# Crystal Structure of the T4 regA Translational Regulator Protein at 1.9 Å Resolution

ChulHee Kang,*† Rodney Chan, Imre Berger, Curtis Lockshin, Louis Green, Larry Gold, Alexander Rich†

The translational regulator protein regA is encoded by the T4 bacteriophage and binds to a region of messenger RNA (mRNA) that includes the initiator codon. RegA is unusual in that it represses the translation of about 35 early T4 mRNAs but does not affect nearly 200 other mRNAs. The crystal structure of regA was determined at 1.9 Å resolution; the protein was shown to have an α-helical core and two regions with antiparallel β sheets. One of these β sheets has four antiparallel strands and has some sequence homology to RNP-1 and RNP-2, which are believed to be RNA-binding motifs and are found in a number of known RNA-binding proteins. Structurally guided mutants may help to uncover the basis for this variety of RNA interaction.

The three-dimensional (3D) structures of many DNA-binding proteins have been solved, and the mechanism of sequence-specific DNA recognition is largely understood. In contrast, little is known about the recognition systems of proteins that bind to RNA in a sequence-specific manner. Most information concerning such interactions has come from crystal structures of the tRNA aminoacyl synthetases bound to their cognate tRNA molecules (*1*). Many proteins that bind to RNAs and influence mRNA splicing or translation have characteristic short amino acid sequences that are believed to participate in RNA recognition (*2*). These proteins are present in eubacteria and eukaryotes, and they usually bind to only one or a few target RNAs. However, the bacteriophage T4 regA protein binds many early T4 mRNAs and diminishes translation by blocking ribosome movement (*3*). Phage T4 encodes nearly 300 proteins, and as many as 35 of the 200 early genes are regulated by the regA protein (*3*). In comparing binding sites of target mRNAs, it has not been possible to identify a consensus sequence (*3*, *4*). How does a protein with 122 amino acids translationally repress some 35 different mRNAs while ignoring the other mRNAs that are found in an infected cell at the same time? Here, we present the 3D structure of the regA protein and show that it shares some features with other proteins that are known to be important in binding to RNA.

The expression, purification (*5*), crystallization, and structure determination of the regA protein are described in Tables 1 and 2. The crystals diffracted to better than 1.9 Å resolution and belong to the orthorhombic space group $P2_12_12_1$ with two molecules in the asymmetric unit (Table 1). The α carbon positions of the two molecules in the asymmetric unit are shown in Fig. 1. The molecule has three large α-helical segments and one turn of a $3_{10}$ helix (Fig. 2). There are two regions with β pleated sheets: Sheet A contains three antiparallel strands and one parallel strand and sheet B contains four antiparallel strands. In the crystal structure, two identical polypeptide chains are brought together to form a dimer (Fig. 1) with a noncrystallographic pseudo-twofold axis. The dimer interface is stabilized by a symmetrical pair of intersubunit salt bridges between $Arg^{91}$ and $Glu^{68}$ as well as by a symmetrical pair of intersubunit backbone hydrogen bonds between the carbonyl group of $Thr^{92}$ of one molecule and the amino group of the corresponding $Thr^{92}$ of the other molecule. These interactions suggest the possibility that the molecule may exist as a dimer when it is biologically active; in fact, the molecule exists as a dimer in dilute solution (*6*). Several RNA-binding proteins are known to exist as dimers, including the viral MS2 coat protein (*7*).

The main structural differences between the two regA molecules in the asymmetric unit are found near residues 95 to 100 (Fig. 1) and at the COOH-terminal region (residues 119 to 122). The two COOH-terminal residues are disordered in one molecule. The root mean square (rms) deviation between the positions of the α carbons of the two molecules is 1.2 Å. When residues 95 to 100 and 119 to 122 are taken out, the rms deviation drops to 0.6 Å.

β sheet structures appear to be important components in RNA recognition. In the crystal structure of glutaminyl-tRNA complexed to its tRNA synthetase, a β pleated

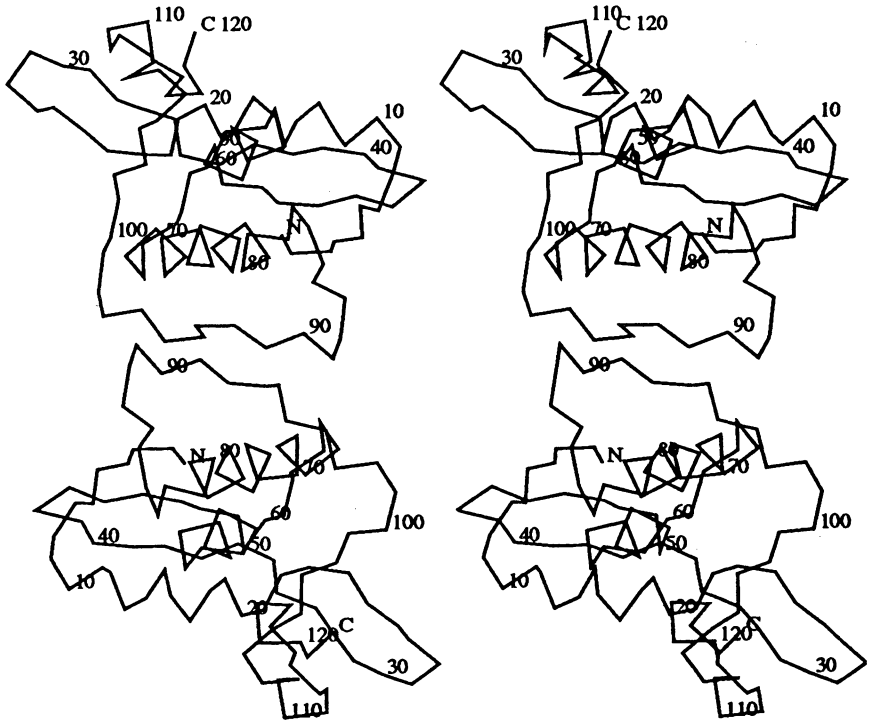


**Fig. 1.** Stereo diagram illustrating the position of α carbon atoms in the two regA molecules found in the asymmetric unit. The last two COOH-terminal residues in the upper molecule are disordered. The molecules are organized around a noncrystallographic pseudo-twofold axis perpendicular to the page.

C. Kang, R. Chan, C. Lockshin, A. Rich, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
I. Berger, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA, and Department of Biophysical Chemistry, Hannover Medical School, 30623 Hannover, Germany.
L. Green, Nexagen Inc., 2869 Wilderness Place, Boulder, CO 80302, USA.
L. Gold, Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, CO 80302, USA.

*Present address: Department of Biochemistry and Biophysics, Washington State University, Pullman, WA 99164, USA.
†To whom correspondence should be addressed.

**Table 1.** Crystallization data. The regA protein was purified as described (*5*). Crystals were grown from a solution containing the protein (3 mg/ml), 12.5% (w/v) polyethylene glycol (PEG; molecular weight 4000), 0.01 M $MgCl_2$, 0.05 M tris buffer (pH 8.5), and 0.7 M $NH_2SO_4$. Droplets (10 μl) were equilibrated against a reservoir with twice the concentration of salts and PEG. Within 5 to 6 days, orthorhombic crystals appeared (approximate dimensions 0.5 by 0.5 by 1.0 mm). The space group was $P2_12_12_1$ with unit cell dimensions $a$ = 83.56 Å, $b$ = 86.02 Å, and $c$ = 43.86 Å. Crystals diffracted to somewhat better than 2 Å resolution. Complete diffraction data to 2.5 Å were collected with the Xuong-Hamlin area detector at the University of California at San Diego Research Center and with a Rigaku R-Axis II imaging plate. Heavy atom derivatives were obtained by soaking the crystals in stabilizing solutions, and 3.5 Å data sets were collected for all derivatives, together with anomalous data. Heavy atom sites were determined by difference Patterson maps and confirmed by anomalous difference Patterson maps.

| Compound | Soaking | | Heavy atom positions | | | Occupancy | Temperature factor | Site | Residue |
|---|---|---|---|---|---|---|---|---|---|
| | Concentration | Time (days) | $x$ | $y$ | $z$ | | | | |
| $K_2PtCl_6$ | 10.0 mM | 3 | 0.5937455 | 0.0275159 | 0.1789431 | 52 | 21 | A | $Cys^{36}$ I |
| | | | 0.2782735 | 0.8734361 | 0.1894905 | 17 | 25 | B | $Cys^{36}$ II |
| | | | 0.5542349 | 0.0183559 | 0.3006860 | 18 | 25 | C | $Arg^{70}$ I |
| | | | 0.7961411 | 0.2351236 | 0.6303344 | 10 | 30 | D | $Asp^{80}$ II |
| *p*-Chloromercuriphenylsulfonic acid | Saturated | 2 | 0.5915480 | 0.0276858 | 0.1767269 | 58 | 21 | A | $Cys^{36}$ I |
| | | | 0.2768740 | 0.8722591 | 0.1667804 | 33 | 25 | B | $Cys^{36}$ II |
| | | | 0.5577730 | 0.0203111 | 0.2946805 | 22 | 27 | C | $Arg^{70}$ I |
| | | | 0.2256121 | 0.0126724 | 0.1248634 | 11 | 36 | E | $Tyr^{33}$ II |
| 2-Chloromercuri-4-nitrophenol | Saturated | 2 | 0.5093997 | 0.026232 | 0.1776977 | 57 | 22 | A | $Cys^{36}$ I |
| | | | 0.278729 | 0.867304 | 0.1623901 | 43 | 26 | B | $Cys^{36}$ II |
| | | | 0.5814127 | 0.0317296 | 0.2958603 | 17 | 31 | C | $Arg^{70}$ I |

sheet system functions as a sequence-specific binding site for the anticodon (*1*). The small RNA binding domain of the ribonuclear protein U1A binds to a stem-loop structure of the U1 RNA, and the structure of the protein-RNA complex has recently been solved (*8*, *9*). The protein-RNA cocrystal reveals the importance of the β sheet in RNA binding. The U1 RNA forms a stem-loop structure with the loop nucleotides splayed out from the center, where they interact with the central two strands (RNP-1 and RNP-2) of the four-stranded β sheet and the COOH-terminal region of the protein. The two sequence motifs, RNP-1 and RNP-2, have been found as components of a large number of known RNA-binding proteins (*2*). We therefore focused our attention on the β sheet regions of the regA protein.

An electron density map of a portion of β sheet B of regA is shown in Fig. 3. Side chains of this four-stranded β sheet and side chains of helices A and C form a hydrophobic core. Two strands of the β sheet region have sequence similarities to RNP-1 and RNP-2, and the $NH_2$-terminal component of the β sheet has some similarities to that of RNP-2. Starting with residue 4, the regA sequence is ITLKK (*10*, *11*); the consensus

**Table 2.** Structure determination data. The multiple isomorphous replacement (MIR) method was used in combination with anomalous data and solvent-flattening techniques. MIR refinement was carried out with the PROTEIN program package (*15*). Reflections with a lack-of-closure error exceeding 2.1 times the rms value were rejected from the phase refinement. The mean figure of merit for the final combined phases of 3491 reflections was 0.71. A noise filtering procedure was applied to improve the 3.5 Å MIR phases (*16*). The resulting electron density map revealed several secondary structure elements, but the connections between some of these were ambiguous. Phasing was thus extended to 3 Å by Fourier back-transformation, which resolved some of the ambiguous regions. A partial backbone model represented by a noncontinuous polyalanine chain was fitted into the electron density map with the use of the program FRODO and the Evans & Sutherland 390 computer graphics system (*17*). Eventually most of the chain could be traced, except for ambiguities in some of the loop regions and in the assignment of side chains near the COOH-terminus. The conventional R factor of this initial model was 0.51%. The model was refined with the program X-PLOR (*18*). In the first several steps of positional refinement, only repulsive energy functions were activated. After regular nonbounded energy functions were used, the $R$ factor fell to 35.9% for data between 10 and 2.7 Å. Subsequent simulated annealing through molecular dynamics and slow cooling was carried out. The $R$ factor of the model then dropped to 24.6%. Then, the phases calculated from the backbone structure of the refined model that used the polyalanine chain were combined with the initial MIR phases at 3 Å. At this point, most of the larger side chains could be identified, which confirmed the accuracy of the chain tracing. This result was further supported by the location of the heavy atoms near $Cys^{36}$, $Arg^{93}$, and $Arg^{70}$. High-resolution data to 1.9 Å were then collected at Brookhaven National Laboratories (beamline X12C). After further refinement, the $R$ factor at present is 18.0% for 18,037 reflections above the 2σ level (based on the structure factor $F$) between 7 and 1.9 Å resolution (data completeness is 74%; $R_{sym}$ is 4.5%, where $R_{sym} = \Sigma_h\Sigma_i \mid I_{h,i} - I_h \mid / \Sigma_h\Sigma_i I_{h,i}$ and $I_h$ is the mean intensity of the $i$ observations of the unique reflection $h$). The rms deviations from ideality are 0.015 Å for bond lengths and 3.25° for bond angles. At the present stage of refinement, the average temperature factor $B$ for nonhydrogen atoms (2030 atoms in 242 residues) is 15.06 $Å^2$ (backbone) and 18.87 $Å^2$ (side chains), and 74 water molecules are included in the model. The coordinates have been deposited in the Brookhaven Data Base (accession number 1 REG).

| Resolution (Å) | Native | | $K_2PtCl_6$ | | *p*-Chloromercuriphenylsulfonic acid | | 2-Chloromercuri-4-nitrophenol | |
|---|---|---|---|---|---|---|---|---|
| | Figure of merit | Reflections | Phasing power* | $R_{Cullis}$† | Phasing power | $R_{Cullis}$ | Phasing power | $R_{Cullis}$ |
| 14.12 | 0.92 | 64 | 2.12 | 0.64 | 2.43 | 0.59 | 4.45 | 0.23 |
| 9.23 | 0.87 | 173 | 2.56 | 0.37 | 3.20 | 0.31 | 3.69 | 0.27 |
| 6.86 | 0.83 | 337 | 2.01 | 0.56 | 2.26 | 0.52 | 3.18 | 0.36 |
| 5.45 | 0.69 | 534 | 1.98 | 0.48 | 1.97 | 0.53 | 1.87 | 0.50 |
| 4.53 | 0.64 | 819 | 1.55 | 0.66 | 1.82 | 0.60 | 1.50 | 0.64 |
| 3.87 | 0.68 | 1117 | 1.34 | 0.72 | 1.96 | 0.51 | 1.57 | 0.68 |
| 3.50 | 0.70 | 1023 | 1.57 | 0.73 | 2.61 | 0.42 | 1.09 | 0.81 |

*Phasing power = $\Sigma f_H^2/\{\Sigma[F_{PH(obs)} - F_{PH(calc)}]^2\}$, where $f_H$ is the atomic scattering factor for the heavy atom and $F_{PH(obs)}$ and $F_{PH(calc)}$ are the observed and calculated structure factors for the heavy atom derivative, respectively. †$R_{Cullis} = \Sigma|F_{PH(obs)} - F_{PH(calc)}|$ (for centric reflections).

sequence for RNP-2 is IYIKG (*2*). There is reasonable agreement in the first four residues of this comparison. The adjacent β strand peptide in regA has the sequence KGLYYIVH, starting with position 42; the consensus sequence for RNP-1, KGFG-FVXF, has several similarities. A difference is that the sequence similarity enters the loop region in the antiparallel β sheet in regA, whereas RNP-1 and RNP-2 are located closer to the center of the β sheet region in U1A (*8*).

It is interesting to compare the RNA recognition chores of U1A and regA. U1A recognizes a stem-loop structure in the U1 RNA in a sequence-specific manner; part of the RNA loop structure interacts with the U1A region containing the β sheet elements RNP-1 and RNP-2 (*9*). The regA molecule, on the other hand, recognizes many transcripts (*3*), as shown by nuclease protection experiments in which ribonuclease digestion of transcripts was inhibited where the protein was bound to the RNA. The segments that were protected in different transcripts varied in length from 16 to 28 nucleotides (*3*); most of them included the AUG initiator codon and the nearby Shine-Dalgarno ribosome recognition sequence. A remarkable finding is the absence of sequence similarity among the group of transcripts that bind to regA. Although the regA protein has been found to bind poly(rU) more than 100 times as tightly as any other ribonucleotide homopolymer (*12*), the abundance of uridines within all T4 translational initiation regions and in the entire genome suggests that recognition by regA of its targets must depend on factors other than simple measurement of uridines. It has been suggested that the regA protein may recognize secondary structural elements as well as sequences (*3*). It is not apparent from viewing the sequences that bind to regA that they can adopt a common secondary structure. Thus, RNA recognition by the regA protein is much more complex than recognition by proteins such as U1A. Nonetheless, U1A and regA have some similarities. The second amino acid in the RNP-1 consensus sequence is glycine; it is also found at the β sheet turn in U1A. Similarly, $Gly^{43}$ in regA is located at a turn. U1A has two basic regions near the RNP-1 β sheet region, called "jaws" (*8*), that straddle the RNA duplex (*9*). RegA has two pairs of lysine residues ($Lys^{41}$ and $Lys^{42}$; $Lys^{7}$ and $Lys^{8}$) that are in the same region as the U1A jaws.

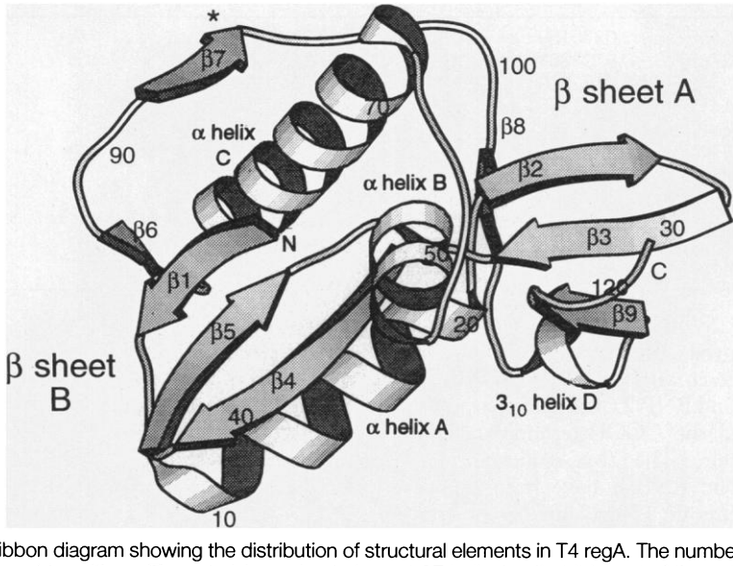


**Fig. 2.** Ribbon diagram showing the distribution of structural elements in T4 regA. The numbers refer to the amino acid positions. The asterisk on sheet element β7 indicates that it is part of the antiparallel β sheet found in the dimerization region of the molecule. The two major β sheet regions are labeled A and B. This figure was generated with the program Molscript (*19*).

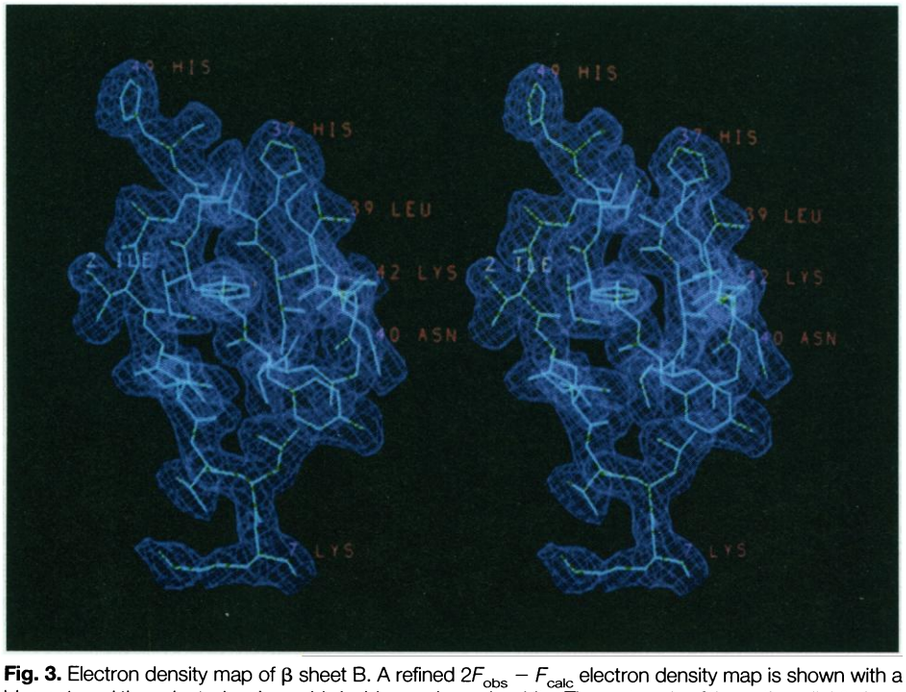


**Fig. 3.** Electron density map of β sheet B. A refined $2F_{obs} - F_{calc}$ electron density map is shown with a blue net, and the selected amino acids inside are shown in white. Three strands of the antiparallel β sheet are shown. The strand on the lower left comprises the $NH_2$-terminal seven amino acids from $Met^{1}$ to $Lys^{7}$ (MIEITLK) (*11*). The upper right contains two connected strands of the antiparallel β sheet extending from $His^{37}$ to $His^{49}$ (HILNKKGLYYIVH). Selected residues are labeled.

The regA protein of RB69 has been isolated and its sequence determined (*13*). The sequence has 78% identity with T4 regA. The regions of difference are scattered throughout the molecule; most of them are in the periphery in unstructured loops on the edge of the molecule. However, the core of the T4 regA molecule is largely conserved (Fig. 1). A few differences occur near the β sheets, and some are incorporated into α helix B.

A number of recent structure-function studies have focused on the regA protein. It has been proposed that the regions from $Val^{15}$ to $Ala^{25}$ and from $Arg^{70}$ to $Ser^{73}$ are particularly sensitive to mutations, both in T4 regA and in RB69 regA (*12*). In the crystal structure, residues 70 to 73 are part of the long α helix C. $Arg^{71}$ is connected by salt bridges to $Glu^{72}$ of the same helix C and also to $Glu^{52}$ of helix B. The regions between $Val^{15}$ and $His^{25}$ have been proposed to form a potential helix-turn-helix motif, as has been found in various DNA-binding proteins (*13*). However, the crystal

structure shows that residues 15 to 20 are part of α helix A and that residues 23 to 25 are involved in β sheet A. In T4 regA, changing $Ala^{25}$ to asparagine results in a significant alteration of the binding affinity (*12*). The side chain of $Ala^{25}$ is located in a hydrophobic environment generated by the side chains of β sheet A and helix D ($Val^{32}$, $Ile^{104}$, $Leu^{114}$, and $Trp^{112}$). Residue 25 is in the center of this β pleated sheet–forming region. It is likely that the integrity of this β sheet is maintained by the hydrophobic interactions involving the $Ala^{25}$ side chain. Mutating $Ala^{25}$ to valine appears to have no effect on the binding ability; this finding reinforces the interpretation that this hydrophobic binding domain is important for maintaining 3D structure.

Photochemical cross-linking to radioactive nucleotides, in combination with cyanogen bromide cleavage of the regA protein, led to the suggestion that two regions—residues 31 to 41 and 96 to 122—may be involved in RNA binding (*14*). Residues 31 to 41 are part of the β sheet system. In the COOH-terminal segment, residues 113 to 116 are involved in the same antiparallel β sheet system A that contains $Ala^{25}$ as well as residues 31 to 35. Thus, several types of experiments have suggested that β sheet system A may be involved in RNA recognition. The mutational experiments mentioned above suggest that sequences in the $NH_2$-terminal region between $Val^{15}$ and $Ala^{25}$ are important for RNA binding (*13*). This region is found in α helix A, which serves to connect β sheet regions A and B.

Understanding the manner in which regA binds RNA in a sequence-specific manner requires the determination of the structure of the protein complexed to one of its RNA substrates. In view of the numerous and complex types of specific interactions exhibited by this protein, it is possible that the region involved in recognition is not simple and may span a large portion of the molecule. For example, the two β sheet regions A and B that are 25 Å apart could both be involved in RNA interactions, as could the α-helical segment connecting them. Although the experiments described above suggest that one or more of these regions of the regA molecule constitute a possible site of RNA recognition, the interactions found between the U1A protein and the U1 RNA (*9*) emphasize the importance of the RNP-1 and RNP-2 motifs. Hence, mutational experiments with regA should be conducted in which, for example, the sequences of the central β sheets are made identical to the RNP-1 and RNP-2 consensus sequences. Because regA binds to various mRNAs with different affinities (*3*), changes induced in binding affinities may yield insight into the regA recognition system.

## REFERENCES AND NOTES


1. M. A. Rould *et al.*, *Nature* **352**, 213 (1991); J. Cavarelli *et al.*, *ibid.*, p. 181; V. Biou *et al.*, *Science* **263**, 1404 (1994).
2. C. G. Burd and G. Dreyfuss, *Science* **265**, 615 (1994).
3. E. S. Miller, J. D. Karam, E. Spicer, in *Molecular Biology of Bacteriophage T4*, J. D. Karam, Ed. (ASM Press, Washington, DC, 1994), pp. 193–205.
4. R. B. Winter *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 7822 (1987).
5. S. Unnithan *et al.*, *Nucleic Acids Res.* **18**, 7083 (1990); E. S. Miller, R. B. Winter, K. M. Campbell, S. D. Power, L. Gold, *J. Biol. Chem.* **260**, 13053 (1985).
6. Sedimentation equilibrium studies were carried out by Y. Kyogoku of Osaka University. In a dilute solution (0.03 mM) the molecular weight was close to the expected value of a dimer (personal communication).
7. K. Valegård *et al.*, *Nature* **345**, 36 (1990).
8. K. Nagai *et al.*, *ibid.* **348**, 515 (1990).
9. C. Oubridge, N. Ito, P. R. Evans, C.-H. Teo, K. Nagai, *ibid.* **372**, 432 (1994).
10. M. Trojanowska, E. S. Miller, J. Karam, G. Stormo, L. Gold, *Nucleic Acids Res.* **12**, 5979 (1984).
11. Single-letter abbreviations for the amino acid residues are as follows: E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; T, Thr; V, Val; Y, Tyr; and X, any amino acid.
12. K. R. Webster and E. K. Spicer, *J. Biol. Chem.* **265**, 19007 (1990).
13. C. E. Jozwik and E. S. Miller, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 5053 (1992).
14. K. R. Webster *et al.*, *J. Biol. Chem.* **267**, 26097 (1992).
15. W. Steigmann, *PROTEIN: A Package of Crystallographic Programs for Analysis of Proteins* (Max Planck Institute for Biochemistry, Martinsried, Germany, 1982).
16. B. C. Wang, *Methods Enzymol.* **115**, 90 (1985).
17. T. A. Jones, in *Computational Crystallography*, D. Sayer, Ed. (Oxford Univ. Press, Oxford, 1982), pp. 303–317; J. W. Pflugrath, M. A. Saper, F. A. Quiocho, in *Methods and Applications in Crystallographic Computing*, S. Hall and E. Ashida, Eds. (Clarendon, Oxford, 1984), pp. 404–407.
18. A. T. Brünger, X-PLOR, Yale University, New Haven, CT.
19. P. J. Kraulis, *J. Appl. Crystallogr.* **24**, 946 (1991).


20. We thank A. M. deVos, N. Xuong, W. Royer, and R. Sweet for assistance in data collection and A. Herbert, K. Lowenhaupt, S. Woelfl, L. Su, J. Spitzner, and Y. Kyogoku for helpful discussions. Supported by grants from NIH, NSF, Human Frontier Science (HFS), and the U.S. Office of Naval Research (at MIT) and from HFS and NSF (at the University of Colorado).

4 November 1994; accepted 14 February 1995


# Distinct Binding Specificities and Functions of Higher Eukaryotic Polypyrimidine Tract–Binding Proteins


Ravinder Singh, Juan Valcárcel, Michael R. Green



In higher eukaryotes, the polypyrimidine-tract (Py-tract) adjacent to the 3′ splice site is recognized by several proteins, including the essential splicing factor $U2AF^{65}$, the splicing regulator Sex-lethal (Sxl), and polypyrimidine tract–binding protein (PTB), whose function is unknown. Iterative in vitro genetic selection was used to show that these proteins have distinct sequence preferences. The uridine-rich degenerate sequences selected by $U2AF^{65}$ are similar to those present in the diverse array of natural metazoan Py-tracts. In contrast, the Sxl-consensus is a highly specific sequence, which can help explain the ability of Sxl to regulate splicing of *transformer* pre-mRNA and autoregulate splicing of its own pre-mRNA. The PTB-consensus is not a typical Py-tract; it can be found in certain alternatively spliced pre-mRNAs that undergo negative regulation. Here it is shown that PTB can regulate alternative splicing by selectively repressing 3′ splice sites that contain a PTB-binding site.


Several eukaryotic RNA-binding proteins preferentially interact with uridine-rich sequences and have thus been classified as Py-tract–binding proteins (*1*). Human $U2AF^{65}$ is an essential splicing factor that recognizes a wide variety of Py-tracts (*2*). *Drosophila* Sxl regulates 3′ splice-site switching of *transformer* (*tra*) pre-mRNA and exon skipping of its own pre-mRNA (*3*, *4*). The U-octamer ($U_8C$) sequence common to the non–sex-specific (NSS) Py-tract of *tra* and the male-specific Py-tract of *Sxl* pre-mRNA has been suggested to be the Sxl-binding site (*4*–*6*). PTB, also known as hnRNP I (*7*), was originally identified by its binding to the Py-tracts of adenoviral major late (Ad ML) and α-tropomyosin pre-mRNAs, and on this basis was proposed to be a splicing factor (*8*).


Howard Hughes Medical Institute, Program in Molecular Medicine, University of Massachusetts Medical Center, Worcester, MA 01605, USA.


To gain insight into RNA recognition and function of these proteins, we performed iterative in vitro genetic selection (*9*). The sequences of 20 to 30 complementary DNA (cDNA) clones from each selected pool revealed that $U2AF^{65}$, Sxl, and PTB had distinct RNA sequence preferences (Fig. 1, A to C). The $U2AF^{65}$-selected sequences were enriched in uridines that were frequently interrupted by two to three cytidines [UUUUUU(U/C)-CC(C/U)UUUUUUCC]. The relative distribution of nucleotides in the $U2AF^{65}$-selected pool (11.5% A, 29.8% C, 7.3%