

Structure of PTB Bound to RNA: Specific Binding and Implications for Splicing Regulation

Florian C. Oberstrass,^{1,4*} Sigrid D. Auweter,^{1,4*} Michèle Erat,^{1*}
Yann Hargous,¹ Anke Henning,¹ Philipp Wenter,³
Luc Raymond,³ Batoul Amir-Ahmady,² Stefan Pitsch,³
Douglas L. Black,² Frédéric H.-T. Allain^{1†}

The polypyrimidine tract binding protein (PTB) is a 58-kilodalton RNA binding protein involved in multiple aspects of messenger RNA metabolism, including the repression of alternative exons. We have determined the solution structures of the four RNA binding domains (RBDs) of PTB, each bound to a CUCUCU oligonucleotide. Each RBD binds RNA with a different binding specificity. RBD3 and RBD4 interact, resulting in an antiparallel orientation of their bound RNAs. Thus, PTB will induce RNA looping when bound to two separated pyrimidine tracts within the same RNA. This leads to structural models for how PTB functions as an alternative-splicing repressor.

The 58-kD polypyrimidine tract binding protein 1 (PTB1) is an abundant eukaryotic RNA binding protein implicated in several aspects of mRNA metabolism, including splicing regulation (1), internal ribosomal entry site

(IRES)-mediated translation initiation (2), 3' end processing (3), and mRNA stability (4). The mechanisms of PTB action in these processes are not well understood. For example, in splicing regulation, PTB may simply compete with other splicing factors in binding RNA or it may prevent the splicing machinery from assembling a spliceosome (5–8). PTB is composed of four RNA binding domains (RBDs) of the RBD/RRM/RNP type (RNA recognition motif/ribonucleoprotein) (9). The structures of all four RBDs of PTB in the free state have previously been determined by nuclear magnetic resonance (NMR) spectroscopy and show that all four domains adopt the typical RBD fold with a $\beta\alpha\beta\alpha\beta$ topology, but RBD2 and RBD3 are extended by an addi-

tional fifth β strand ($\beta 5$) (10, 11). Three linkers of 51, 91, and 23 amino acids separate the RBDs along the polypeptide sequence. PTB binds with an affinity of 1 to 10 nM (K_d) to sequences containing 15 to 25 pyrimidines, with a preference for pyrimidine tracts containing a UCUU sequence element (12, 13). We aim here at elucidating how PTB recognizes RNA so as to understand its different biological functions.

To study the interaction between human PTB and RNA by NMR spectroscopy, we chose the oligonucleotide 5' CUCUCU 3' that is found in several intronic regulatory sequences; for example, several copies surround the alternatively spliced c-src N1 exon (14, 15). Titration experiments show that PTB binds to the oligonucleotide in the fast-exchange regime relative to the NMR time scale. Protein saturation is reached at four equivalents of the hexanucleotide, which suggests that each RBD of PTB binds to one RNA molecule. Indeed, isolated RBD1 or RBD2 bind one RNA molecule. Similarly, the C terminus of PTB, containing RBD3, RBD4, and their interdomain linker (protein construct RBD34) binds to two CUCUCU sequences. ¹⁵N-TROSY (transverse relaxation-optimized spectroscopy) spectra of the individual domains (RBD1, RBD2, and RBD34) complexed with RNA are identical to the ¹⁵N-TROSY spectrum of full-length PTB in complex with four equivalents of CUCUCU in the dispersed regions of the spectra (fig. S1). Thus, RBD1, RBD2, and RBD34 are independent, and the interdomain linkers (except between RBD3 and RBD4) are flexible and do not participate in RNA binding. The difference between PTB1 and its other human isoforms lies in the length of these two linkers; therefore, their functional difference is

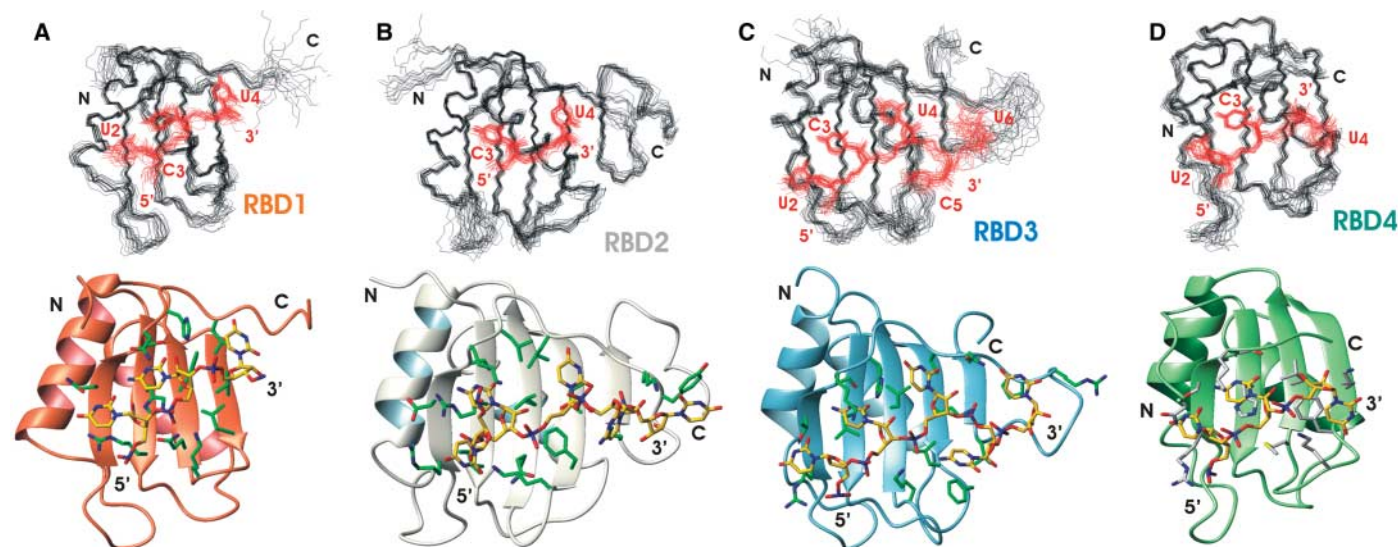


Fig. 1. Overlay of the structural ensemble and view of the most representative structure for each RBD of PTB in complex with RNA. Overlay of the 20 lowest energy structures of PTB RBD1 (A, top), RBD2 (B, top), RBD3 (C, top), and RBD4 (D, top) in complex with CUCUCU, the protein

backbone, and the RNA heavy atoms of the bound nucleotides are shown. View of the most representative structure of each RBD in complex with CUCUCU: RBD1 (A, bottom), RBD2 (B, bottom), RBD3 (C, bottom) and RBD4 (D, bottom).

likely not to reside in their RNA binding specificity.

To determine the NMR structures of PTB RBD1, RBD2, and RBD34, each in complex with RNA, we used three oligonucleotides: CUCUCU, UCUCU, and CUCU. The structure of the RBDs in complex was initially determined with the hexanucleotide, because the affinity of the RBDs is higher for this RNA. However, each individual domain binds to the hexamer in multiple registers. Therefore, the complex with shorter oligonucleotides helped to identify the intermolecular nuclear Overhauser effects (NOE) crosspeaks. A total of 1574 interproton distance restraints for RBD1 (including 46 intermolecular), 1950 restraints for RBD2 (including 53 intermolecular), and 4587 restraints for RBD34 (including 64 and 54 intermolecular for RBD3 and RBD4, respectively) were used to obtain precise structures (Fig. 1 and table S1).

In all four RBD-RNA complexes, the nucleotides are spread across the β -sheet surface. Both RBD1 and RBD4 bind the $U_2C_3U_4$ triplet, RBD2 binds the C_3U_4 doublet and U_6 , and RBD3 binds the $U_2C_3U_4C_5U_6$ quintet (Fig. 1). When bound, U_2 , C_3 , and U_4 are each positioned on a separate β strand [4, β 1, and β 2, respectively (Fig. 2A)]. The PTB-RNA interactions differ from all other RBD-RNA structures in that the β 3 strand of each

RBD, which includes part of the RNP1 consensus sequence, participates only weakly in RNA binding (fig. S2). In particular, position 3 and position 5 of the RNP1 motif are usually occupied by aromatic residues that make extensive hydrophobic interactions with the RNA bases and sugars (9). Neither of these residues is involved in such interactions in the RBDs of PTB. Instead, this role is fulfilled by the hydrophobic side chains located in β 2 (Fig. 2A and fig. S2). PTB RBD2 and RBD3 share an unusual C-terminal extension of the RBD (10, 11) with an additional fifth β strand that has been seen only in these two RBDs. These extensions participate in RNA binding and allow the domain to bind one (RBD2) or two (RBD3) additional nucleotides (Fig. 2, B and C). The overall fold of each PTB RBD is identical in the free (10, 11) and bound state. However, a few loops become more ordered upon RNA binding, such as the β 4- β 5 loop in RBD2 and RBD3. A detailed description of the protein-RNA interactions can be found in (16).

PTB is a sequence-specific RNA binding protein. C_3 is specifically recognized by all four RBDs of PTB, and U_4 by all RBDs except RBD4 (Fig. 2A). U_2 is specifically recognized by RBD1, RBD3, and RBD4. The U_2 O2 is hydrogen-bonded with the C_3 amino group, and its imino group is hydrogen-bonded to Asn, Thr, and Ser side chains in RBDs 1, 3,

and 4, respectively (Fig. 2A). These side chains act as hydrogen-bond acceptors but also have the ability to act as donors. Therefore, a cytosine (C_2) could be accommodated instead of U_2 . The structure shows that each RBD of PTB recognizes a different RNA sequence that defines its binding register, YCU for RBD1, CU(N)N for RBD2, YCUNN for RBD3, and YCN for RBD4 (Y indicating a pyrimidine and N any nucleotide). These four different RNA recognition consensi for PTB explain the difficulties in finding a clear RNA consensus sequence using SELEX (systematic evolution of ligands by developmental enrichment) (12, 13). However, because 10 of the 14 bound pyrimidines (71%) in our complex are recognized by specific base contacts, the structure reflects the enrichment (72%) of pyrimidines found with SELEX (12). PTB sequence specificity is biologically critical, because mutating PTB binding sites to poly(C) or poly(U) alters PTB function in splicing considerably (12, 14, 15, 17). This preference of PTB for poly(CU) sequences may explain how PTB can compete effectively with the essential splicing factor U2AF (U2 accessory factor) for 3' splice sites containing CU sequences, because U2AF prefers a pure poly(U) sequence (13). However, this competition mechanism does not fully explain the role of PTB as a splicing repressor (5, 8, 18).

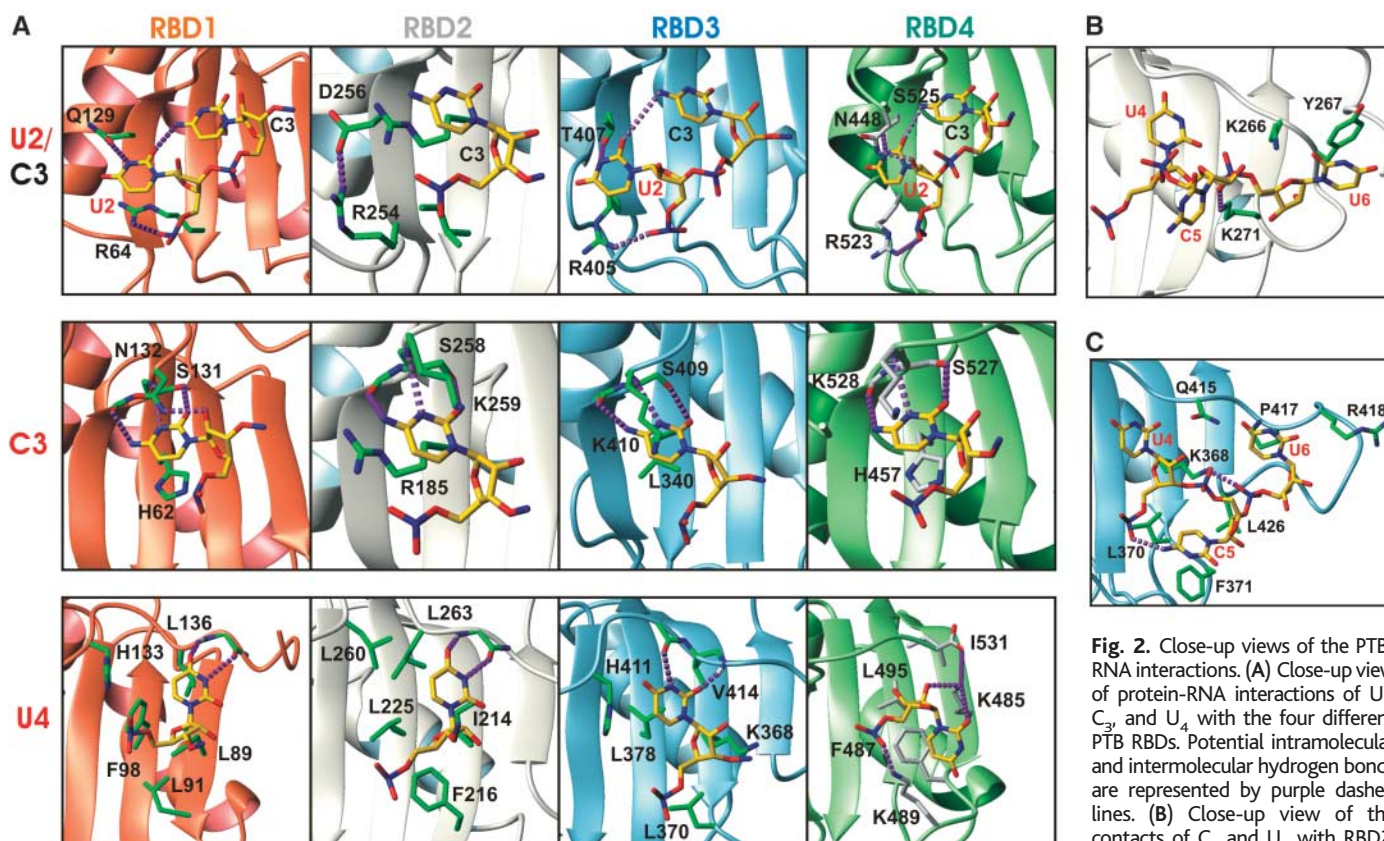


Fig. 2. Close-up views of the PTB-RNA interactions. (A) Close-up view of protein-RNA interactions of U_2 , C_3 , and U_4 with the four different PTB RBDs. Potential intramolecular and intermolecular hydrogen bonds are represented by purple dashed lines. (B) Close-up view of the contacts of C_5 and U_6 with RBD2. (C) Close-up view of the contacts between C_5 and U_6 with RBD3.

Most interestingly, we found that PTB RBD3 and RBD4 interact extensively and have a fixed orientation relative to one another. More than 20 protein side chains located in helices 1 and 2 of RBD3, in helix 2 of RBD4, and in the interdomain linker form a hydrophobic core between the two domains (Fig. 3). Mutations of only three side chains within helix2 of RBD4 are sufficient to abolish this interdomain interaction (fig. S3). As a result of this orientation, RBD34 can bind two pyrimidine tracts within the same RNA only if they are separated by a linker sequence. We found that an RNA with a spacer of 15 nucleotides between two CUCUCU hexamers is bound with the highest affinity (16) (fig. S4). This distinguishes PTB RBD34 from other proteins with two consecutive RBDs, such as poly(A) binding protein (PABP) (19), Sex-lethal (20), or nucleolin (21), in which both domains bind

immediately adjacent stretches on a continuous RNA with high affinity. Hence, PTB RBD34 has the capacity to bring two distantly located pyrimidine tracts to a distance as close as 30 Å (the minimum distance that separates the two RNAs across the protein) and induce RNA looping (Fig. 4A).

PTB has been implicated in the splicing repression of several short alternative exons (1). RNA binding sites for PTB tend to be clustered in the surrounding intron sequences but can be found within the regulated exon itself. The location and the number of these binding sites differ, but in most cases (7, 22), it is important that multiple PTB binding sites are present. These multiple PTB binding sites can be positioned both upstream and downstream of the alternative exon (14, 23–27) or of a branch point (28). One model for how PTB represses splicing is that PTB loops out the

exon or the branch point by binding several pyrimidine tracts upstream and downstream. This in turn prevents binding of spliceosomal components or their subsequent assembly with later components (1). These models usually involve the dimerization of PTB. However, recent studies indicate that PTB may not always dimerize to repress splicing (7). It was also shown that PTB is a monomer in solution (5, 10). Our structural work and the footprinting results obtained for PTB on exon 9 of the γ -aminobutyric acid- γ 2 (GABA- γ 2) pre-mRNA (6) let us propose a model in which only a single PTB molecule is necessary to loop out a branch-point adenine (Fig. 4B).

Based on this initial topology, we constructed several other models based on identified PTB binding sites, showing how PTB loops out alternative exons (Fig. 4, C to F). The models with two PTB molecules explain the cooperativity observed in binding studies with such RNAs (5). All these structural models highlight the critical role shown for RBD4 in splicing repression (6), because removing RBD4 prevents RNA looping around the branch point or the alternative exon.

Our findings suggest that it would be relatively easy for a PTB-repressible exon to emerge during evolution. Because each constitutive intron contains a long uracil-rich pyrimidine tract as part of its 3' splice-site consensus sequence, spontaneous mutations into cytosines (three U to C mutations would be sufficient) would create strong binding sites for PTB RBD1, 2, and 3, and a weaker one for U2AF that binds preferentially to poly(U) sequences (13). This mutated 3' splice site combined with a short pyrimidine sequence (containing UC or CC dinucleotides that serve as binding sites for RBD4) in the nearby exon or intron would result in an ideal binding site for PTB to form an RNA loop. Such 3' splice-site mutations could convert a constitutive exon into an alternative exon repressed by PTB (Fig. 4, C and G). It is notable that alternative exons containing PTB binding sites in the 3' splice site can often be made constitutive by C to U changes in the polypyrimidine tract (15, 24, 29–31).

We found that PTB is both a sequence-specific RNA binding protein with a preference for CU tracts and an RNA remodeler with an ability to bring separated pyrimidine tracts into proximity. This dual capacity could allow PTB to antagonize U2AF binding, as seen in some systems, and also to loop out essential RNA elements, as seen in other systems. These structural findings are also interesting in relation to the role of PTB in IRES-mediated translation, where PTB is thought to have an RNA chaperone activity (32, 33).

References and Notes

1. E. J. Wagner, M. A. Garcia-Blanco, *Mol. Cell. Biol.* 21, 3281 (2001).
2. C. U. Hellen, P. Sarnow, *Genes Dev.* 15, 1593 (2001).

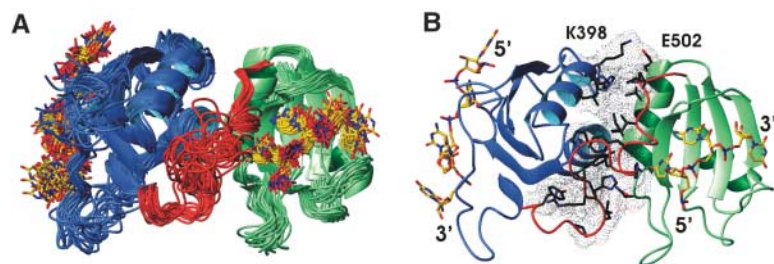
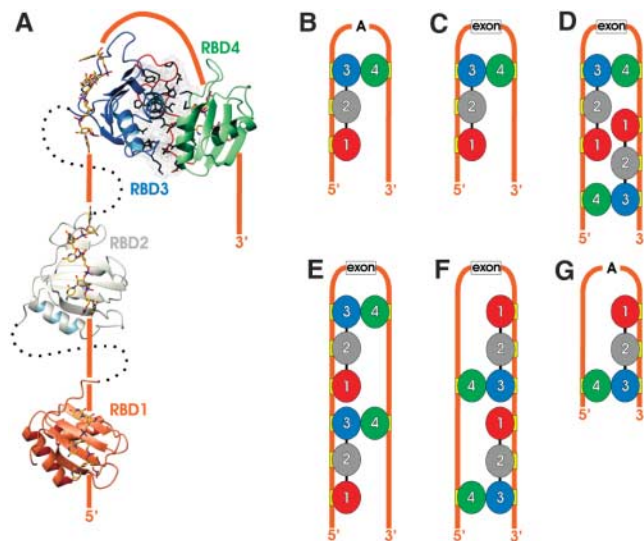


Fig. 3. PTB RBD3 and RBD4 interdomain interactions. (A) Overlay of the 20 lowest energy structures of PTB RBD34 in complex with CUCUCU. RBD3 is in blue, RBD4 in green, the interdomain linker in red, and the RNA in yellow. (B) View of the most representative structure. The side chains contributing to the interdomain interaction are shown by black sticks and by black dots representing their surfaces.

Fig. 4. Structural models representing how PTB loops out RNA (branch point or exon) and in turn represses splicing of alternative exons. (A) Model of the structure of PTB with a long RNA containing four pyrimidine tracts based on the structures of the three complexes we have determined. The dotted black line represents the flexible interdomain linkers between RBD1 and RBD2 and between RBD2 and RBD3. The thick orange line represents the long RNA. Note that the orientation of RBD1 and RBD2 along the RNA is not yet determined. (B) Model based on the example of GABA- γ 2 exon 9 repression (6). The branch-point adenine is looped out. (C) Model illustrating how one molecule of PTB can repress a short alternative exon (represented by a box). Model based on splicing regulation of the calcitonin and calcitonin gene-related peptide (CT/CGRP) 3' terminal exon (26) and the cTNT exon 5 (27). (D to F) Models involving two or more PTB molecules. The models are based on data obtained for α -tropomyosin exon 3 (30), c-src N1 exon (34), and FGF-R2 exon IIIb (25), respectively. (G) Hypothetical model illustrating how PTB loops out a branch point with PTB binding to the 3' splice site and a short pyrimidine tract upstream of the branch point. The RNA is represented as a thick orange line. Red, gray, blue, and green ovals represent the position of RBD1, RBD2, RBD3, and RBD4 of PTB bound to RNA, respectively.



3. P. Castelo-Branco *et al.*, *Mol. Cell Biol.* **24**, 4174 (2004).
4. K. P. Knoch *et al.*, *Nat. Cell Biol.* **6**, 207 (2004).
5. B. Amir-Ahmady, P. L. Boutz, V. Markovstov, M. Phillips, D. L. Black, *RNA* **11**, 699 (2005).
6. H. Liu, W. Zhang, R. B. Reed, W. Liu, P. J. Grabowski, *RNA* **8**, 137 (2002).
7. J. M. Izquierdo *et al.*, *Mol. Cell* **19**, 475 (2005).
8. S. Sharma, A. Falick, D. Black, *Mol. Cell* **19**, 485 (2005).
9. C. Maris, C. Dominguez, F. H. T. Allain, *FEBS J.* **272**, 2118 (2005).
10. P. J. Simpson *et al.*, *Structure* **12**, 1631 (2004).
11. M. R. Conte *et al.*, *EMBO J.* **19**, 3132 (2000).
12. I. Perez, C. H. Lin, J. G. McAfee, J. G. Patton, *RNA* **3**, 764 (1997).
13. R. Singh, J. Valcarcel, M. R. Green, *Science* **268**, 1173 (1995).
14. R. C. Chan, D. L. Black, *Mol. Cell Biol.* **15**, 6377 (1995).
15. R. C. Chan, D. L. Black, *Mol. Cell Biol.* **17**, 4667 (1997).
16. Additional results, materials, and methods are available as supporting material on Science Online.
17. R. F. Roscigno, M. Weiner, M. A. Garcia-Blanco, *J. Biol. Chem.* **268**, 11222 (1993).
18. N. Gromak *et al.*, *EMBO J.* **22**, 6356 (2003).
19. R. C. Deo, J. B. Bonanno, N. Sonenberg, S. K. Burley, *Cell* **98**, 835 (1999).
20. N. Handa *et al.*, *Nature* **398**, 579 (1999).
21. F. H. Allain, P. Bouvet, T. Dieckmann, J. Feigon, *EMBO J.* **19**, 6870 (2000).
22. H. Shen, J. L. Kan, C. Ghigna, G. Biamonti, M. R. Green, *RNA* **10**, 787 (2004).
23. C. Gooding, G. C. Roberts, G. Moreau, B. Nadal-Ginard, C. W. Smith, *EMBO J.* **13**, 3861 (1994).
24. J. Southby, C. Gooding, C. W. Smith, *Mol. Cell Biol.* **19**, 2699 (1999).
25. E. J. Wagner, M. A. Garcia-Blanco, *Mol. Cell* **10**, 943 (2002).
26. H. Lou, D. M. Helfman, R. F. Gagel, S. M. Berget, *Mol. Cell Biol.* **19**, 78 (1999).
27. N. Charlet-B., P. Logan, G. Singh, T. A. Cooper, *Mol. Cell* **9**, 649 (2002).
28. M. Ashiya, P. J. Grabowski, *RNA* **3**, 996 (1997).
29. N. Gromak, A. J. Matlin, T. A. Cooper, C. W. Smith, *RNA* **9**, 443 (2003).
30. C. Gooding, G. C. Roberts, C. W. Smith, *RNA* **4**, 85 (1998).
31. R. P. Carstens, E. J. Wagner, M. A. Garcia-Blanco, *Mol. Cell Biol.* **20**, 7388 (2000).
32. E. V. Pilipenko *et al.*, *Genes Dev.* **14**, 2028 (2000).
33. S. A. Mitchell, K. A. Spriggs, M. J. Coldwell, R. J. Jackson, A. E. Willis, *Mol. Cell* **11**, 757 (2003).
34. M. Y. Chou, J. G. Underwood, J. Nikolic, M. H. Luu, D. L. Black, *Mol. Cell* **5**, 949 (2000).
35. We are grateful to R. Stefl (ETH Zürich) for help with

the structure calculation and to L. Skrisovska, C. Maris, and G. Wider (ETH Zürich), S. Curry, and S. Matthews (Imperial College, London) for helpful discussions. We also thank R. Peterson and J. Feigon (UCLA) for sharing recently developed NMR pulse programs. This investigation was supported by grants from the Swiss National Science Foundation, the Structural Biology National Center of Competence in Research to S.P. and F.H.T.A. and by the Roche Research Fund for Biology at ETH Zürich to F.H.T.A. F.H.T.A. is a European Molecular Biology Organization Young Investigator. The coordinates of the structures of PTB RBD1, RBD2, and RBD34 in complex with CUCUCU have been deposited in the Protein Data Bank with accession codes 2AD9, 2ADB, and 2ADC, respectively.

Supporting Online Material

www.sciencemag.org/cgi/content/full/309/5743/2054/DC1

Materials and Methods

SOM Text

Figs. S1 to S4

Table S1

References

26 April 2005; accepted 24 August 2005

10.1126/science.1114066

Direct Observation of the Three-State Folding of a Single Protein Molecule

Ciro Cecconi,^{1,2*} Elizabeth A. Shank,^{1*} Carlos Bustamante,^{1,2,3†}
Susan Marqusee^{1†}

We used force-measuring optical tweezers to induce complete mechanical unfolding and refolding of individual *Escherichia coli* ribonuclease H (RNase H) molecules. The protein unfolds in a two-state manner and refolds through an intermediate that correlates with the transient molten globule-like intermediate observed in bulk studies. This intermediate displays unusual mechanical compliance and unfolds at substantially lower forces than the native state. In a narrow range of forces, the molecule hops between the unfolded and intermediate states in real time. Occasionally, hopping was observed to stop as the molecule crossed the folding barrier directly from the intermediate, demonstrating that the intermediate is on-pathway. These studies allow us to map the energy landscape of RNase H.

Protein folding remains a major unsolved challenge for modern molecular biology. Theoretical studies emphasize the potential heterogeneous nature of the process; however, traditional bulk biochemical experiments often mask this complexity in their inherent ensemble averaging. For instance, many proteins are observed to populate partially structured conformations early during the folding process (1, 2). These so-called burst-phase intermediates (I) are often formed within the dead time of

the measuring instrument (usually milliseconds) and therefore cannot be characterized directly. Thus, controversy remains about whether these intermediates are on-pathway and productive to protein folding, or are off-pathway and do not lead directly to the folded state (2, 3). It is also unclear whether these intermediates constitute distinct thermodynamic states or simply represent a redistribution of the unfolded ensemble when exposed to native conditions (4). These issues can be addressed with the use of single-molecule manipulation to follow in real time the trajectories of individual protein molecules during mechanically induced unfolding/refolding processes.

Previous single-molecule manipulation studies have used the atomic force microscope (AFM) to characterize the mechanical unfolding and refolding of tandem repeats of globular protein domains (5–11). Although informative,

this approach has been limited to characterizing the high-force unfolding behavior of proteins. It has been difficult to observe the refolding process directly or to monitor the equilibrium between the folded and unfolded states using this method, in part because of the high spring constants of commercially available cantilevers and the correspondingly high loading rates they exert. In contrast, optical tweezers, with considerably reduced mechanical stiffness and loading rates, can overcome these limitations, as shown in RNA unfolding studies (12). Optical tweezers have been used to mechanically unfold the multiglobular protein titin (5, 13); although the resulting data revealed the overall mechanical unfolding response of the molecule, the heterogeneous composition of this protein obscured the unfolding of the individual globular domains and refolding could not be observed directly.

Here, we report the complete force-induced unfolding and refolding trajectory of a single molecule of the *E. coli* protein ribonuclease H* Q4C/V155C (14) (referred to herein as RNase H) with the use of force-measuring optical tweezers (15). RNase H is a 155-residue, single-domain protein whose structure, stability, and folding mechanism have been extensively characterized using bulk biochemical techniques. The central portion of the polypeptide forms the core of the protein (Fig. 1A), being both the first region to fold (16) and the most stable region of the native structure (17). This observation and additional kinetic and mutagenesis studies support the hypothesis that RNase H folds via a hierarchical mechanism, in which the most stable regions of the native structure form first during folding (3, 18). It has been impossible, however, to demonstrate this behavior directly with traditional approaches. This wealth of both equilibrium

¹Department of Molecular and Cell Biology and Institute for Quantitative Biology, ²Department of Physics, ³Howard Hughes Medical Institute, University of California, Berkeley, CA 94720, USA.

*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: carlos@alice.berkeley.edu; marqusee@berkeley.edu