Cell, Vol. 105, 281–289, April 20, 2001, Copyright ©2001 by Cell Press

Structure of Pumilio Reveals Similarity between RNA and Peptide Binding Motifs

**Thomas A. Edwards,* Scott E. Pyle,†
Robin P. Wharton,† and Aneel K. Aggarwal*‡**
*Structural Biology Program
Department of Physiology and Biophysics
Mount Sinai School of Medicine
Box 1677
1425 Madison Avenue
New York, New York 10029
†Department of Genetics
Howard Hughes Medical Institute
Box 3657
Duke University Medical Center
Durham, North Carolina 27710

Summary

Translation regulation plays an essential role in the differentiation and development of animal cells. One well-studied case is the control of *hunchback* mRNA during early *Drosophila* embryogenesis by the trans-acting factors Pumilio, Nanos, and Brain Tumor. We report here a crystal structure of the critical region of Pumilio, the Puf domain, that organizes a multivalent repression complex on the 3′ untranslated region of *hunchback* mRNA. The structure reveals an extended, rainbow shaped molecule, with tandem helical repeats that bear unexpected resemblance to the armadillo repeats in β-catenin and the HEAT repeats in protein phosphatase 2A. Based on the structure and genetic experiments, we identify putative interaction surfaces for *hunchback* mRNA and the cofactors Nanos and Brain Tumor. This analysis suggests that similar features in helical repeat proteins are used to bind extended peptides and RNA.

Introduction

Translation regulation plays a vital role in the lives of most organisms (Gray and Wickens, 1998). It provides an important checkpoint in the pathways for cell growth and differentiation (Gray and Wickens, 1998; Willis, 1999), and a link to the pathology of several diseases (Conne et al., 2000). The impact of translation is perhaps most evident during early development (Curtis et al., 1995; Macdonald and Smibert, 1996). In early *Drosophila* embryos, for instance, a cascade of translation regulatory events helps to give rise to the protein gradients that organize the body pattern along the anterior–posterior axis (Curtis et al., 1995; Macdonald and Smibert, 1996). This regulation is generally mediated by cis-acting elements in the 3′-untranslated regions (3′UTR) of target mRNAs. One such target is the maternally derived *hunchback* (*hb*) mRNA, which is uniformly distributed initially, but whose translational repression at the posterior leads to a Hb protein gradient with highest concentration at the anterior (Tautz, 1988). Failure of this repression disrupts the formation of abdominal segments.

Trans-acting factors, such as Pumilio (Pum) and Nanos (Nos), that target 3′UTR regulatory elements have now been identified in several organisms. Pum binds to a pair of 32 nucleotide sequence motifs, the so-called Nanos response elements (NREs), within the 3′UTR of *hb* mRNA (Wharton and Struhl, 1991; Murata and Wharton, 1995). Pum is distributed evenly throughout the embryo (Macdonald, 1992), whereas Nos is distributed as a gradient emanating from the posterior due to the regulation of its own mRNA by Smaug and Oskar (Gavis and Lehmann, 1994; Dahanukar et al., 1999; Smibert et al., 1999). Nos is recruited by a combination of weak protein–protein and protein–RNA contacts to the Pum/NRE complex (Sonoda and Wharton, 1999). Thus, in a broad sense, Pum provides the specificity for NRE recognition, while Nos provides the positional information for repression at the posterior. Recently, an additional component of the repression complex was identified, Brain Tumor (Brat), which is a member of the evolutionarily conserved RBCC-NHL class of proteins (Sonoda and Wharton, 2001). Brat is required to regulate *hb* in early embryos, and is recruited to the repression complex via contacts with both Pum and Nos (Sonoda and Wharton, 2001). The mechanism by which the Pum/Nos/Brat/NRE quaternary complex blocks *hb* mRNA translation is unclear, though it is thought to target a component(s) of the polyadenylation and/or the translation machinery (Wharton and Struhl, 1991; Wreden et al., 1997).

To understand how Pum organizes this repression complex, we undertook a structural analysis of the region critical for complex formation. Although Pum is a large protein of ~156 kDa, much of it is unnecessary for regulating *hb* mRNA translation. In a striking finding, expression of just the minimal RNA binding domain (RBD), defined as a 37 kDa fragment close to the C terminus, is sufficient to rescue abdominal segmentation defects in *pum* mutant embryos (Wharton et al., 1998). Thus, the Pum RBD alone appears to contain all of the residues required for regulation of *hb*, including RNA binding and recruitment of Nos and Brat.

Pum is a founder member of a novel class of RNA binding proteins (Zamore et al., 1997; Wharton et al., 1998). The similarity between Pum RBD and that of another translation regulator FBF (Zhang et al., 1997), which binds to the 3′UTR of *fem*-3 mRNA in *C. elegans*, defines a Puf (*Pu*m and *F*BF) domain, which is conserved in organisms as diverse as plants, yeast, and humans. The Puf domain is characterized by eight imperfect repeats of ~36 amino acids (Puf repeats), followed by a C-terminal extension. All eight repeats appear to be required for proper folding of the Puf domain, as limited proteolysis fails to yield stable smaller fragments (unpublished data). The Puf domain is thus amongst the largest sequence-specific RNA binding motifs to be discovered; the RRM (70–90 residues), the KH domain (70 residues), and the dsRBD (65 residues), are much smaller (Nagai, 1996). Only the fly Pum and not the human Pum is capable of recruiting Nos, even though the

‡To whom correspondence should be addressed (e-mail: aggarwal@inka.mssm.edu).

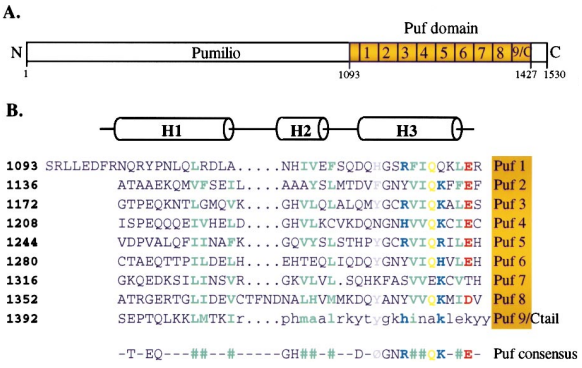

Figure 1. Pum Primary and Secondary Structure

(A) A cartoon to show the relative location of the Puf domain and the individual Puf repeats (highlighted in brown). The full Puf domain contains residues 1093–1427, while the protein in the crystal contains residues 1093–1413.

(B) Structure based alignment of the Puf repeats. The secondary structure elements (helices H1, H2, and H3) are shown above the sequence alignment. Conserved residues are colored green for hydrophobic, red for acidic, blue for basic, and yellow for glutamine. The consensus Puf repeat sequence is shown below the alignment, with the symbol # designating hydrophobic residues and the symbol ϕ marking aromatic residues. Residues 1404–1427 of the Puf domain, which are not defined in the crystal structure, are shown in lower case.

human Pum is capable of binding specifically to the *hb* NRE (Zamore et al. 1997; Sonoda and Wharton, 1999).

We report here the crystal structure of the Pum Puf domain, which encompasses all eight Puf repeats as well as part of the C-terminal extension. We find that the Puf domain falls into a superfamily of α-helical repeat proteins, characterized by armadillo (Arm) repeats in the signaling protein β-catenin (Huber et al., 1997) and the nuclear import protein karyopherin-α (also called importin-α) (Conti et al., 1998) and HEAT repeats in protein phosphatase 2A (pp2A) (Groves et al., 1999). The same features that allow these proteins to bind peptides/proteins within an extended groove appear to provide the basis for Pum mRNA binding. We identify the putative interaction surfaces for Nos and Brat from the crystal structure, supported by mutational analysis of the Pum RBD. Taken together, these structural and genetic studies provide an initial view of the evolutionarily conserved Puf domain and offer a framework for understanding the assembly of an mRNA translation repression complex.

Results and Discussion

Structure Determination

A fragment of Pum, encompassing residues 1092 to 1411 (Figure 1), was expressed, purified, and crystallized as described (Edwards et al., 2000). In brief, hexagonal crystals from ammonium sulfate solutions, containing 2 molecules per asymmetric unit, were used to collect native and multiwavelength anomalous dispersion (MAD) data. The best MAD data were collected from a mercury derivative, which together with the native data were used to compute experimental phases to 3.0 Å resolution (Table 1). These phases were extended to 2.3 Å and a solvent-flattened electron density map calculated at that resolution. The map showed excellent electron density, allowing the construction of a molecular model for both copies of the molecule without a need for noncrystallographic averaging. The current model includes residues 1093–1404 with good stereochemistry (Table 1)

Genetic Analysis

In parallel with the structural work, we undertook a genetic analysis of the Pum RBD to identify residues involved in RNA binding as well as recruitment of Nos and Brat. Each of these activities was assayed in yeast using various protein–RNA and protein–protein interaction techniques. In brief, we randomly mutagenized a gene encoding the Pum RBD in vitro by error prone PCR, introduced the pool of mutant genes into yeast, and identified derivatives that bind to the NRE. These mutants were subsequently screened for the ability to recruit Nos and Brat (Sonoda and Wharton, 1999, 2001).

Overall Architecture

The Puf repeats align tandemly to form an extended, curved molecule (Figure 2). Each repeat is related to the next by a rotation of $\sim$18°, resulting in a rainbow-like arc that covers approximately one third (124°) of the circumference of an $\sim$42 Å radius circle (Figure 3). Each Puf repeat is a trihelical bundle made up of two long α helices (H1 and H3) and one short helix (H2). These α helices align with equivalent helices of neighboring repeats to give three parallel layers that run the length of the Pum arc (Figure 2). The H1 layer covers the outer convex surface, the H2 layer forms the ridge, and the H3 layer coats the inner concave surface. The overall dimensions of the Pum Puf domain are approximately 26 $\times$ 26 $\times$ 90 Å.

The Puf repeats show unexpected similarity to the Arm repeats in β-catenin and karyopherin-α (Huber et al., 1997; Conti et al., 1998). Like a Puf repeat, each Arm repeat contributes 3 α helices (H1, H2, and H3) in shaping a contiguous, nonglobular domain with distinct convex and concave surfaces (Figure 3). However, the Arm repeats align with both rotational and translational components in β-catenin and karyopherin-α, giving rise to superhelical structures in which the groove (or concave surface) winds around the rotation axis (Huber et al., 1997; Conti et al., 1998). In contrast, Pum has a concave surface that lacks the twist observed in the Arm repeat proteins. In overall topology, the Pum arc resembles the PR65/A subunit of pp2A (Figure 3), containing 15 bihelical (H1 and H2) HEAT repeats (Groves et al., 1999). In particular, HEAT repeats 4 to 12 trace a Pum-like arc, though with a somewhat deeper groove.

Puf Repeats

The choice of which three helices constitute a single Puf (or Arm) repeat is somewhat arbitrary. However, for consistency, we have chosen a register in which helix H3 lines the concave surface in both Pum and β-catenin (Figure 3). Overall, the Puf repeats are more uniformly stacked than either Arm or HEAT repeats, with no major discontinuities or kinks in the circular path. This uniformity in packing is reflected in regularity of structure, with rmsds between Puf repeats varying form 0.85 to 1.6 Å. This is significantly narrower than the range of rmsds observed between Arm repeats in β-catenin

Table 1. Data Collection and Refinement Statistics

Data Collection	Native	PHMB[c]			$OsCl_3$
		λ1	λ2	λ3	
Wavelength (Å)	1.1	1.009	1.0084	0.990	1.14
Max. Resolution (Å)	2.6	2.3	3.2	2.85	3.0
Independent Reflections	32,387	49,178	18,498	25,938	20,997
No. of measurements	152,956	437,974	187,620	230,957	237,555
R_{merge}[a,b]	0.066	0.089	0.085	0.086	0.066
	(0.201)	(0.311)	(0.418)	(0.350)	(0.179)
Completeness (%)[a]	93.1 (56.0)	95.8 (73.2)	94.3 (81.1)	94.2 (75.4)	90.4 (62.6)
I/σ[a]	16.7 (2.7)	20.7 (2.4)	15.4 (2.0)	17.9 (2.0)	22.1 (3.0)
No. of sites		10	10	10	4
FoM[d] (centric/acentric) 3.0 Å		0.518/0.379			
FoM[e] (SOLOMON) 2.3 Å		0.740			
Refinement					
Resolution Range (Å)		20–2.3			
R_{cryst}/R_{free}[f]		0.244/0.269			
No. of atoms					
Protein		2584			
Water; Hg		187;10			
rms deviations					
Bonds (Å)		0.008			
Angles (°)		1.33			
Avg B factor (Å^2)		41.7			

[a] Values for outermost shell are given in parentheses.
[b] $R_{merge} = \Sigma |I - <I>|/\Sigma I$, where I is the integrated intensity of a given reflection.
[c] PHMB = *p*-hydroxymercury benzoate.
[d] FoM = Mean figure of merit computed to the 3.0 Å limit for MAD phasing in SHARP.
[e] FoM = Overall mean figure of merit at 2.3 Å after solvent flattening.
[f] $R_{cryst}/R_{free} = \Sigma ||F_o| - |F_c||/\Sigma|F_o|$. R_{free} was calculated using 5% of data excluded from refinement.

(0.75–2.25 Å) and HEAT repeats in pp2A (1.0–2.8 Å). Nonetheless, there is clearly some structural variation between the Puf repeats, because the RMSD between the two Pum monomers in the crystallographic asymmetric unit is only 0.8Å. Visually, the most pronounced deviation from a regular 36 aa Puf repeat structure is a 4 aa insert in the loop between helices H1 and H2 in repeat 8. This extra long loop contains a solvent exposed phenylalanine (F1367) close to the inferred Nos and Brat binding sites (discussed below). Moreover, this region is poorly defined in our electron density map, implying flexibility that may be important for interactions with these cofactors. Helix H1 in repeats 5 and 6 is also slightly distorted, resulting in a small twist in the middle of the molecule. Consequently, the H3 helices from repeats 6–8 are not quite parallel to those from repeats 1–5 (Figure 2).

The conservation of hydrophobic residues at strategic positions across the Puf repeats forms the basis of a contiguous hydrophobic core running through the molecule (Figure 2). At the N terminus, the hydrophobic core is capped by repeat 1 and an additional N-terminal α helix (residues 4 to 12). Three phenylalanine residues, emanating from repeat 1 and the N-terminal helix, coalesce to cap the hydrophobic core. At the C terminus, the important capping residues come from the C-terminal tail, and not repeat 8. Because the protein in the crystals contains only a partial C-terminal tail, the hydrophobic core at the C terminus remains largely uncapped. This may be the reason why Pum forms a "tail to tail" dimer in our crystals, with a hydrophobic core running through the entire S-shaped dimer. Neither we, nor others (Zamore et al., 1999), have found any evidence of Pum dimers in solution, either free or in complex with RNA. Thus, it is unlikely that Pum forms the kind of dimer seen in the crystals in vivo.

The Pum C-terminal tail may fold to form a ninth Puf repeat. The tail sequence contains a pattern of hydrophobic residues which, as seen in the structure, are important for both inter- and intrarepeat contacts (Figure 1). Accordingly, the first 11 residues of the partial C-terminal tail in our crystals fold into an H1-like α helix. However, the ensuing ten residues, which have the potential to form helix H2, are disordered. This suggests a requirement for H3 for proper folding, the residues for which are missing from our protein construct. Because the C-terminal tail sequence is more variable than repeats 1–8, within the same domain as well as across different species, it has not been described as a putative Puf repeat. Based on our structure, the complete Puf domain encompasses residues from the N-terminal capping helix, through repeats 1–8, and to the ninth repeat/C-terminal tail (Figure 1). This is consistent with a deletion analysis that suggests the minimal RNA binding domain extends from P1105 to K1426.

RNA Interaction Surface

The concentration of positive charge along the concave surface suggests that it may be the binding site for mRNA. The positive charge is distributed across most

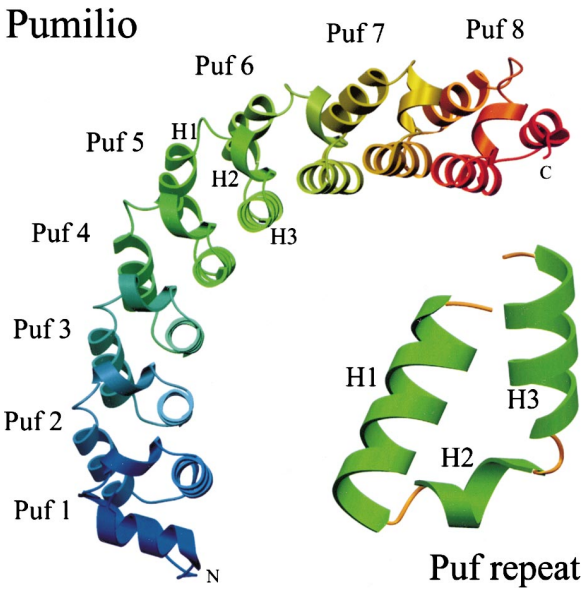

Figure 2. Pum Puf Domain Structure

The Puf domain contains eight tandem Puf repeats (shown in different colors) that together comprise a single contiguous domain. Inset is a magnification of a single repeat. Each repeat is composed of three α helices (H1, H2, and H3). The figure was generated using the program SETOR (Evans, 1993).

of the inner surface and is contributed by conserved lysine and arginine residues on the H3 ladder lining the concave surface (Figure 4A). The H3 ladder also shows the highest sequence conservation (Figure 4B). Intriguingly, β-catenin, karyopherin-α, and pp2A also show the highest conservation of sequence within their inner concave surfaces (Huber et al., 1997; Conti et al., 1998; Groves et al., 1999). The Pum H3 consensus F/Y-X_5-Q-K/R-X_2-E is conserved between Puf repeats within the same domain (Figure 1), as well as across different species (Figure 4). These ladders of conserved residues (Figure 4C) projecting into the solvent are reminiscent of the asparagine and tryptophan arrays on the H3 surface of karyopherin-α (Conti et al., 1998). These tryptophans (in karyopherin-α) form grooves that bind the aliphatic parts of lysine side chains of the nuclear localization signal (NLS) peptide, while the asparagines keep the peptide in an extended conformation by hydrogen bonding to the peptide backbone (Conti et al., 1998). By analogy, it is possible that the glutamine array in Pum is important in maintaining the RNA in an extended conformation, while the aromatics stack with the bases and the basic residues neutralize the sugar-phosphate backbone (Figure 4C). The conservation of glutamates is more mysterious, though it is possible they bind metals and form bridging ionic interactions with the backbone (although no metal dependency has been observed in RNA binding).

Two lines of evidence from mutagenesis studies support the idea that the Pum concave surface binds RNA. First, we randomly mutagenized a gene encoding the 322 residue minimal Pum RBD and isolated variants that bind normally to the wild-type NRE in yeast (Figures 5A and 5B). Collectively, these variants bear substitutions at 61 residues, 55 of which map to the structure (Figure

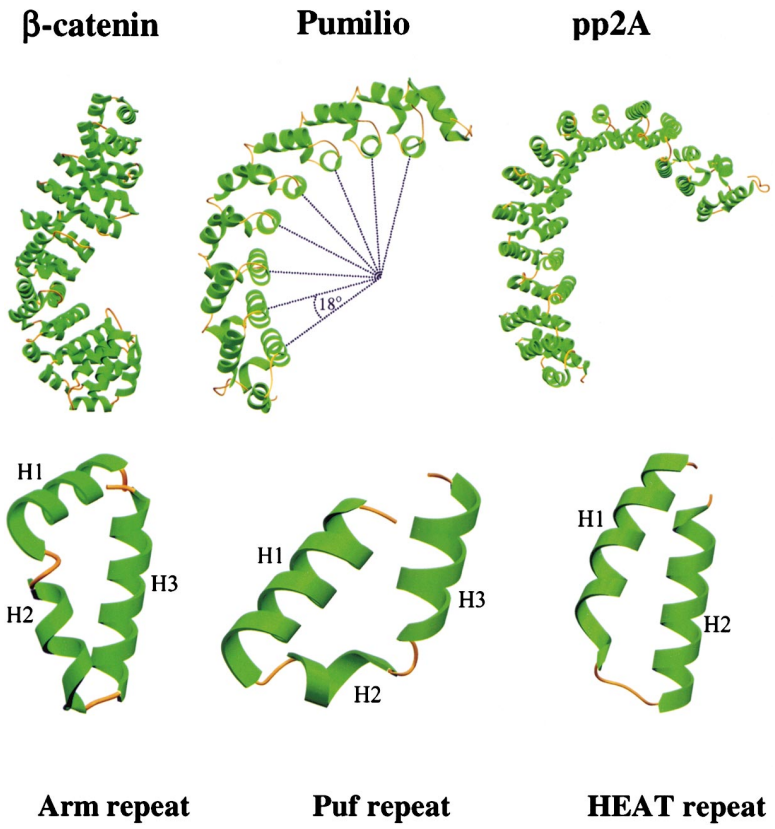

Figure 3. Pum Is a Member of the Helical Repeat Protein Family

Typical members of the family, β-catenin with arm repeats (left) and pp2A with HEAT repeats (right), are shown alongside Pum with Puf repeats (middle). Shown below is a single repeat from each structure, aligned with functionally equivalent helices—H3 for Arm and Puf repeats and H2 for HEAT repeat—in a similar orientation.

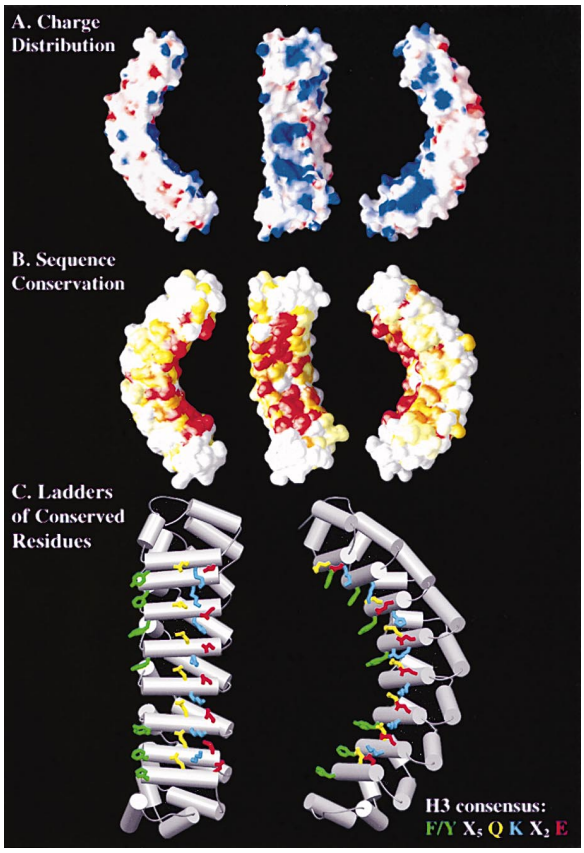

Figure 4. Surface of the Puf Domain

(A) Electrostatic surface. Blue depicts regions of positive potential and red shows regions of negative potential. The three views of the Puf domain are related by successive 90° rotations about the vertical axis. The highest concentration of positive charge occurs on the concave surface. The N terminus of the molecule is toward the bottom of the figure.

(B) Conservation of sequence across species. The color ranges from 0% (white) to 100% (red) sequence identity, based on the alignment of fly, human, *C. elegans*, *S. cerevisiae*, and *S. Pombe* Puf domain sequences. The Puf domain is oriented as in (A). The highest conservation of sequence occurs at the concave surface. The figure was generated using GRASP (Nicholls et al., 1991).

(C) Ladders of conserved residues. The Puf domain is shown in two orientations related by a 90° rotation about the vertical axis. Highlighted are the conserved, solvent exposed residues on the H3 helices, lining the concave surface. This repetitive arrangement is similar to the asparagine and tryptophan arrays seen in the H3 lined groove of karyopherin-α.

5C); the remaining six are in the putative 9th repeat, not in this crystal structure. Of these, only 3 (presumably silent) substitutions fall on the solvent exposed concave surface, with the remaining 52 lying elsewhere (Figure 5D). The relative paucity of substitutions within the inner surface is consistent with this being the area that contacts the RNA (Figure 6). Second, based on the structure, we introduced single substitutions in solvent-exposed residues along the inner surface in five of the eight Puf domains and tested RNA binding activity in yeast (Figures 5B and 5E). Each of these mutants is inactive. Thus, the concentration of positive charge and the distribution of both silent and inactivating substitutions together suggest that the RNA interacts with the inner concave surface.

We propose that *hb* mRNA binds to this inner surface in an extended single-stranded conformation. Algorithms that predict RNA structure suggest the NRE does not adopt a stable secondary or tertiary structure. The minimal NRE for high affinity Pum binding consists of nucleotides 3–27, which bracket specific contacts with nucleotides 9, 11–13, and 21–24 (Murata and Wharton, 1995; Wharton et al., 1998; Zamore et al., 1997). The length of this minimal NRE, in an extended single-stranded conformation (112 Å), agrees approximately with the contour length (90 Å) of the concave surface of the Puf domain (Figure 2). It is noteworthy that β-catenin also has the highest concentration of positive charge within its concave surface (or groove), which is the proposed binding site for segments of cadherins, APC, and members of the LEF-1/TCF family of transcription factors (Huber et al., 1997; von Kries et al., 2000). A recent crystal structure of a β-catenin/TCF complex shows the TCF segment tethered along the positively charged groove (Graham et al., 2000). In the case of karyopherin-α, the concave surface is the binding site for the NLS peptide (Conti et al., 1998). Taken together, the binding of ligands to concave surfaces is a recurring theme in helical repeat proteins. The Pum Puf domain shows that this type of extended surface can be used to bind RNA, as well as peptides.

Interactions with Nos and Brat

Repression of *hb* mRNA depends not only on Pum, but also on the recruitment of Nos and Brat to form a quaternary complex (Sonoda and Wharton, 1999, 2001). Previous work suggested that Nos is recruited via residues in Puf repeat 8 (Sonoda and Wharton, 1999). These residues map to the extra long loop between helices H1 and H2 in repeat 8, that is the main protrusion from an otherwise relatively smooth outer Pum surface (Figure 2). Two different insertions into this loop have no effect on Pum–RNA binding but eliminate recruitment of Nos (Sonoda and Wharton, 1999). To further define the Nos interaction surface, we tested the collection of Pum mutants that bind normally to RNA (described above) for Nos recruitment in yeast (Figure 5). Of the 61 substitutions distributed throughout the domain, only two abrogate interaction with Nos. One is a substitution in the putative ninth Puf repeat that is not represented in our structure, while the other changes the solvent exposed phenylalanine on the H1/H2 loop to a serine (F1367S) (Figures 5B, 5C, and 5E). Thus, the Pum surface that interacts with Nos appears to be limited to a small region that includes the eighth repeat and the C-terminal tail (Figures 5E and 6). If this tail indeed does fold into a ninth Puf repeat as discussed above, then the Pum-Nos interface would span a length of ~15–20 Å on the outer convex surface. It is tempting to think that the C-terminal tail may only fold when Pum binds to the RNA, thereby explaining why Nos is only recruited to the Pum/NRE binary complex and not to Pum alone (Sonoda and Wharton, 1999). The insertions into the long flexible loop in repeat 8 may modify its conformation such that F1367 is no longer exposed for interaction with Nos. The proposed Phe-Nos interaction is reminiscent of the way in

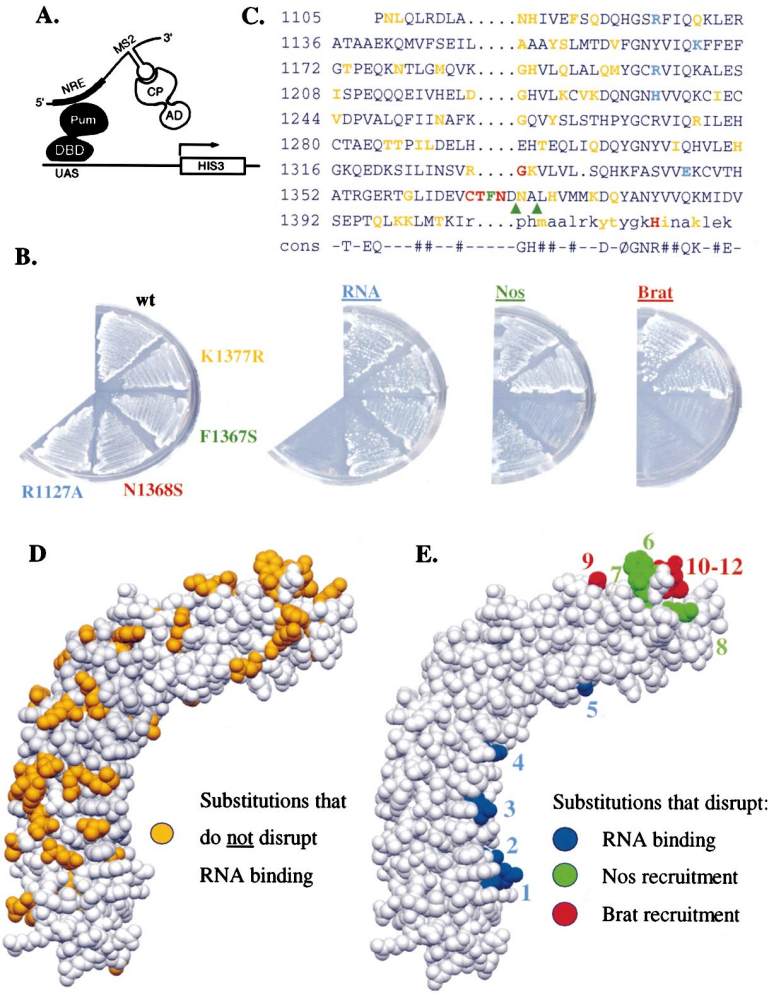

Figure 5. Analysis of the Pum/NRE/Nos/Brat Interaction Surfaces

(A) A cartoon to show the yeast interaction assay used to assess RNA interaction. Binding of Pum to the NRE allows growth in the absence of histidine.

(B) Yeast interaction assays showing the activities of the wild-type (wt) Puf domain, and 4 singly substituted mutants corresponding to 4 different phenotypes: K1377R is an example of a mutant that binds RNA, Nos, and Brat normally. F1367S is an example of a mutant that binds RNA normally but does not recruit Nos. N1368S is an example of a mutant that binds RNA and Nos normally but does not recruit Brat. R1127A is an example of a substitution within the concave surface, engineered on the basis of structure, that abolishes RNA binding, and as a consequence, Nos recruitment.

(C) Mapping of mutations onto the primary and secondary structure. In orange are residues where randomly generated substitutions do not affect interaction with RNA, Nos, or Brat. In blue are residues where substitutions engineered on the basis of structure disrupt RNA binding. In green are residues where substitutions disrupt Nos binding. Nos binding is also disrupted by two insertions (triangles). In red are residues at which substitutions disrupt Brat recruitment.

(D) Mapping onto three-dimensional structure the mutations that do not affect RNA binding (orange). Pum is drawn in the same orientation as Figure 2. The relative paucity of substitutions within the concave surface is consistent with this being the area that contacts *hb* mRNA.

(E) Mapping onto three dimensional structure the mutations and insertions that disrupt RNA (blue), Nos (green), or Brat (red) binding. The highlighted substitutions are: 1, R1127A; 2, K1167A; 3, R1199A; 4, H1235A; 5, E1346K; 6, F1367S; 7, GPH insert at 1369; 8, QICA insert at 1372; 9, G1330D; 10, C1365R; 11, T1366D; and 12, N1368S.

which a solvent exposed phenylalanine on the receptor CD4 interacts with the HIV gp120 glycoprotein (Kwong et al., 1998; Wang et al., 1990; Ryu et al., 1990).

The surface that interacts with Brat appears to be limited to repeats 7, 8, and 9, based on analysis of the collection of Pum mutants that bind normally both to the NRE and to Nos (Figure 5). Five single mutants and one double mutant bearing substitutions in this region of the protein do not interact with Brat. The mutations in repeats 7 and 8 map to the loops, between helices H1 and H2, that are exposed on the convex surface (Figure 5E), consistent with our earlier studies. The Brat binding site is localized immediately adjacent to the Nos binding site on the outer Pum surface (Figure 6), raising the possibility of cooperative interactions between the two cofactors. The close proximity of the sites may explain why Brat is only recruited once Nos has joined the Pum/NRE complex (Sonoda and Wharton, 2001).

The Pum/Nos partnership extends beyond the regulation of *hb* mRNA to the correct development of the germline. In addition to *hb* mRNA, Pum and Nos jointly repress translation of maternal *cyclinB* mRNA in the germline precursor cells (Asaoka-Taguchi et al., 1999). Although the sequences required for this regulation have not yet been defined, Pum binds to an element in the *cyclinB* 3′UTR that is similar in sequence to the NRE (Dalby and Glover, 1993; Sonoda and Wharton, 2001). While the Pum/*cyclinB* RNA complex can recruit Nos, the resulting ternary complex does not bind Brat efficiently. This suggests a structural difference between Pum/Nos bound to the *hb* NRE versus the *cyclinB* RNA, allowing the former to recruit Brat and the latter to recruit a different cofactor present in the germ line. It is noteworthy that much of the Pum outer surface is "empty" or "unspecified" (Figure 6), and it may be this portion of the molecule that interacts differently with Nos (and other cofactors) when bound to *cyclin B* mRNA. To understand the basis of this geometric difference will require cocrystallization of Pum/Nos with different RNA sequences. While the allosteric effects of closely related DNA sites on the conformation of transcription factors are well documented (Lefstin and Yamamoto, 1998; Scully et al., 2000), this issue is largely unexplored with RNA binding proteins.

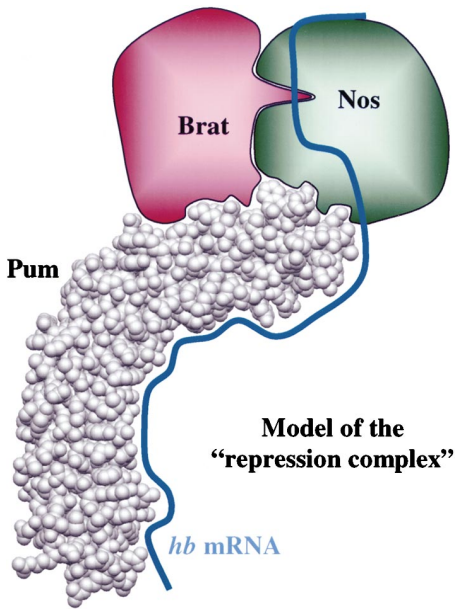

Figure 6. A Hypothetical Model of the Pum/NRE/Nos/Brat Repression Complex

The 3′UTR of *hb* mRNA (NRE) is postulated to bind the inner concave surface of Pum, while the cofactors Nos and Brat are hypothesized to bind the outer convex surface. The close proximity of Nos and Brat sites (c.f., Figure 5E) raises the possibility of cooperative interactions between the two cofactors.

Pum Is an Unexpected Member of the Helical Repeat Family

Pum joins a family of helical repeat proteins that includes β-catenin and karyopherin-α with Arm repeats (Huber et al., 1997; Conti et al., 1998), pp2A with HEAT repeats (Groves et al., 1999), karyopherin-β (also called importin-β) with a mixture of HEAT and Arm repeats (Chook and Blobel, 1999; Cingolani et al., 1999), and protein phosphatase 5 with tetratricopeptide repeats (Das et al., 1998). A broader definition of the family would include proteins with a repeating α/β substructure, such as ribonuclease inhibitor with leucine-rich-repeats (Kobe and Deisenhofer, 1993) and IκB with ankyrin repeats (Huxford et al., 1998; Jacobs and Harrison, 1998). All of these family members are characterized by an extended surface that until now had been thought to be ideally suited for protein–protein interactions. The Pum structure shows that the same kind of surface can also be used to recognize RNA. It is curious that several members of the family, including β-catenin, karyopherin-α, and karyopherin-β, (and I-κB) are involved in movements in and out the cell nucleus. It is tempting to speculate that Puf domains may have roles beyond regulation of mRNA translation (Tadauchi et al. 2001) and degradation (Olivas and Parker, 2000), perhaps involving the trafficking of RNA out of the nucleus in some species.

Experimental Procedures

Protein Preparation and Crystallization

The *Drosophila* Puf domain (residues 1092–1411 of Pumilio) was expressed in *E. coli* and purified as described in Edwards et al. (2000). Crystals were obtained by hanging drop vapor diffusion against 2.1 M ammonium sulfate, 100 mM HEPES (pH 6.8), 4% DMSO, and 2 mM DTT at 4°C. The crystals grow to maximum size after about one week in space group $P6_3$ (a = b = 94.5Å, c = 228.2Å, α = β = 90°, and γ = 120°) with 2 molecules per asymmetric unit.

Data Collection, Structure Determination, and Refinement

Native and MAD data were collected (110 K) at beamline X25 of the National Synchrotron Light Source at Brookhaven National Laboratory (BNL). Crystals were frozen in liquid nitrogen after transfer through solutions containing the mother liquor plus increasing concentrations of MPD to a final concentration of 20%. Soaking the crystals in 1 mM solutions of $OsCl_3$ or *p*-hydroxymerury benzoate respectively, during cryoprotection, produced osmium and mercury derivatives. X-ray fluorescence scans were obtained for both derivatives, in order to determine the wavelengths around the osmium and mercury L-III absorption edge profiles for MAD data collection. All data were indexed and integrated using DENZO and reduced using SCALEPACK (Otwinowski and Minor, 1997).

Anomalous and isomorphous difference Patterson maps calculated with the PHASES package (Furey and Swaminathan, 1997) showed strong peaks corresponding to 2 Os sites. The program SOLVE (Terwilliger and Berendzen, 1999) found the same 2 sites plus another 2 weak sites. These sites were refined with MLPHARE (Otwinowski, 1991) and the phases used to find mercury positions by difference Fourier analysis. A total of 9 mercury sites were identified from the difference Fourier maps. Because the crystals tend to decay in the X-ray beam, the resolution limit for each wavelength (collected in the order λ1, λ3, and λ2) is different. Thus, useful data extend to 2.3 Å for λ1, but are limited to 3.3 Å for λ2. The MAD analysis was performed with the program SHARP (de La Fortelle and Bricogne, 1997), using the native data and the three wavelength mercury MAD data (Table 1). (For reasons that are unclear, the inclusion of osmium MAD data did not improve phasing.) The SHARP phases were extended to the maximum Bragg spacing of 2.3Å (corresponding to λ1 data—Table 1) with density modification using SOLOMON (CCP4, 1994). This yielded an experimental electron density map that was readily interpretable without the need for noncrystallographic averaging. The model of both molecules was built into this map.

Structure refinement was carried out with a 96% complete data set that had an overall R_{merge} = 0.089 at 20–2.3Å resolution. After an initial rigid body refinement, the crystallographic R factor was 44.2% (R_{free} 44.8%). The R factor dropped to 31.4% (R_{free} 41.6%) after a round of simulated annealing. Iterative rounds of model building and positional and B factor refinement were carried out, using the programs O (Jones et al., 1991) and CNS (Brunger et al., 1998). The waters were added after the R_{free} dropped to below 32%. The final model consists of residues 1093–1404 and 187 water molecules, with an R factor of 24.4% and an R_{free} of 26.9% (Table 1). The program PROCHECK (Laskowski et al., 1993) revealed only 11 residues in unfavorable (ϕ, ψ) regions with 98.1% of residues in favorable and allowed regions.

Mutagenesis

A gene encoding the minimal Pum RBD was mutagenized by error prone PCR essentially as described by Vidal et. al. (1996) Yeast strain PJ69-4A was transformed with pM665, which encodes a chimeric RNA containing tandem NREs and binding sites for MS2 CP, pJ2531 that encodes a fusion of CP to the GAL4 transcriptional activation domain (AD), a gapped plasmid that encodes the GAL4 DNA binding domain, and the PCR product. Yeast colonies that encode the three components of Figure 5A (the DBD–PumRBD plasmid reconstituted by gap repair in vivo) were isolated on Trp^- Leu^- Ura^- medium, and NRE binding was assayed by streaking on medium that also lacked His and contained 3-amino triazole (Sonoda and Wharton, 1999). Approximately 6% of the transformants proved to harbor *pum* genes with functional RNA binding domains. Forty-four *pum* mutant candidates were sequenced; 12 encode wild-type proteins and 32 encode proteins with between 1 and 6 substitutions (average = 2.0). Recruitment of Nos into a ternary complex was assayed by retransformation with pJ2486, which encodes a Nos-AD fusion, in place of pJ2531 (Sonoda and Wharton, 1999). Recruitment of Brat was assayed by transferring each mutant *pum* gene

into the four-hybrid vector described by Sonoda and Wharton (2001). The collection of mutants shown in Figure 5 includes the substitutions described by Sonoda and Wharton (2001). Site-directed mutations were prepared by standard PCR methods using mismatched primers; the expression in yeast of each inactive mutant was verified by Western blot.

Acknowledgments

We are grateful to Lonnie Berman, Hal Lewis, and Michael Becker for facilitating X-ray data collection at NSLS. We thank Larry Shapiro, Jun Sonoda, and Tom Burke for advice, Millie McAdams and Judy Phelps for sequencing, and José Trincão for help in preparing figures. We also thank Larry Shapiro for comments on the manuscript. A. K. A is supported by a grant from the NIH (R01 GM62947), R. P. W. is an Assistant Investigator of the HHMI.

Received January 8, 2001; revised March 16, 2001.

References

Asaoka-Taguchi, M., Yamada, M., Nakamura, A., Hanyu, K., and Kobayashi, S. (1999). Maternal Pumilio acts together with Nanos in germline development in Drosophila embryos. Nat. Cell Biol. *1*, 431–437.

Brunger, A.T., Adams, P.D., Clore, G.M., Delano, W.L., Gros, P., Grosse-Kunstleve, R., Jiang, W., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography & NMR system: a software suite for macromolecular structure determination. Acta Crystallogr. *D54*, 905.

CCP4 (1994). Collaborative computing project no. *4*, The CCP4 suite: programs for protein crystallography. Acta Cryst D50, 760–763.

Chook, Y.M., and Blobel, G. (1999). Structure of the nuclear transport complex karyopherin-β2-Ran x GppNHp. Nature *399*, 230–237.

Cingolani, G., Petosa, C., and Muller, C.W. (1999). Structure of importin-β bound to the IBB domain of importin-α. Nature *399*, 221–229.

Conne, B., Stutz, A., and Vassalli, J.-D. (2000). The 3′ untranslated region of messenger RNA: a molecular 'hotspot' for pathology. Nat. Med. *6*, 637–641.

Conti, E., Uy, M., Leighton, L., Blobel, G., and Kuriyan, J. (1998). Crystallographic analysis of the recognition of a nuclear localisation signal by the nuclear import factor karyopherin-a. Cell *94*, 193–204.

Curtis, D., Lehmann, R., and Zamore, P.D. (1995). Translational regulation in development. Cell *81*, 171–178.

Dahanukar, A., Walker, J.A., and Wharton, R.P. (1999). Smaug, a novel RNA-binding protein that operates a translational switch in Drosophila. Mol. Cell *4*, 209–218.

Dalby, B., and Glover, D.M. (1993). Discrete sequence elements control posterior pole accumulation and translational repression of maternal cyclin B RNA in Drosophila. EMBO J. *12*, 1219–1227.

Das, A.K., Cohen, P.W., and Barford, D. (1998). The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. EMBO J. *17*, 1192–1199.

de La Fortelle, E., and Bricogne, G. (1997). Maximum-likelihood heavy atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. Methods Enzymol. *276*, 472–494.

Edwards, T.A., Trincao, J., Escalante, C.R., Wharton, R.P., and Aggarwal, A.K. (2000). Crystallization and characterization of Pumilio: a novel RNA binding protein. J. Struct. Biol. *132*, 251–254.

Evans, S.V. (1993). SETOR: hardware lighted three dimensional solid model representations of macromolecules. J. Mol. Graphics *11*, 134–138.

Furey, W., and Swaminathan, S. (1997). PHASES-95: a program package for the processing and analysis of diffraction data from macromolecules. Methods Enzymol. *277*, 590–629.

Gavis, E.R., and Lehmann, R. (1994). Translational regulation of nanos by RNA localization. Nature *369*, 315–318.

Graham, T.A., Weaver, C., Mao, F., Kimelman, D., and Xu, W. (2000). Crystal structure of a b-catenin/Tcf complex. Cell *103*, 885–896.

Gray, N.K., and Wickens, M. (1998). Control of translation initiation in animals. Annu. Rev. Cell. Dev. Biol. *14*, 399–458.

Groves, M.R., Hanlon, N., Turowski, P., Hemmings, B.A., and Barford, D. (1999). The structure of a protein phosphatase 2a PR65/A subunit reveals the conformation of its 15 tandemly repeated HEAT motifs. Cell *96*, 99–110.

Huber, A.H., Nelson, W.J., and Weis, W.I. (1997). Three dimensional structure of the armadillo repeat region of b-catenin. Cell *90*, 871–882.

Huxford, T., Huang, D.B., Malek, S., and Ghosh, G. (1998). The crystal structure of the IκBα/NF-κB complex reveals mechanisms of NF-kappaB inactivation. Cell *95*, 759–770.

Jacobs, M.D., and Harrison, S.C. (1998). Structure of an IκBα/NF-κB complex . Cell *95*, 749–758.

Jones, T.A., Zou, J.-Y., and Cowan, S.W. (1991). Improved methods for building models in electron density maps and the location of errors in these models. Acta Crystallogr. A*47*, 110–119.

Kobe, B., and Deisenhofer, J. (1993). Crystal structure of porcine ribonuclease inhibitor, a protein with leucine-rich repeats. Nature *366*, 751–756.

Kwong, P.D., Wyatt, R., Robinson, J., Sweet, R.W., Sodroski, J., and Hendrickson, W.A. (1998). Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature *393*, 648–659.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. *A47*, 110–119.

Lefstin, J.A., and Yamamoto, K.R. (1998). Allosteric effects of DNA on transcriptional regulators. Nature *392*, 885–888.

Macdonald, P.M. (1992). The Drosophila pumilio gene: an unusually long transcription unit and an unusual protein. Development *114*, 221–232.

Macdonald, P.M., and Smibert, C.A. (1996). Translational regulation of maternal mRNAs. Curr. Opin. Genet. Dev. *6*, 403–407.

Murata, Y., and Wharton, R.P. (1995). Binding of pumilio to maternal hunchback mRNA is required for posterior patterning in Drosophila embryos. Cell *80*, 747–756.

Nagai, K. (1996). RNA-protein complexes. Curr. Opin. Struct. Biol. *6*, 53–61.

Nicholls, A., Sharp, K.A., and Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins *11*, 281–296.

Olivas, W., and Parker, R. (2000). The puf3 protein is a transcript-specific regulator of mRNA degradation in yeast. EMBO J. *19*, 6602–6611.

Otwinowski, Z. (1991). Maximum likelihood refinement of heavy atom parameters. In Isomorphous Replacement and Anomalous Scattering, W. Wolf, P.R. Evans and A.G.W. Leslie, eds. (Warrington, UK: Science and Engineering Research Council, Daresbury Laboratory).

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.

Ryu, S.E., Kwong, P.D., Truneh, A., Porter, T.G., Arthos, J., Rosenberg, M., Dai, X.P., Xuong, N.H., Axel, R., Sweet, R.W., et al. (1990). Crystal structure of an HIV-binding recombinant fragment of human CD4. Nature *348*, 419–426.

Scully, K.M., Jacobson, E.M., Jepsen, K., Lunyak, V., Viadiu, H., Carriere, C., Rose, D.W., Hooshmand, F., Aggarwal, A.K., and Rosenfeld, M.G. (2000). Allosteric effects of pit-1 DNA sites on long-term repression in cell type specification. Science *290*, 1127–1131.

Smibert, C.A., Lie, Y.S., Shillinglaw, W., Henzel, W.J., and Macdonald, P.M. (1999). Smaug, a novel and conserved protein, contributes to repression of nanos mRNA translation in vitro. RNA *5*, 1535–1547.

Sonoda, J., and Wharton, R.P. (1999). Recruitment of Nanos to hunchback mRNA by Pumilio. Genes Dev. *13*, 2704–2712.

Sonoda, J., and Wharton, R.P. (2001). Drosophila Brain Tumor is a translational repressor. Genes Dev. *15*, 762–773

Tadauchi, T., Matsumoto, K., Herskowitz, I., and Irie, K. (2001). Post-transcriptional regulation through the HO 3′UTR by Mpt5, a yeast homolog of Pumilio and FBF. EMBO J. *20*, 552–561.

Tautz, D. (1988). Regulation of the Drosophila segmentation gene hunchback by two maternal morphogenetic centres. Nature *332*, 281–284.

Terwilliger, T.C., and Berendzen, J. (1999). Automated MAD and MIR structure solution. Acta Crystallogr. D*55*, 849–861.

Vidal, M., Braun, P., Chen, E., Boeke, J.D., and Harlow, E. (1996). Genetic characterization of a mammalian protein-protein interaction domain by using a yeast reverse two-hybrid system. Proc. Natl. Acad. Sci. USA *93*, 10321–10326.

von Kries, J.P., Winbeck, G., Asbrand, C., Schwarz-Romond, T., Sochnikova, N., Dell'Oro, A., Behrens, J., and Birchmeier, W. (2000). Hot spots in β-catenin for interactions with LEF-1, conductin and APC. Nat. Struct. Biol. *7*, 800–807.

Wang, J.H., Yan, Y.W., Garrett, T.P., Liu, J.H., Rodgers, D.W., Garlick, R.L., Tarr, G.E., Husain, Y., Reinherz, E.L., and Harrison, S.C. (1990). Atomic structure of a fragment of human CD4 containing two immunoglobulin-like domains . Nature *348*, 411–418.

Wharton, R.P., Sonoda, J., Lee, T., Patterson, M., and Murata, Y. (1998). The Pumilio RNA-binding domain is also a translational regulator. Mol. Cell *1*, 863–872.

Wharton, R.P., and Struhl, G. (1991). RNA regulatory elements mediate control of Drosophila body pattern by the posterior morphogen nanos. Cell *67*, 955–967.

Willis, A.E. (1999). Translational control of growth factor and proto-oncogene expression. Int. J. Biochem. Cell. Biol. *31*, 73–86.

Wreden, C., Verrotti, A.C., Schisa, J.A., Lieberfarb, M.E., and Strickland, S. (1997). Nanos and pumilio establish embryonic polarity in Drosophila by promoting posterior deadenylation of hunchback mRNA. Development *124*, 3015–3023.

Zamore, P.D., Bartel, D.P., Lehmann, R., and Williamson, J.R. (1999). The PUMILIO-RNA interaction: a single RNA-binding domain monomer recognizes a bipartite target sequence. Biochemistry *38*, 596–604.

Zamore, P.D., Williamson, J.R., and Lehmann, R. (1997). The Pumilio protein binds RNA through a conserved domain that defines a new class of RNA-binding proteins. RNA *3*, 1421–1433.

Zhang, B., Gallegos, M., Puoti, A., Durkin, E., Fields, S., Kimble, J., and Wickens, M.P. (1997). A conserved RNA-binding protein that regulates sexual fates in the C. elegans hermaphrodite germ line. Nature *390*, 477–484.