct sequence combinations. In cotransformations, the occurrence of double transformation was calculated as the number of colonies growing under selective pressure with trimethoprim (described below) divided by the number growing in its absence, when cotransformed with equal amounts of each DNA of a given, preselected pair.

Selection. Selective pressure for DHFR was maintained throughout all steps by inhibiting the bacterial DHFR with trimethoprim (1 μg/ml) in minimal medium. Ampicillin and kanamycin (100 μg/ml and 50 μg/ml, respectively) were also included in all steps to retain the library plasmids and the *lacI*^{q} repressor-encoding plasmid (pRep4), respectively. Expression of the proteins was induced with 1 mM isopropylthiogalactoside (IPTG). When selecting on solid medium, growth was allowed for 45 h at 37°C. When selecting in liquid medium, the starting OD_{600nm} was either 0.0005 or 0.0001. Cells were propagated either in Erlenmeyer flasks or in a 10 l; New Brunswick Scientific fermentor (Edison, NJ), depending on the volume required to ensure adequate representation of all clones present, at 37°C with shaking, or stirring at 250 r.p.m. After 10–24 h, OD_{600nm} reached 0.2–1.0, and cells were harvested. In competition selections, liquid culture was directly used to inoculate the next passage. We used BL21 cells with a transformation efficiency of no less than 5×10^7 transformants per μg of DNA using 200 pg of DNA, or 2×10^7 transformants per μg, using 500 ng of DNA. In cotransformations, the occurrence of double transformation was calculated as the number of colonies growing under selective pressure with trimethoprim divided by the number growing in its absence, when cotransformed with equal amounts of each DNA of a given, preselected pair.

Competition selection. When it was necessary to control precisely the starting number of cells in a competition, the number of viable cells in the starter cultures was quantified as follows. The appropriate clones were prop-

agated in liquid media under selective conditions, and dilute aliquots were frozen at -80°C with 15% glycerol. One aliquot for each clone was thawed and plated under selective conditions, and the colonies were counted after 45 h. The volume of cells to use for P0 was then calculated, such that each clone should be overrepresented by a factor of at least 2,000. Colony sizes (Fig. 4) were evaluated using the NIH Image Particle Analysis Facility.

Chain shuffling. DNA from the WinZip-A1B1 clone was isolated and retransformed into bacteria in order to obtain clones carrying either plasmid WinZip-A1-DHFR[1] or WinZip-B1-DHFR[2:I114A]. A pure clone (for each) was electroporated with the appropriate library. Library representation was calculated by comparison with control transformations of the same cells with DNA from the other WinZip-A1B1 polypeptide (calculated as the number of colonies growing in the presence of trimethoprim divided by the number growing in the absence). Single-step and competition selection were undertaken as described. It should be noted that cotransformation of bacteria at high DNA concentrations (0.5 μg per library) can lead to multiple plasmid transformation, where many survivors harbor more than one of either library sequence (data not shown). However, in no case was more than one sequence pair identified per clone after any competition selection, suggesting that multiply transformed cells retained only the pair of plasmids optimal for survival throughout the competition selection.

Acknowledgments

J.N.P. was a recipient of a fellowship from les Fonds de la Recherche en Santé du Québec, and currently holds a fellowship from le Conseil de Recherche en Sciences Naturelles et en Génie du Canada. K.M.A. is a recipient of a doctoral grant from the Stipendienfonds der Basler Chemischen Industrie. This work was funded by The Burroughs-Wellcome Fund (S.W.M.) and by grant 0311628 from the Bundesministerium für Bildung und Forschung (AP). We thank Günter Wellnhofer (Morphosys AG, Munich, Germany) for the synthesis of the oligonucleotides with the mixed trinucleotide building blocks. We thank Michel Denault for his painstaking help in probabilistic analysis, and François-Xavier Campbell-Valois, Alexis Vallée-Belisle, and Patrick Forrer for critical reading of the manuscript.

1. Fields, S. & Song, O. A Novel genetic system to detect protein-protein interactions. *Nature* **340**, 245–246 (1989).
2. Chien, C.T., Bartel, P.L., Sternglanz, R. & Fields, S. The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. *Proc. Natl. Acad. Sci. USA* **88**, 9578–9582 (1991).
3. Smith, G.P. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **228**, 1315–1317 (1985).
4. Pelletier, J.N., Remy, I. & Michnick, S.W. Protein-fragment complementation assays: a general strategy for the in vivo detection of protein-protein interactions. *J. Biomol. Techniques,* http://www.abrf.org/JBT/Articles/JBT0012/jbt0012.html.
5. Pelletier, J.N., Campbell-Valois, F.X. & Michnick, S.W. Oligomerization domain-directed reassembly of active dihydrofolate reductase from rationally designed fragments. *Proc. Natl. Acad. Sci. USA* **95**, 12141–12146 (1998).
6. Remy, I. & Michnick, S.W. Clonal selection and in vivo quantitation of protein interactions with protein fragment complementation assays. *Proc. Natl. Acad. Sci. USA* **96**, 5394–5399 (1999).
7. Remy, I., Wilson, I.A. & Michnick S.W. Erythropoietin receptor activation by a ligand-induced conformation change. *Science* **283**, 990–993 (1999).
8. Sydor, J.R., Engelhard, M., Wittinghofer, A., Goody, R.S. & Herrmann, C. Transient kinetic studies on the interaction of Ras and the Ras-binding domain of c-Raf-1 reveal rapid equilibration of the complex. *Biochemistry* **37**, 14292–14299. (1998).
9. Chen, J., Zheng, X.F., Brown, E.J. & Schreiber, S.L. Identification of an 11-kDa FKBP12-rapamycin-binding domain within the 289-kDa FKBP12-rapamycin-associated protein and characterization of a critical serine residue. *Proc. Natl. Acad. Sci. USA* **92**, 4947–4951 (1995).
10. O'Shea, E.K., Lumb, K.J. & Kim, P.S. Peptide 'Velcro': design of a heterodimeric coiled coil. *Curr. Biol*. **3**, 658–667 (1993).
11. Jelesarov, I. & Bosshard, H.R. Thermodynamic characterization of the coupled folding and association of heterodimeric coiled coils (leucine zippers). *J. Mol. Biol*. **263**, 344–358 (1996).
12. Zhou, N.E., Kay, C.M. & Hodges, R.S. The role of interhelical ionic interactions in controlling protein folding and stability. De novo designed synthetic two-stranded alpha-helical coiled-coils. *J. Mol. Biol*. **237**, 500–512 (1994).
13. Müller, K.M., Arndt, K.M., Strittmatter, W. & Plückthun, A. The first constant domain (C_H1 and C_L) of an antibody used as heterodimerization domain for bispecific miniantibodies. *FEBS Lett*. **422**, 259–264 (1998).
14. Virnekäs, B. *et al*. Trinucleotide phosphoramidites: ideal reagents for the synthesis of mixed oligonucleotides for random mutagenesis. *Nucleic Acids Res*. **22**, 5600–5607 (1994).
15. Zeng, X., Herndon, A.M. & Hu, J.C. Buried asparagines determine the dimerization specificities of leucine zipper mutants. *Proc. Natl. Acad. Sci. USA* **94**, 3673–3678 (1997).
16. Lumb, K.J. & Kim, P.S. A buried polar interaction imparts structural uniqueness in a designed heterodimeric coiled coil. *Biochemistry* **34**, 8642–8648 (1995).
17. O'Shea, E.K., Klemm, J.D., Kim, P.S. & Alber, T. X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science* **254**, 539–544 (1991).
18. Zhou, N.E., Kay, C.M. & Hodges, R.S. The net energetic contribution of interhelical electrostatic attractions to coiled-coil stability. *Protein Eng*. **7**, 1365–1372 (1994).
19. Monera, O.D., Kay, C.M. & Hodges, R.S. Electrostatic interactions control the parallel and antiparallel orientation of alpha-helical chains in two-stranded alpha-helical coiled-coils. *Biochemistry* **33**, 3862–3871 (1994).
20. John, M., Briand, J.P., Granger-Schnarr, M. & Schnarr, M. Two pairs of oppositely charged amino acids from Jun and Fos confer heterodimerization to GCN4 leucine zipper. *J. Biol. Chem*. **269**, 16247–16253 (1994).
21. Lumb, K.J. & Kim, P.S. Measurement of interhelical electrostatic interactions in the GCN4 leucine zipper. *Science* **268**, 436–439 (1995).
22. Buchwalder, A., Szadkowski, H. & Kirschner, K. A fully active variant of dihydrofolate reductase with a circularly permuted sequence. *Biochemistry* **31**, 1621–1630 (1992).
23. Hu, J.C., O'Shea, E.K., Kim, P.S. & Sauer, R.T. Sequence requirements for coiled-coils: analysis with lambda repressor-GCN4 leucine zipper fusions. *Science* **250**, 1400–1403 (1990).
24. Zeng, X., Zhu, H., Lashuel, H.A. & Hu, J.C. Oligomerization properties of GCN4 leucine zipper e and g position mutants. *Protein Sci*. **6**, 2218–2226 (1997).
25. Spada, S. & Plückthun, A. Selectively infective phage (SIP) technology: a novel method for in vivo selection of interacting protein-ligand pairs. *Nat. Med*. **3**, 694–696 (1997).
26. Rudert, F., Woltering, C., Frisch, C., Rottenberger, C. & Ilag, L.L. A phage-based system to select multiple protein-protein interactions simultaneously from combinatorial libraries. *FEBS Lett*. **440**, 135–140 (1998).
27. Bartel, P.L., Roecklein, J.A., SenGupta, D. & Fields, S. A protein linkage map of Escherichia coli bacteriophage T7. *Nat. Genet*. **12**, 72–77 (1996).
28. Fromont-Racine, M., Rain, J.C. & Legrain, P. Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nat. Genet*. **16**, 277–282 (1997).
29. SenGupta, D.J. *et al*. A three-hybrid system to detect RNA-protein interactions in vivo. *Proc. Natl. Acad. Sci. USA* **93**, 8496–8501 (1996).