Unbiased selection of localization elements reveals cis-acting determinants of mRNA bud localization in *Saccharomyces cerevisiae*

Ashwini Jambhekar*, Kimberly McDermott†, Katherine Sorber*, Kelly A. Shepard‡, Ronald D. Vale‡§¶, Peter A. Takizawa†, and Joseph L. DeRisi*§

Departments of *Biochemistry and Biophysics and ‡Cellular and Molecular Pharmacology, and §Howard Hughes Medical Institute, University of California, San Francisco, CA 94107; and †Department of Cell Biology, Yale University School of Medicine, New Haven, CT 06520

Contributed by Ronald D. Vale, October 22, 2005

Cytoplasmic mRNA localization is a mechanism used by many organisms to generate asymmetry and sequester protein activity. In the yeast *Saccharomyces cerevisiae*, mRNA transport to bud tips of dividing cells is mediated by the binding of She2p, She3p, and Myo4p to coding regions of the RNA. To date, 24 bud-localized mRNAs have been identified, yet the RNA determinants that mediate localization remain poorly understood. Here, we used nonhomologous random recombination to generate libraries of sequences that could be selected for their ability to bind She-complex proteins, thereby providing an unbiased approach for minimizing and mapping localization elements in several transported RNAs. Analysis of the derived sequences and predicted secondary structures revealed short sequence motifs that mediate binding to the She complex and RNA localization to the bud tip *in vivo*. A predicted single-stranded core CG dinucleotide appears to be an important component of the RNA–protein interface, although other nucleotides contribute in a context-dependent manner. Our findings further our understanding of RNA recognition by the She complex, and the methods used here should be applicable for elucidating minimal RNA motifs involved in many other types of interactions.

three-hybrid selection | nonhomologous random recombination | RNA zipcode

Localization of mRNA is commonly used to target proteins to specific regions within a cell. In most cases, this process requires recognition by RNA-binding protein(s) and linkage of the resulting RNA–protein complex directly or indirectly to molecular motors (1). The determinants of recognition, transport factor binding, and subsequent targeting are cis-acting sequences often found in untranslated regions. Precise characterization of these RNA "zipcodes" has proven to be cumbersome for several reasons. The reported length of the minimal sequence requirements for transport ranges from 50 nucleotides (nt) to several hundred, and this apparent complexity is compounded by functional redundancy among zipcodes and a diversity of cellular recognition components (2–5).

The yeast *Saccharomyces cerevisiae* provides a tractable model system to characterize the determinants of zipcode recognition. To date, 24 bud-localized mRNAs have been identified, and coding regions were shown to mediate transport (6). Localization depends on the She complex, which comprises She2p, a putative RNA binding protein, Myo4p, a type V myosin motor, and She3p, which interacts directly with both Myo4p and She2p (7–9).

Independent studies of one transported RNA, *ASH1*, identified three (N, C, and U) (10) or four (E1, E2A, E2B, and E3) (11) zipcodes based on their ability to mediate localization of a reporter. Only one of the elements lies in the 3′ UTR; the remaining are located within the coding region. These elements bear no obvious primary sequence or secondary structural similarity to each other, and mutational analysis suggested that secondary structure was required for activity (10, 11). Recently, Olivier *et al.* (12) reported that a CGA triplet in a loop, along with a single-stranded cytosine six bases away and opposite to the triplet, was necessary for bud localization of *ASH1* and two other RNAs. However, these criteria are insufficient to identify zipcodes in other RNAs localized by the She complex (6).

To extend our understanding of the She-complex–RNA interaction, we used an unbiased approach to select zipcode-containing fragments from pools of known localized RNAs. The fragments were tested for localization *in vivo*, and bona fide zipcodes were subjected to further analysis, which revealed a highly degenerate motif predicted to lie in single-stranded regions and is necessary for She-complex-dependent transport. Highlighting the complexity of the She2/3p–RNA interaction, we also found that the precise sequences mediating recognition and transport depend on the context of the adjacent sequence and structural features in the mRNA.

Materials and Methods

Nonhomologous Random Recombination (NRR). NRR was carried out as described in ref. 13 (see *Supporting Materials and Methods*, which is published as supporting information on the PNAS web site, for oligonucleotide sequences). Thirty picomoles of T7hairpin, or 15 pmol each of XmaHairpin1 and SphHaripin1, were used for NRR ligations. Ligated DNA was digested with PmeI to remove hairpin ends. Ten percent of the restriction digest was used for PCR with 1 μM XmaT7 primer or NRRprimer1. NRR was carried out with T7hairpin and *ASH1*, *YLR434c*, *ERG2*, or *MID2* sequences separately. XmaHairpin1 and SphHairpin1 were used for separate NRR reactions with *CPS1*, *DNM1*, *WSC2*, *MMR1*, or *YGR046w*, or with a pool of *ERG2*, *MID2*, and base pairs 1–1000 and 1500–1761 of *TPO1*. Base pairs 1000–1500 of *TPO1* were excluded because they do not to contain any zipcodes (data not shown). The full coding region of each gene was used for NRR (unless noted), except for *ASH1*, which included 99 bp of downstream sequence.

Three-Hybrid RNA-Expression Library Construction. NRR products were ligated into the XmaI site (for XmaT7 products) or asymmetrically into XmaI and SphI sites (for NRRprimer1 products) of pIIIΔA/MS2.2. This vector consists of pIIIA/MS2.2 (14) with a deletion of the AatII-Tth111-I fragment encoding *ADE2*. In all cases, the library size was sufficient to ensure that every sequence was represented at least once. Separate libraries were constructed for *YLR434c* and *ASH1*. Another library contained NRR products derived from a pool of *ERG2*, *MID2*, and bases 1–500 and 1500–1761 of *TPO1*. A fourth

Conflict of interest statement: No conflicts declared.

Freely available online through the PNAS open access option.

Abbreviations: HA, hemagglutinin; NRR, nonhomologous random recombination; 3-AT, 3-aminotriazole.

¶To whom correspondence should be addressed. E-mail: vale@cmp.ucsf.edu.

CELL BIOLOGY

library contained *ERG2*, *MID2*, *WSC2*, *DNM1*, *CPS1*, *YGR046W*, *MMR1*, *YMR171C*,and *SRL1* NRR products. Each library was screened separately by three-hybrid analysis (15).

For randomization experiments, complementary oligonucleotides fully degenerate at the indicated positions were annealed and cloned into the NotI and XhoI sites of pAJ232, which consists of pIIIΔA/MS2.2 with an insertion of NotI and XhoI in the XmaI site. Plasmid library members were selected at 10 mM 3-aminotriazole (3-AT) as described below.

Three-Hybrid Selection. DNA encoding the carboxyl terminus of She3 (base pairs 706-1278) was cloned into XmaI/SacI sites of Gal4-AD expression vector pACT2 and introduced into the three-hybrid L40 coat host strain (15). Where indicated, *SHE2* was deleted in L40 coat as described in ref. 16. RNA plasmid libraries (12–20 μg) were transformed into the She3-L40 coat strain by using the lithium acetate method (17). Transformants were plated on SD-HIS-URA medium containing 6.67 mg/liter adenine and 0, 0.5, 1, 5, 10, or 15 mM 3-AT. All transformants were screened for dependence on the *SHE3* plasmid for three-hybrid activity (*Supporting Materials and Methods*). Thirty to eighty *SHE3*-dependent colonies selected at the highest 3-AT concentrations were tested for LacZ expression by X-Gal filter assay (14). All tested candidates expressed LacZ (data not shown). Plasmids were rescued, and inserts were fully sequenced from the 5′ end. Quantitative β-gal assays were performed as described in ref. 14, except that cells were lysed with Yeast Protein Extraction Reagent (Pierce).

Visualization of RNA. The U1A-GFP system was used for visualizing RNA localization *in vivo* (6, 8). RNAs >150 nt were cloned directly into the pGAL-U1A vector (6) containing NotI and XhoI cloning sites. Shorter RNAs were assayed by fusing to the 3′ end of the unlocalized *ADH1* gene. For RNAs <75 nt, a linker containing a 13-bp inverted repeat separated by NotI and XhoI sites was inserted downstream of *ADH1*. Synthetic oligos (Operon Biotechnologies, Alameda, CA) encoding the target RNA sequences were ligated into the NotI and XhoI sites so that the RNA was expressed with flanking inverted repeats that formed a stable helix.

For visualization of RNA, the pGAL-U1A plasmid containing the RNA of interest was introduced into a W303 yeast strain harboring the U1A-GFP plasmid (6, 8). More than 50 premitotic cells expressing RNA were counted from two independent transformants for each RNA as described in ref. 6.

RNA Structure Predictions. All RNA structure predictions were computed by using MFOLD (18, 19)

Protein Purification and Gel Shifts. She2p-HA (HA, hemagglutinin) contains a single HA epitope at its C terminus. She2p-HA was overexpressed in *S. cerevisiae* and isolated from cell extracts with anti-HA antibodies coupled to protein A Sepharose (Sigma). She2p-HA was eluted from the resin with excess HA peptide, dialyzed to remove free peptide and concentrated in a Microcon YM-10 (Millipore). His-She3p 251–425 contains a His-6 tag at the N terminus of amino acids 251–425 of She3p. His-She3p 251–425 was expressed in BL21 RIPL (Stratagene) and purified with Ni-NTA agarose (Qiagen) according to the manufacturer's instructions.

To generate ^{32}P-labeled RNAs for mobility shifts, annealed oligos containing a T7 promoter followed by a particular zipcode sequence were used as templates in an *in vitro* transcription reaction. The oligo templates were added to a Maxiscript T7 (Ambion) reaction containing UTP-^{32}P (Amersham Pharmacia). Full-length RNAs were gel purified from the reactions. Each gel shift reaction contained 0.5 nM labeled RNA, 0.1 mg/ml tRNA in 25 mM Hepes·KOH (pH 7.5), 100 mM KCl, 2 mM $MgCl_2$, and 1 mM DTT. Purified She2p-HA and His-She3 251–425 were added at varying concentrations. Reactions were incubated at room temperature for 30 min and then run on a 5% acrylamide gel (37.5:1) in 90 mM Tris/64.6 mM boric acid/2.5 mM EDTA, pH 8.3 at 4°C. The gel was fixed, dried, and exposed to film.

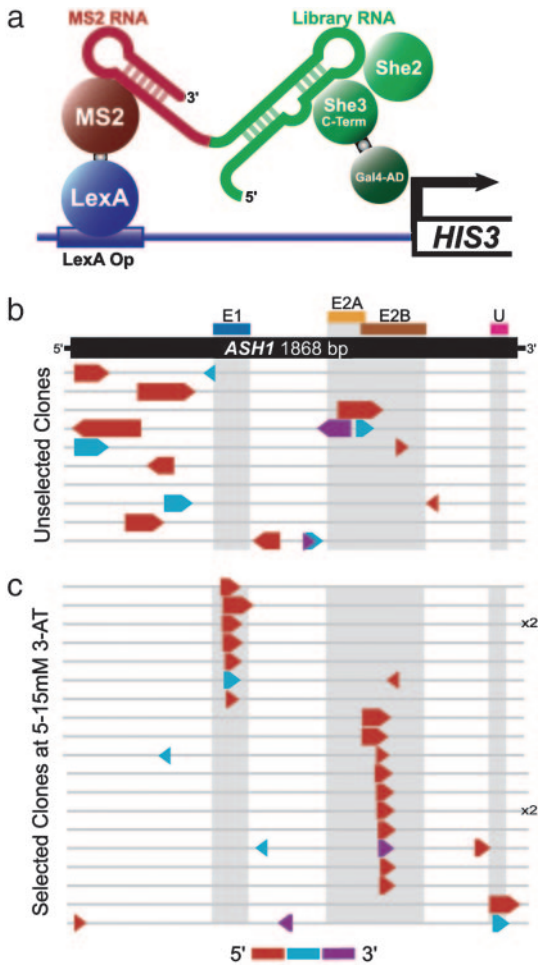

Fig. 1. Three-hybrid scheme for selection of She3p-interacting RNA fragments. (*a*) Schematic of three-hybrid assay and representation of *ASH1* NRR library members before (*b*) and after (*c*) three-hybrid selection. Each arrow represents a fragment from *ASH1*. The direction of the arrowhead indicates whether the fragment is expressed in the sense (right) or antisense (left) orientation from the three-hybrid RNA expression vector. The position of each arrow corresponds to the location of the fragment within the gene, and arrow colors indicate the connectivity of the fragments in the clone. Clones recovered in more than one independent yeast transformant are indicated.

Results

Identification of She-Complex-Dependent Localization Sequences. We sought to identify short zipcodes from known transported RNAs in a high-throughput manner without making assumptions about exact zipcode length, orientation, or connectivity. For this reason, we used NRR (13) to generate libraries of sequences that could be selected for their ability to bind to She-complex proteins. We reasoned that the region of overlap of multiple, independently selected clones would define a short zipcode.

To generate a library by NRR, DNA encoding a target RNA was digested with DNaseI, and 20–200 bp fragments were isolated and ligated in the presence of hairpin linkers to generate products containing one to three tandem fragments of various sizes and connectivities flanked by hairpins. The products were PCR amplified with primers complementary to the linker sequence and selected for interaction with the She complex by three-hybrid assay (Fig. 1*a*). As bait, we used the carboxyl terminus of She3p, which

Table 1. Summary of elements identified by NRR/three-hybrid selection

Zipcode	Coordinates	Length, nt	Three-hybrid activity	No. of times recovered	Percent localized
E1min*	635–683	49	+++	8	>90
E2Bmin*	1279–1314	36	+++	11	>90
Umin*	1766–1819	54	+	2	>90
Other*	1684–1719R	36	++	1	N/D
WSC2N	418–471	54	++	14	>90
WSC2C	1313–1384	72	++	6	>90
ERG2N	180–250	71	++	24	>90
DNM1N	605–805	201	+	1	70–80
DNM1C	1656–1752	97	+	1	>90
SRL1C	419–596	178	+	6	>90
YLR434-1	[*21–55*][195–209]	50	+	15	70–80
YLR434-2	[*138–186*][*56–90*]	76	+	11	>90
TPO1N	2–178	177	±	6	70–80
CPS1CR	1305–1456R	152	+	1	<60

Coordinates indicate the smallest overlapping fragment common to all sequences isolated for each zipcode. Nucleotides are numbered with the adenosine of the start codon as +1. *, sequences derived from *ASH1*. When multiple fragments were contained in one clone, the fragments are listed in 5′ to 3′ order. Fragments in italics were cloned in the antisense orientation. The length of each clone is given in nucleotides. Activity in the three-hybrid assay was assessed by the highest 3-AT concentration at which the sequence was recovered. ±, 1mM; +, 5mM; ++, 10mM; +++, 15mM 3-AT. Also shown is the number of recovered clones containing the indicated sequence. Percent localized refers to the percent of cells with exclusively bud-localized RNA. N/D, not determined.

interacts with She2p (7) and displays proper specificity for RNA targets (9) (vector and IRE controls, Fig. 6, which is published as supporting information on the PNAS web site). For the two RNAs tested, the three-hybrid interaction also required endogenous She2p (*she2* WSC2N and *she2* Umin, Fig. 6), indicating the formation of a tripartite RNA–protein complex.

To validate this approach, we subjected *ASH1* to NRR and three-hybrid selection. Sequencing of NRR-generated clones before selection revealed fragments derived from various parts of the gene (Fig. 1*b*). After selection, almost all clones fell within previously identified localization elements (Fig. 1*c*). Although no sequences were recovered from E2A, this zipcode is active in the three-hybrid system (12), so its absence in our selection most likely resulted from insufficient sequencing of positive transformants. Only one selected clone did not contain a fragment overlapping known localization elements and was not pursued further. In all cases, the sequences defined by selected overlapping clones were shorter than the zipcodes from which they were derived (10, 11). To verify that the shorter sequences localized *in vivo*, we used the U1A-GFP system (6, 8) to visualize RNA distribution in live cells. Sequences <150 nt in length were fused to the 3′ end of *ADH1* and assayed for their ability to direct bud localization of the RNA. All *ASH1* sequences defined by the NRR/three-hybrid selection localized to bud tips in >90% of cells (Table 1 and Fig. 2 *b*, *c*, and *e Insets*).

Ten other genes encoding localized RNAs were screened in this manner individually (*YLR434c*) or in pools (*ERG2*, *MID2*, *TPO1*, *WSC2*, *MMR1*, *SRL1*, *CPS1*, *DNM1*, and *YGR046w*), and 10 more putative zipcodes were identified ranging from 50 (YLR434–1) to 201 (DNM1N) nt in length (Table 1). All sequences defined by overlapping clones were tested for localization *in vivo*. Although the control *ADH1* reporter was localized in only 20% of cells, our experience with testing various constructs has revealed that, in rare cases, unlocalized RNAs can produce dim, bud-localized particles in up to 60% of cells in a She2p-independent manner. Thus, we classified any RNA that was localized in <60% of cells as unlocalized. Only one selected RNA, CPS1CR, failed to localize by this criterion. Of the remainder, nine sequences localized in >90% of cells in a She2p-dependent manner (Table 1 and data not shown). Two others, TPO1N and DNM1N, localized less efficiently (in 70–80% of cells). In general, sequences recovered multiple times at high 3-AT concentrations were more likely to localize than those recovered once or only at low 3-AT concentrations (Table 1; see also Table 2, which is published as supporting information on the PNAS web site). Although some zipcodes were recovered numerous times, we failed to recover any zipcodes from *CPS1*, *MID2*, *MMR1*, or *YGR046w*, suggesting that the screen was not saturating.

Identification of a Conserved She2/3p-Dependent Localization Motif. We used MEME analysis, which identifies statistically overrepresented sequence motifs within a data set (20), to find any motifs shared by the newly identified zipcodes. The data set consisted of nine zipcodes displaying >90% localization activity, including two (WSC2N and YLR434–2) that had been minimized by deletion

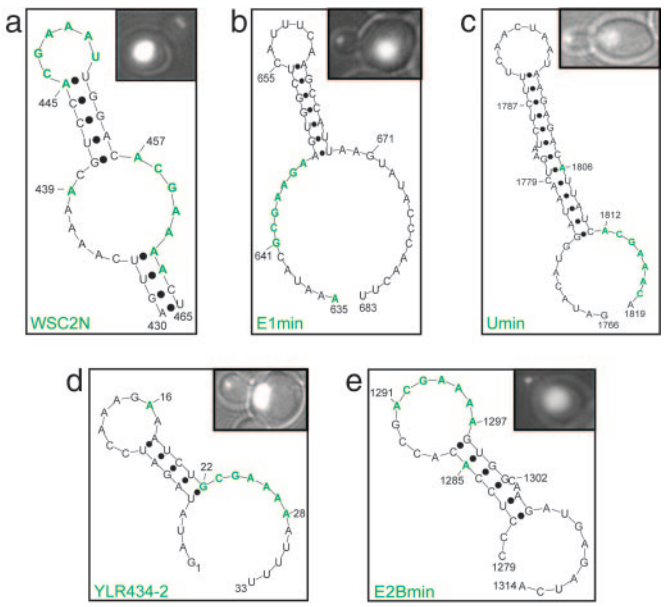

Fig. 2. Sequences and predicted structures of WT zipcodes. Bases identified by MEME analysis are green. (*a*) WSC2N. (*b*) E1min. (*c*) Umin. (*d*) YLR434–2. (*e*) E2Bmin. Bases are numbered with the adenosine of the start codon as +1, with the exception of YLR434-2, which is numbered with the 5′ base as +1. Insets contain representative GFP-RNA localization images. RNA particles are cytoplasmic; excess, unbound U1A-GFP is sequestered in the nucleus.

CELL BIOLOGY

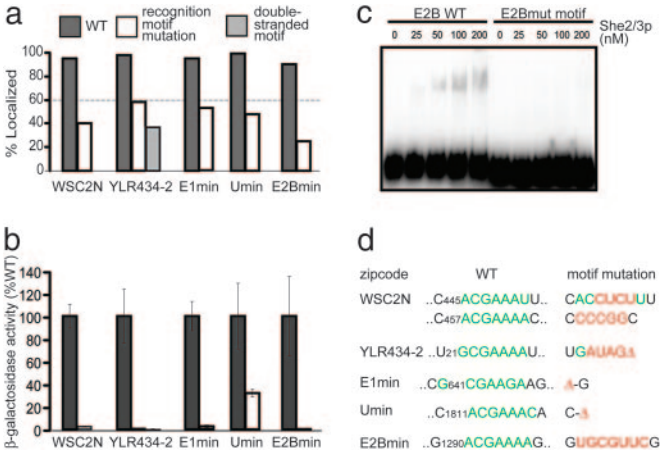

Fig. 3. The recognition motif mediates She2/3-dependent localization and binding. (*a* and *b*) Localization ability (*a*) and three-hybrid β-gal activity (*b*) of zipcodes in Fig. 2 containing WT bases or mutations in the recognition motif. ''Double-stranded motif' indicates that the recognition motif is in an ectopic duplex. Dashed line in *a* indicates the threshold below which RNAs were considered unlocalized. (*c*) *In vitro* binding of She2p and She3p to E2Bmin. RNA mobility shift assay consists of WT or mutant RNA lacking the recognition motif with increasing concentrations of purified She2p-HA and His-She3p carboxyl terminus. (*d*) WT and mutant motifs sequences used in *a–c*. Motif bases are green and mutations are red.

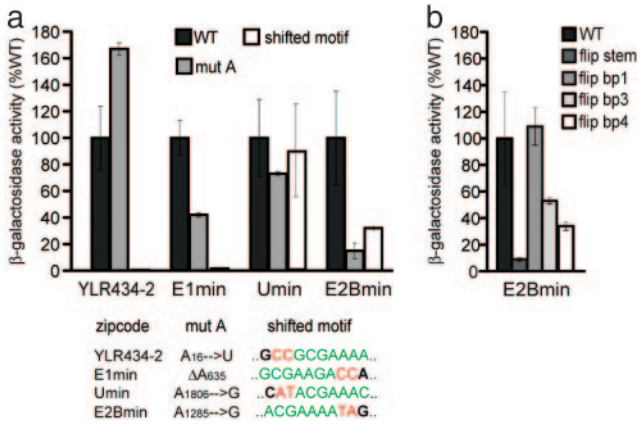

Fig. 4. Context-dependency of recognition motif is revealed by mutational analysis. (*a*) Three-hybrid β-gal activity of zipcodes bearing mutations in the upstream adenosine (mutA) or 2-nt insertions between the motif and adjacent helix (shifted motif). Mutations are defined for each zipcode. Motif bases are in green, insertions are red, and duplex bases are bold. (*b*) β-gal activity of E2Bmin sequences containing mutations in the stem. ''Flip'' indicates that each base of the indicated pair was changed to its partner. The ''flip stem'' RNA contains compensatory mutations along the entire stem. Base pairs are numbered with the bottom of the stem as 1.

mapping (Fig. 2). Of several candidates, one degenerate motif (RCGAADA) was present in all input sequences and mapped almost exclusively (in seven of eight cases) to single-stranded regions of the secondary structures predicted by MFOLD (18, 19) (Fig. 2; see also Fig. 7, which is published as supporting information on the PNAS web site). One zipcode, WSC2N, displayed two copies of the motif, a more degenerate version in the terminal loop and a consensus sequence in the 3′ bulge (Fig. 2*a*). Additionally, seven zipcodes contained an adenosine six bases upstream of the motif. This sequence pattern was observed in three other zipcodes not included in the MEME analysis (E2A in *ASH1* and zipcodes in *IST2* and *YMR171c*; ref. 12).

Five zipcodes were selected for further analysis based on the fact that the seven-base motif could be mutated or deleted in these RNAs without affecting the predicted structure of the remainder of the molecule (Fig. 2; see also Table 3, which is published as supporting information on the PNAS web site). Wild-type (WT) zipcodes localized in >90% of budded cells (Fig. 3), and displayed β-gal activities >200 Miller units (Fig. 6). All zipcodes required the motif for localization and LacZ expression (Fig. 3 *a*, *b*, and *d*). Deletions or mutations of the motif in E1min, E2Bmin, and YLR434–2 abolished activity in both assays. Deletion of the motif in Umin also abolished localization but decreased β-gal activity by only 65% (Fig. 3 *a* and *b*). WSC2N, which contains two copies of the motif, required mutations in both to abolish localization and β-gal activity (Fig. 3 *a* and *b*; see also Fig. 8, which is published as supporting information on the PNAS web site).

The ability of purified She2p and the carboxyl terminus of She3p (251–425) to bind WT and mutant zipcodes directly was tested also by RNA mobility shift. Nanomolar concentrations of She2p and She3p retarded the mobility of all WT zipcodes, indicating that She2/3 bind directly to each zipcode (Fig. 3*c*; see also Fig. 9, which is published as supporting information on the PNAS web site). Furthermore, the protein complex displayed sequence-specific binding, because mutations of the motif in Umin, YLR434–2, E2Bmin, and E1min decreased or abolished the shift (Figs. 3*c* and 9). Although a large amount of WT RNAs remained unbound at the highest protein concentrations, it is unlikely that additional proteins facilitate She-complex binding to RNA *in vivo*, because a limited number of proteins, like She2p (21), are present in both the nucleus and cytoplasm to facilitate bud localization and three-hybrid activity. It is more likely that some of the RNA misfolds and cannot bind She2/3 *in vitro*. Nevertheless, we conclude that the degenerate motif is essential for RNA binding of She2/3 and that activity in localization and β-gal assays reflects binding of the RNA to the She complex.

In addition to the recognition motif, MEME analysis identified an adenosine six bases upstream in seven zipcodes. Mutation or deletion of this base caused varying effects on three-hybrid activity ranging from an increase (YLR434–2min) to a 5-fold reduction (E2Bmin) (Fig. 4*a*). Some base substitutions may be more favorable than others at this position, resulting in the range of phenotypes displayed by the mutations in different zipcodes. Although this adenosine was highly conserved among the zipcodes, its contribution to binding was context-dependent.

Although the mutational analyses revealed that the primary sequence of the motif was essential for zipcode activity, they did not address structural requirements for She2/3 recognition. To determine whether the single-stranded nature of the motif was necessary for She-complex recognition, the 5′ end of YLR434-2 was changed to complement the motif at the 3′ end, placing the motif in a predicted duplex. The resulting RNA (YLR434-2 double-stranded motif) failed to localize *in vivo* and did not display significant three-hybrid β-gal activity (Fig. 3 *a* and *b*), indicating that the She complex cannot bind its recognition site in a stable helix. We also observed that the recognition motifs bordered predicted helices in most zipcodes. To test whether this juxtaposition was essential, 2 nt were inserted between the stems and motifs of four zipcodes. The resulting mutant phenotypes ranged from no decrease in β-gal activity (Umin) to a complete abolition of She-complex interaction (E1min, YLR434–2) (Fig. 4*a*).

Although the above results implied that the stems of zipcodes were important for She2/3 binding, no primary sequence similarities were observed in these regions. The current models (10–12) proposed that stems play only a structural role in the RNA–protein interaction. In support of the model, compensatory mutations in the stem of YLR434–2 preserved zipcode function (Fig. 10, which is published as supporting information on the PNAS web site), but similar mutations in E2Bmin abolished three-hybrid activity (Fig. 4*b*). Therefore, each base pair in the E2Bmin stem was individually

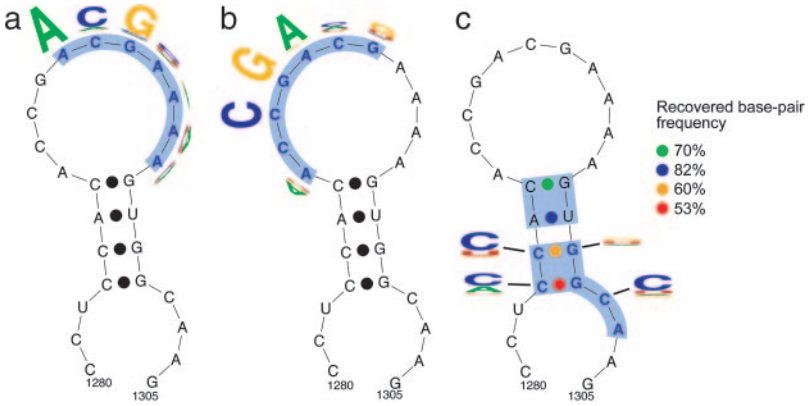

Fig. 5. Predicted secondary structure for E2Bmin and sequence logos derived from randomization and three-hybrid selection of bases 1291–1297 (*a*), 1287–1293 (*b*), or 1283–1286 and 1298–1303 (*c*). The height of each letter is proportional to the fraction of the observed frequency relative to the expected frequency at each position (22, 23). The color of each dot in *c* indicates the frequency of base-pairing among the selected clones.

mutated to identify essential bases. Mutation of each of the two base pairs adjacent to the loop decreased β-gal activity 2- to 4-fold, whereas mutating the pair at the base of the stem had no effect. (The C_{1284}·G_{1300} pair was not tested because substitutions were predicted to disrupt the entire stem.) Surprisingly, no single base-pair mutation decreased activity to the extent that mutation of the entire stem did. These results indicated that the primary sequence of the stem contributes to She2/3 binding in some cases and that bases in the stem of E2Bmin contribute in an additive manner. Collectively, these results support the role of the degenerate, single-stranded motif in mediating She-complex recognition; however, the precise sequence and topological requirements appear to be context-dependent.

Analysis of Base Contributions Within a Single Zipcode. Because it seemed that conserved bases in the recognition motif and other, less-conserved bases contributed to She2/3 binding, we investigated the sequence requirements for She2/3 binding to a single zipcode. Four- to seven-base regions of a further-minimized E2Bmin zipcode were fully randomized, and resulting sequences were selected for She-complex binding by the three-hybrid system.

The contribution of each base in the loop of E2Bmin was determined via two separate, overlapping randomization/selection experiments. One position in the loop (1288) displayed no base preferences for She-complex recognition (Fig. 5*b*). In contrast, six of seven bases in the WT motif were significantly overrepresented upon selection (Fig. 5*a*), but the importance of each base within the motif for She-complex recognition appeared to vary. The 5′ A_{1291}CG triplet was highly overrepresented in the selected clones, whereas a lesser bias toward adenosines at the 3′ end was detected (Fig. 5*a*; see also Table 4, which is published as supporting information on the PNAS web site). In support of these observations, mutation of the guanosine (G1293C) in the context of a selected E2Bmin clone (A_{1291}CGUUUU → ACCUUUU) decreased activity 10-fold (data not shown). The motif randomization was repeated in zipcode YLR434-2, and although similar results were obtained, the strength of the base preferences varied at some positions (Fig. 11 and Table 5, which are published as supporting information on the PNAS web site). Surprisingly, the strength of the bias for C_{1292}G varied even between the two overlapping E2B experiments (Fig. 5*b*; see also Table 6, which is published as supporting information on the PNAS web site), indicating that the requirements for She-complex binding are influenced by the variability of the surrounding region. We noticed that most selected sequences were predicted to form the same secondary structure as WT E2Bmin. Although the observed sequence biases may have resulted from structural constraints, the recovered clones represented only a small fraction of sequences predicted to form the same structure as the natural zipcode (data not shown), suggesting that secondary structure alone cannot mediate She2/3p recognition.

In addition to the overrepresentation of bases in the recognition motif, we also detected a bias toward the adenosine at the 5′ end of the loop (A_{1287}) and a stronger requirement for the C_{1289}G dinucleotide upstream of the recognition motif (Fig. 5*b* and Table 6). Olivier *et al.* (12) recently reported that the C_{1289}GA triplet was essential for She2p binding; our results supported the importance of these bases and the downstream C_{1292}G. Taken together, our results show that a repeated CG dinucleotide promotes She-complex binding: The consensus sequence, by base frequency, of positions 1289–1293 of E2Bmin was CGACG, and CGACGA was most frequently selected in the context of YLR434-2. However, the CG dinucleotide followed by adenosines occurs most frequently in natural zipcodes, and this pattern is sufficient for bud localization.

The sequence and structural requirements in the stem of E2Bmin were also analyzed by randomization and selection. The bias toward base-pairing was strongest at the second position from the top of the loop, whereas the base of the stem was paired only somewhat more often than was expected at random (Fig. 5*c*). Although targeted mutatgenesis had revealed weak sequence preferences in the two loop-proximal base pairs, no biases were observed by randomization/selection (Fig. 5*c*; see also Table 7, which is published as supporting information on the PNAS web site), possibly because 3-AT selection does not discriminate between modest differences in three-hybrid activity (24). Surprisingly, we recovered a bias toward the C_{1283}C dinucleotide in the 5′ strand of the stem and a weaker bias for G_{1300} (Fig. 5*c*; see also Table 8, which is published as supporting information on the PNAS web site). The bias toward this guanosine likely results from the need to base pair with C_{1284}. These results further support our conclusion that stems can contribute both sequence and structural information for She-complex recognition.

Sequence requirements in the 3′ tail were also revealed. Although Olivier *et al.* (12) reported that C_{1302} was essential for She2 binding, only a modest bias toward this cytosine was detected (Fig. 5*c* and Table 8). Eleven of 12 clones that contained substitutions at this position had a UC dinucleotide immediately upstream, even though this pattern was not observed in native zipcodes lacking an analogous cytosine. It is apparent that the requirements for She-complex recognition are flexible, and that the cytosine described by Olivier *et al.* is not essential for all zipcodes.

Using the requirements elucidated by the mutational and randomization analyses, we sought to identify zipcodes in other localized RNAs. One candidate zipcode (bases 798–839 of *MID2*), which contains a single-stranded ACGAAAU motif adjacent to a stem and an adenosine six bases upstream, was localized above background levels (in 65–70% of budded cells) but less efficiently than other zipcodes isolated by three-hybrid assay. Candidate zipcodes in *IST2* and *BRO1*, however, failed to be localized above background levels (data not shown). WSC2C was the only isolated zipcode that did not contain the recognition motif in a single-stranded region and required two stem-loops for WT activity (Fig.

12, which is published as supporting information on the PNAS web site). These results suggest that RNA recognition by She2/3 is complex and that current knowledge of binding requirements and/or prediction tools is insufficient for accurately identifying new zipcodes.

Discussion

We have used a high-throughput selection for mapping She-complex binding sites in RNA targets. This methodology uses NRR to prepare DNA-encoding localized RNAs, followed by three-hybrid selection to identify small fragments containing binding sites. Unlike other *in vitro* evolution techniques, NRR does not alter WT binding sites, making it easier to deconvolute the sequences after selection. Next, NRR covers sequence space efficiently because every starting pool contains a She2/3-binding site, eliminating the need to sample every nucleotide at every position and, thus, generating positive results from low-complexity libraries. Unlike conventional deletion mapping approaches, NRR samples all orientations and connectivities of input sequences.

By subjecting the NRR-derived pool to an *in vivo* three-hybrid selection, we could recover potentially lower-affinity and lower-abundance library members that may be missed by *in vitro* SELEX-style selection or candidate mutagenesis approaches. At the same time, the three-hybrid selection resulted in a low rate of false positives, because higher-abundance library members did not have a significant selective advantage. Finally, the *in vivo* selection ensured that the She proteins retained any posttranslational modifications that may be necessary for WT activity.

Complex Sequence and Structural Features Mediate She2/3 Binding. Initial analysis of the NRR-derived zipcodes revealed a conserved single-stranded, seven-base motif lying proximal to a duplex region. Targeted mutagenesis confirmed that the motif sequence was necessary in different zipcodes for RNA transport and for direct binding to She2/3. The structural context of the motif was also important for She-complex recognition: Positioning the motif in a duplex abolished activity, and increasing the distance between the motif and adjacent stem decreased activity in three of four zipcodes. A simple sequence motif stabilized by surrounding secondary structure appears to be a common theme of many protein binding sites in mRNAs, e.g., the Smg binding site in *nos* RNA (25). The She2/3 recognition site defined in this work expands on the CGA triplet reported by Olivier *et al.* (12) by virtue of a larger set of zipcodes that allowed us to identify the more degenerate bases downstream of the triplet as part of the recognition site. An additional single-stranded cytosine defined by Olivier *et al.* does not appear to be essential for She-complex recognition, because several natural zipcodes do not contain this nucleotide.

Quantitative analysis (by randomization/selection) of the nucleotide requirements for She-complex binding contributed to a more thorough description of the RNA–protein interaction. Nucleotides at the 5′ end of the motif, particularly a CG dinucleotide, were most important for binding, whereas the 3′ adenosines made a weaker contribution. All natural zipcodes contained an adenosine after the CG dinucleotide, and this base was strongly favored in two of three randomization experiments, suggesting that it too plays a major role in binding. Bases outside of the conserved motif also facilitated She-complex binding: Some bases in the stem and 3′ tail of E2Bmin were overrepresented in the selected clones, even though these sequences were not present in other zipcodes and one zipcode (YLR434-2) did not contain essential stem sequences.

The randomization/selection experiments revealed an unexpected plasticity in the sequence requirements for She-complex recognition. When the four adenosines at the 3′ end of the E2Bmin motif were held constant, there was only a weak bias for the upstream CG dinucleotide; but when these adenosines were allowed to vary, the CG dinucleotide was strongly required, suggesting that some motif bases can bypass the requirement for others. Surprisingly, the two CG dinucleotides in E2Bmin do not function redundantly, because the requirement for the downstream CG was strongest when the upstream CG was invariable. A second example of sequence flexibility is that a UC dinucleotide can suppress mutations of a downstream cytosine identified by Olivier *et al.* (12) as essential for She2p binding. Some of these context-dependent effects may result from the RNA adopting a suboptimal fold upon binding She2/3. The extensive sequence and structural plasticity, however, suggests that the She complex recognizes a precise 3D structure in its target RNAs: The complex may bind specifically to the key CG dinucleotide, with the surrounding bases simply maintaining the required structure.

One goal of defining a minimal RNA motif is to generate a predictive model whereby zipcodes could be identified in other RNAs *in silico*. We found that the core motif appears in She2/3 targets and other RNAs known not to be localized, confirming that the motif alone does not confer specificity to the RNA–protein interaction. When the motif and other accessory features (e.g., an upstream adenosine and/or a cytosine 6 nt away from the motif) were used to identify new zipcodes, many localized RNAs did not contain any sequences that fit these criteria. From our analyses of known zipcodes, we conclude that RNA recognition likely involves complex structural features that cannot be appreciated with current tools of searching linear sequences and prediction of secondary structures. Thus, accurate prediction of zipcodes in other localized RNAs awaits a 3D structure of the She complex bound to a target RNA and methods for predicting this structural fold in other RNAs. Meanwhile, the combination of NRR and three-hybrid selection provides a rapid and accurate way to isolate bona fide localization signals, and additional minimized zipcodes will help to elucidate the range of sequences/structures bound by the She complex.

We thank M. Wickens for providing yeast strains and plasmids for the three-hybrid assay; David Liu, Josh Bittker, and Jane Liu for advice on NRR; Joel Credle for assistance with sequencing; and members of the DeRisi laboratory (University of Wisconsin) for comments on the manuscript. This work was supported by grants from the National Science Foundation (to A.J.), The David and Lucille Packard Foundation (to J.L.D.), the Searle Scholars Program (to P.A.T.), the Jane Coffin Childs Memorial Fund for Medical Research (to K.A.S.), and National Institutes of Health Grant 38499 (to R.D.V.).

1. Oleynikov, Y. & Singer, R. H. (1998) *Trends Cell Biol.* **8,** 381–383.
2. Betley, J. N., Frith, M. C., Graber, J. H., Choo, S. & Deshler, J. O. (2002) *Curr. Biol.* **12,** 1756–1761.
3. Kim-Ha, J., Webster, P. J., Smith, J. L. & Macdonald, P. M. (1993) *Development (Cambridge, U.K.)* **119,** 169–178.
4. Macdonald, P. M. & Kerr, K. (1998) *Mol. Cell Biol.* **18,** 3788–3795.
5. Gautreau, D., Cote, C. A. & Mowry, K. L. (1997) *Development (Cambridge, U.K.)* **124,** 5013–5020.
6. Shepard, K. A., Gerber, A. P., Jambhekar, A., Takizawa, P. A., Brown, P. O., Herschlag, D., DeRisi, J. L. & Vale, R. D. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 11429–11434.
7. Bohl, F., Kruse, C., Frank, A., Ferring, D. & Jansen, R. P. (2000) *EMBO J.* **19,** 5514–5524.
8. Takizawa, P. A. & Vale, R. D. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 5273–5278.
9. Long, R. M., Gu, W., Lorimer, E., Singer, R. H. & Chartrand, P. (2000) *EMBO J.* **19,** 6592–6601.
10. Gonzalez, I., Buonomo, S. B., Nasmyth, K. & von Ahsen, U. (1999) *Curr. Biol.* **9,** 337–340.
11. Chartrand, P., Meng, X. H., Singer, R. H. & Long, R. M. (1999) *Curr. Biol.* **9,** 333–336.
12. Olivier, C., Poirier, G., Gendron, P., Boisgontier, A., Major, F. & Chartrand, P. (2005) *Mol. Cell Biol.* **25,** 4752–4766.
13. Bittker, J. A., Le, B. V. & Liu, D. R. (2002) *Nat. Biotechnol.* **20,** 1024–1029.
14. Bernstein, D. S., Buter, N., Stumpf, C. & Wickens, M. (2002) *Methods* **26,** 123–141.
15. SenGupta, D. J., Zhang, B., Kraemer, B., Pochart, P., Fields, S. & Wickens, M. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 8496–8501.
16. Longtine, M. S., McKenzie, A., 3rd, Demarini, D. J., Shah, N. G., Wach, A., Brachat, A., Philippsen, P. & Pringle, J. R. (1998) *Yeast* **14,** 953–961.
17. Gietz, R. D. & Woods, R. A. (2002) *Methods Enzymol.* **350,** 87–96.
18. Mathews, D. H., Sabina, J., Zuker, M. & Turner, D. H. (1999) *J. Mol. Biol.* **288,** 911–940.
19. Zuker, M. (2003) *Nucleic Acids Res.* **31,** 3406–3415.
20. Bailey, T. L. & Elkan, C. (1994) *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2,** 28–36.
21. Kruse, C., Jaedicke, A., Beaudouin, J., Bohl, F., Ferring, D., Guttler, T., Ellenberg, J. & Jansen, R. P. (2002) *J. Cell Biol.* **159,** 971–982.
22. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. (2004) *Genome Res.* **14,** 1188–1190.
23. Schneider, T. D. & Stephens, R. M. (1990) *Nucleic Acids Res.* **18,** 6097–6100.
24. Hook, B., Bernstein, D., Zhang, B. & Wickens, M. (2005) *RNA* **11,** 227–233.
25. Crucs, S., Chatterjee, S. & Gavis, E. R. (2000) *Mol. Cell* **5,** 457–467.