Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Combinatorial crystallization of an RNA–protein complex

Danielle Bodrero Hoggan,[a] Jeffrey A. Chao,[a,b] G. S. Prasad,[a]† C. David Stout[a] and James R. Williamson[a,b]*

[a]Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA, and [b]Department of Chemistry and The Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA

† Present address: Syrrx, 10410 Science Center Drive, San Diego, California 92121, USA.

Correspondence e-mail: jrwill@scripps.edu

Received 10 October 2002
Accepted 19 December 2002

One of the most difficult steps in X-ray crystallography of a ribonucleoprotein (RNP) complex is obtaining crystals that diffract to high resolution. This paper describes a procedure for identifying the optimal lengths of the nucleic acid components that provide high-quality crystals of the RNP. Both strands of an RNA duplex were varied in a systematic manner to generate a large number of unique RNPs that were screened for crystallization behavior. As observed in the crystallization of other nucleic acids and their complexes, the exact length of the RNA chains was found to be critical in obtaining diffraction-quality crystals, even though the relative molecular weights of the protein and RNA components were ∼50 and ∼10 kDa, respectively. In particular, the helix–loop–helix structure in the mRNA for the *Saccharomyces cerevisiae* ribosomal protein L30, which functions as an autoregulatory element for L30 expression, was synthesized as two separate RNA chains of variable length (12–14 and 15–17 nucleotides). Duplex formation of these RNAs formed the asymmetric, internal loop-binding site for L30. 16 such RNA duplexes, varying by ±1 residue at the 5′ or 3′ end of either chain, were used to prepare 16 unique complexes with a maltose-binding protein-L30 fusion protein. The complexes were screened against 48 standard crystallization conditions in 2304 experiments, yielding 30 conditions with single crystals in the initial screen. The most promising of these is being used for structure determination.

1. Introduction

The critical first step in X-ray crystal structure determination of biological macromolecules is to obtain suitable crystals. While crystallographic methods have advanced spectacularly in recent years with the advent of synchrotron-radiation sources, cryogenic techniques for data collection, area detectors and new procedures for phase determination and refinement, the problem of obtaining diffraction-quality crystals often remains a bottleneck. This is particularly true of complexes containing RNA, which are inherently more difficult to prepare, less stable and more conformationally variable. This challenge remains, even though the field of RNP structural biology has seen enormous recent progress with the structure determination of the ribosome and its subunits (Yusupov *et al.*, 2001; Wimberly *et al.*, 2000; Ban *et al.*, 2000; Harms *et al.*, 2001). The objective of this paper is to show that success in obtaining diffraction-quality crystals of an RNP complex can be realised if the effort is made to prepare a series of related complexes that systematically vary in one of the components. This approach takes advantage of modern methods of expression, synthesis and purification and of the fact that RNP complexes are comprised of at least two

components, so that preparation of n variants of each component allows n^2 complexes to be screened. Because all combinations are utilized, we term this approach 'combinatorial crystallization'.

It is well known that a critical factor in obtaining suitable RNP crystals is the specific RNA construct or protein-sequence variant used to make the RNP. To crystallize a 58-nucleotide rRNA–L11 complex, an rRNA mutant was used that stabilizes the tertiary structure (Conn *et al.*, 1998, 1999); the rRNA from another species also contains this stabilizing base triple (Wimberly *et al.*, 1999). Crystallization of the polyA-binding protein–poly(A)$_8$ complex required screening five lengths of poly(A)$_n$ (Deo *et al.*, 1999). Several constructs of the Sx1 protein were screened and one was further mutated to obtain crystals of the *tra* mRNA complex (Handa *et al.*, 1999). Diffraction-quality crystals of a U1A protein–snRNA complex were obtained only after 23 RNA constructs, 38 protein mutants and 90 unique complexes were prepared (Oubridge *et al.*, 1995). The hepatitis δ virus (HDV; Ferré-D'Amaré & Doudna, 2000) and hairpin (Rupert & Ferré-D'Amaré, 2001) ribozymes were crystallized by first introducing the U1A protein-binding domain in order to prepare RNPs. Diffraction-quality crystals of a 104-nucleotide domain of 16S rRNA in complex with three ribosomal proteins were obtained only after biochemical characterization of several rRNA constructs (Agalarov *et al.*, 2000). These examples demonstrate that definition of a stable construct in both the protein and RNA is required to obtain RNP crystals, but that in addition to this, the introduction of single-residue variations can also be critical. Although several intensive crystallization screens have been carried out, the effect of systematic variation in the components is seldom reported.

It is common as well as convenient to vary the length and sequence of double helices to improve the quality of nucleic acid-containing crystals. The sequence of the terminal nucleotides influences stacking and the length of the helix dictates the distance between neighboring molecules in the crystal lattice. Inclusion of overhangs at the ends of helices can further affect intermolecular base pairing or interactions within the major or minor groove (Cruse *et al.*, 1994). Helix engineering has been applied successfully for the crystallization of protein–DNA complexes (Aggarwal *et al.*, 1988; Jordan *et al.*, 1985; Schultz *et al.*, 1990), small structured RNAs (Anderson *et al.*, 1996), the hammerhead ribozyme (Scott *et al.*, 1995) and the U1A–snRNA complex (Oubridge *et al.*, 1995).

An efficient strategy for generating a pool of duplexes was devised for the crystallization of the CAP protein complexed with its DNA target (Schultz *et al.*, 1990). In this approach, five complementary DNA strands were annealed in a combinatorial fashion, resulting in 25 DNA duplexes that differed in length and the presence or absence of overhanging nucleotides at either end of the helix. Although this approach only allowed the generation of DNAs with symmetrical ends, it proved successful in generating high-quality crystals of the CAP–DNA complex. A complete combinatorial screen of paired oligonucleotides was used to obtain crystals of an RNA substrate–DNA enzyme complex (Nowakowski *et al.*, 1999). In this case, the 5′ and 3′ ends of the DNA and RNA components were systematically varied by +0, +1 or +2 residues, creating 3^4 or 81 unique duplexes based on the parent sequence (nine RNAs and nine DNAs). In the study reported here, the latter approach has been employed to facilitate crystallization of an RNP. Both strands of an RNA duplex have been systematically varied by +0 or +1 residue at both the 5′ and 3′ ends to generate 16 unique duplexes and each of these has been complexed to the target protein, generating 16 RNPs. Thus, the sequence variation occurs at the ends of the RNA duplex while the protein component of the RNP has not been mutated; however, the protein, L30, has been expressed as a fusion protein with maltose-binding protein (MBP).

L30 interacts with its own RNA transcript to regulate splicing and translation by binding to a helix–loop–helix structure in its own pre-mRNA and mRNA (Dabeva & Warner, 1993; Vilardell & Warner, 1994; Li *et al.*, 1995). Inhibition of spliceosome assembly occurs because L30-bound pre-mRNAs form abortive splicing complexes with the U1 snRNP (Vilardell & Warner, 1994; Vilardell, Chartrand *et al.*, 2000). L30 also inhibits translation of mature mRNA (Dabeva & Warner, 1993) by preventing the mRNA from associating with 40S subunits (Vilardell, Chartrand *et al.*, 2000). L30 autoregulation of splicing and translation is critical for biological fitness because it ensures that all ribosomal proteins are produced in equimolar amounts (Li *et al.*, 1996).

The atomic resolution structure of the L30–RNA complex was solved using NMR techniques (Mao *et al.*, 1999; Mao & Williamson, 1999*a*,*b*). NMR assignments for the L30–RNA complex were difficult to obtain owing to spectral complexity and broadened resonances. RNA mutants and $^{13}C/^{15}N$ labeling were necessary in order to determine the correct assignments for nucleotides in the internal loop (Mao & Williamson, 1999*a*). The binding of the RNA to the protein is coupled with local folding, since the purine-rich internal loop is unstructured in the absence of L30 but structured in its presence (Mao & Williamson, 1999*a*). The region containing residues 72–88 is largely disordered in the free L30 protein, but becomes well structured when bound to RNA (Mao & Williamson, 1999*b*). The RNA is sharply bent when bound to L30 (Mao *et al.*, 1999), supporting the suggested kinking function of L30 in the 60S subunit (Vilardell, Yu *et al.*, 2000). There is a 130° angle between the two helices bent at the asymmetric internal loop (Mao *et al.*, 1999). The RNA–protein contacts are almost exclusively located between the internal loop of the RNA and loops on one face of the protein.

Recently, structures of two proteins homologous to L30, the 15.5 kDa spliceosomal protein and L7Ae, have been solved bound to their respective RNA targets (Vidovic *et al.*, 2000; Ban *et al.*, 2000). The RNA targets of these three proteins are all purine-rich internal loops, which have been shown to belong to the K-turn RNA secondary-structure motif (Klein *et al.*, 2001). The hydrogen-bonding pattern observed in the L30–RNA structure, however, is different to that of the consensus K-turn. Low NOE density in the internal loop as well as difficulty in making assignments in this region provided the

Table 1
RNA oligonucleotides used for crystallization.

	Name	MW (Da)	Extinction coefficients (260 nm) (mM^{-1} cm^{-1})
Top strands			
GACCGGAGUGUC	T-1	3922	131.8
CGACCGGAGUGUC	T-2	4228	139.2
GACCGGAGUGUCC	T-3	4228	139.2
CGACCGGAGUGUCC	T-4	4534	146.6
Bottom strands			
GACGCAGAGAUGGUC	B-1	4928	174.4
GGACGCAGAGAUGGUC	B-2	5274	186.2
GACGCAGAGAUGGUCG	B-3	5274	186.2
GGACGCAGAGAUGGUCG	B-4	5620	198.0

impetus for undertaking an X-ray crystallography study of the L30–RNA complex.

2. Materials and methods

2.1. RNA oligonucleotides

Eight synthetic RNA oligonucleotides of length 12–17 residues (Table 1) were obtained from Dharmacon Research, Inc., Lafayette, CO, USA (1.0 µmol syntheses, yield $\simeq$ 0.8 µmol), deprotected according to the manufacturer's instructions and purified by preparative gel electrophoresis (20% acrylamide) in 8 *M* urea. The RNAs were eluted in 300 µ*M* sodium acetate pH 5.2, 1 m*M* EDTA (yield $\sim$0.6 µmol), desalted by reverse-phase chromatography (Sep-Pak C-18 column, yield $\simeq$ 0.5 µmol) and buffer exchanged in 10 m*M* sodium phosphate pH 7.2 (final concentration 1 m*M*, yield $\simeq$ 0.4 µmol). The nomenclature, MW and extinction coefficients for the nucleotides are summarized in Table 1.

2.2. MBP-L30 protein

The maltose-binding protein-L30 fusion protein was expressed and purified as previously described (Mao & Williamson, 1999*b*). The clarified supernatant from cell lysate was applied to a CM-cellulose 650M column (Toyopearl) and MBP-L30 was eluted using a 0.2–0.5 *M* NaCl gradient. MBP-L30 was further purified by amylose affinity chromatography (New England Biolabs) and eluted at 30–40 µ*M* maltose. The yield of MBP-L30 was $\sim$5 mg per litre of culture. The MW of MBP-L30 is 52 963 Da (*Escherichia coli* MBP with C-terminal factor Xa cleavage site linked to *S. cerevisiae* L30) and the extinction coefficient at 280 nm is $\sim$72 400 $M^{-1}cm^{-1}$.

2.3. MBP-L30–RNA complexes

The concentration of MBP-L30 relative to the RNA oligonucleotides was determined by band-shift assay using equal concentrations of complementary oligonucleotides and varying concentrations of MBP-L30. RNA duplexes were formed by annealing at 368 K in 20 m*M* Tris–HCl pH 7.0, 200 m*M* KCl, followed by cooling to 308 K in 30 min. Complexes were prepared by adding 80 nmol of duplex to 67 nmol of MBP-L30 at RT for 20 min at final concentrations of <50 µ*M*. Complexes were dialyzed against 10 m*M* potassium phosphate pH 6.0, 1 m*M* maltose, 0.02% sodium azide, assayed by native gel electrophoresis, concentrated to 158 µ*M* and stored at 277 K. Molecular weights for the MBP-L30–RNA complexes range from 61 813 to 63 117 Da.

2.4. Crystallization screens

Crystallization was carried out by vapor diffusion using the hanging-drop method. Two 24-well plates were set up with each of the 16 complexes using conditions Nos. 1–24 of the Hampton Natrix screen for the first plate and conditions Nos. 1–24 of Hampton Crystal Screen 2 (Hampton Research) for the second plate. Three drops were set up over each 0.5 ml reservoir at MBP-L30–RNA complex/buffer volumes of 0.5/0.5, 0.5/1.0 and 1.0/0.5 µl. Therefore, the 16 complexes were screened against 48 reservoir solutions at three concentrations, yielding 2304 unique conditions. The plates were maintained at 295.5 K for at least two weeks before evaluation of crystal growth. The crystals were examined with a binocular microscope (57×) equipped with a polarizable filter.

2.5. X-ray diffraction screening

The diffraction properties of the crystals were evaluated at 100 K using a Rigaku FRD X-ray generator equipped with Osmic focusing mirrors and a 30 cm MAR Research image-

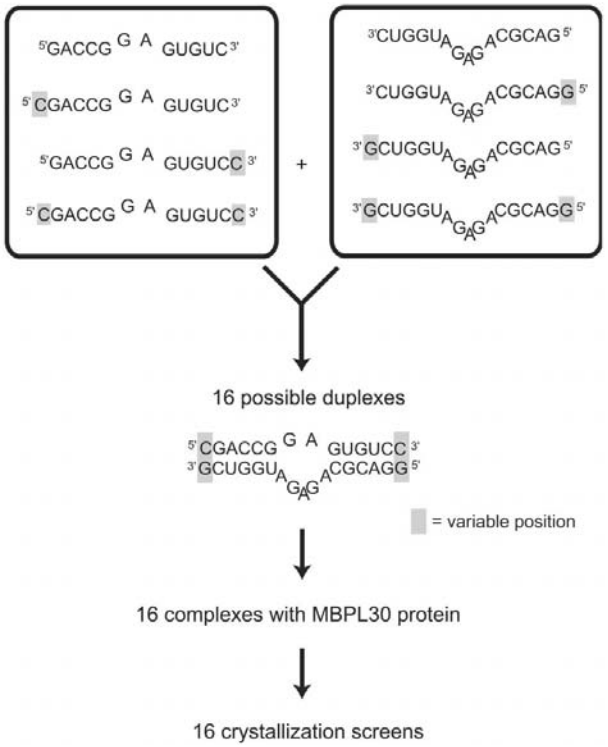

Figure 1
Scheme for combinatorial screening of paired oligonucleotides and crystallization strategy for the MBP-L30–RNA complex. Eight sequences of RNA oligonucleotides were mixed to form 16 unique duplexes containing the secondary structure of the internal loop recognized by L30.

plate scanner or using a Bruker SRA X-ray generator equipped with a graphite monochromator and a 34.5 cm MAR Research image-plate scanner (Cu $K\alpha$ radiation, $\lambda = 1.54$ Å). Cryoprotectants were prepared from the reservoir solutions by the addition of suitable cryosolvents determined by first testing the freezing behavior of the mother-liquor solutions in the absence of the crystals. Data were collected at the Stanford Synchrotron Radiation Laboratory on beamline 11-1 ($\lambda = 0.98$ Å) at 100 K using an ADSC Quantum 4 CCD detector and were processed using the *CCP*4 programs *MOSFLM* and *SCALA* (Collaborative Computational Project, Number 4, 1994).

3. Results

3.1. Design of the combinatorial screen

The concept of the combinatorial screen of paired strands is illustrated in Fig. 1. In this experiment, we have adopted the approach used to crystallize a DNA–RNA complex (Nowakowski *et al.*, 1999) by applying the use of paired oligonucleotides, only in this case to two RNA strands. At the same time, we have extended the method to the crystallization of an RNP. Each of the paired strands is capable of forming the internal loop recognized by L30, which is expressed as a fusion protein with maltose-binding protein. Each RNA strand is synthesized with a different length and purified individually. The RNA strands are mixed in all possible combinations and the resulting unique duplexes are mixed with the protein to generate the same number of unique RNP complexes. Each complex is then subjected to a standard set of crystallization conditions and evaluated for crystal formation. Because the crystallization solutions are identical for the various complexes, this method searches for the best combination of paired strands rather than the best crystallization conditions. Once the best complexes or sequences have been identified, the crystallization conditions can be optimized to increase the size and quality of the crystals.

The sequences and nomenclature for the RNA oligonucleotides used for the 'top' and 'bottom' strands are given in Table 1. Each combination of top and bottom strands forms a duplex with five or six base pairs on either side of the asymmetric seven-nucleotide internal loop (Fig. 1). For this experiment, an additional C residue was added to the 5′ or 3′ end of the top-strand sequence and an additional G residue was added to the 5′ or 3′ end of the bottom-strand sequence, generating eight total sequences while preserving Watson–Crick complementarity. All combinations of the two strands generate 16 unique complexes: four with two blunt ends, eight with one blunt end and one overhanging residue, and four with two overhanging residues (Fig. 2). The sequence could in principle be varied, but the number of resulting complexes would increase enormously, making it difficult to complete the screen. Previous experience suggests that the length of double-helical stems and the presence of overhanging residues alone generate sufficient diversity to provide high-quality crystals (*e.g.* Schultz *et al.*, 1990; Nowakowski *et al.*, 1999). An advantage of the combinatorial mixing strategy is the reduced time needed to synthesize individual strands. Mixing n sequences of one strand with m sequences of the other generates a total of $n \times m$ complexes, but requires synthesis and purification of only $n + m$ strands.

Each oligonucleotide duplex was designated by two numbers, T-1 to T-4 for the top strands and B-1 to B-4 for the bottom strands (Table 1; Fig. 2). Therefore, the T-3/B-3 duplex contains an overhanging C at the 3′ end of the top strand and an overhanging G at the 3′ end of the bottom strand. All RNA molecules were synthesized chemically and combined in a 1:1 molar ratio to make the 16 unique complexes and each of these was combined with the MBP-L30 fusion protein in an 80:67 molar ratio (§2). A single synthesis of each oligonucleotide on the 1.0 µmol scale and a single preparation of MBP-L30 (∼60 mg protein) provided a sufficient amount of material to carry out the screen.

3.2. Crystallization screening

Each MBP-L30–RNA complex was evaluated for its ability to crystallize using a set of 48 reservoir solutions. Previous experience indicates that the choice of molecule is more important than the exact condition, *i.e.* favorable sequences tend to crystallize under a variety of conditions (Nowakowski *et al.*, 1999). Hence, the 48 solutions used, *i.e.* combinations of precipitant, buffer and additive, were solutions Nos. 1–24 of the Hampton Natrix screen, customized for nucleic acids, and solutions Nos. 1–24 of Hampton Crystal Screen 2, customized for proteins. The choice of these screening kits and of which of the 48 solutions per kit to use was otherwise arbitrary. Once promising results were obtained for the complex containing the T-3/B-2 duplex (see below), additional screening was performed with this RNP.

3.3. Crystallization results

The results of the screen are summarized in histograms in Fig. 3. Each of the 2304 drops were scored on a arbitrary scale of 1–9 where '6' represents spherulites, '7' microcrystals, '8' crystals in clusters or many small crystals and '9' single crys-

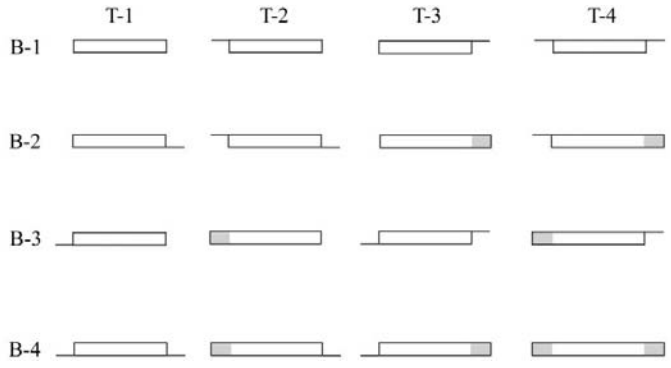

Figure 2
Schematic representation of the positions of overhanging residues and blunt ends in the 16 unique duplexes formed by pairing oligonucleotides T-1 to T-4 with B-1 to B-4. Shaded areas represent variable positions that form complementary base pairs in eight instances.

tals. Although this evaluation of the results is somewhat subjective, the appearance of single crystals is unmistakable. An interesting feature of the data is that for any complex with drops scoring '9' the histogram is also bimodal. This result is consistent with the idea that certain sequences are favored for crystallization and these sequences will form crystals under a variety of conditions, eliminating the need for further screening to initially identify large single crystals. Thirty, or 1.32%, of the drops, scored '9', *i.e.* yielded single crystals that could be evaluated directly for X-ray diffraction. Consistent with the overall bimodal distribution of the yield of crystals, 1.54% of drops scored '8', 1.45% scored '7' and 2.24% scored '6', while the next best score, '5', occurred for 10.7% of the drops (Fig. 3).

Two additional ways to evaluate the results of the screen are with respect to the solutions (conditions) and with respect to the sequences. As observed in combinatorial crystallization of a DNA–RNA complex (Nowakowski *et al.*, 1999), certain solutions are much more effective in producing crystals. For the 65 drops scoring '8' or '9', Hampton Crystal Screen 2 conditions Nos. 1–24 yielded 11 drops and Natrix conditions Nos. 1–24 yielded 54 drops. The latter screen, optimized for nucleic acids, was clearly favored for this RNP, even though the mass of protein was in large excess over that of the RNA. Nevertheless, both screens showed the effectiveness of a very limited number of solutions. For the Hampton Crystal Screen 2, just seven of the 24 conditions yielded any crystals at all, with two of the solutions (Nos. 7 and 13) yielding crystals in seven drops. For the Hampton Natrix screen, over half of the solutions used (13 out of 24) yielded crystals, but just three of these (Nos. 12, 13 and 19) yielded crystals in 33 drops. Although it cannot be known in advance which solutions might be favored for a particular molecule, the results demonstrate that only a limited number of solutions need to be employed.

It is also of interest to evaluate the screen in terms of the number of overhanging residues, especially as this variation is the basis of the experimental design. Complexes containing RNA duplex with two blunt ends yielded on average 12 drops scoring '6' or better. Complexes containing RNA duplex with one overhang averaged nine drops scoring '6' or better, while complexes with two overhangs averaged only six drops scoring '6' or better. Thus, the complexes containing blunt ends are favored in lattice formation (Fig. 2). Based on these results, limited additional screening was performed with one blunt-end complex ('T-3/B-2') and a different set of conditions.

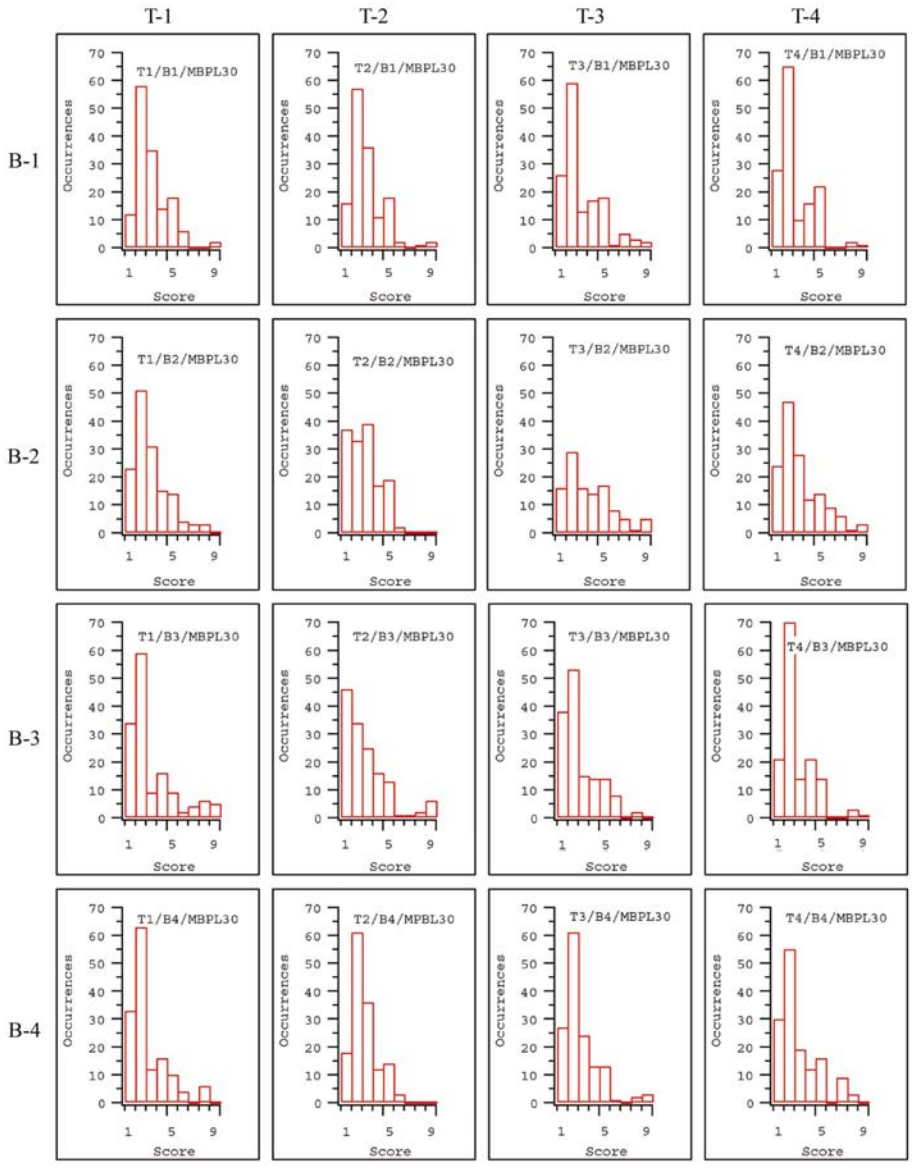

Figure 3
Results of the crystallization screen. Each histogram in the matrix corresponds to a particular MBP-L30–RNA complex comprised of a 'top' (T-1 to T-4) and a 'bottom' (B-1 to B-4) oligonucleotide arranged as in Fig. 2. For each of the 16 complexes the results of the screen are scored on a scale of 1–9 for 144 drops (48 conditions each at three variable concentrations). The score was defined as follows: 1, clear; 2, trace precipitate; 3, oil drops; 4, light precipitate; 5, precipitate; 6, spherulites; 7, microcrystals; 8, clusters of crystals or many small crystals; 9, single crystals. Note that for T-3/B-2/MBPL30, 32 of the 144 conditions were not screened owing to sample limitations.

3.4. Diffraction screening

The identification of diffraction-quality crystals is a two-step process, first finding crystallization conditions and then evaluating crystal forms for cryoprotectants and diffraction resolution. Crystals that were large enough

Table 2
Summary of diffraction results from the combinatorial crystallization screen.

Results are arranged in order of decreasing resolution.

MBP-L30 complex	Screen†	Condition No.	Drop ratio‡ (µl)	Crystal size§ (µm)	Diffraction limit¶ (Å)
T-1/B-3	Natrix	6	1:0.5	200	3.8
T-3/B-2	Natrix	11	1:1	500	3.9
T-3/B-2	Natrix	23	0.5:0.5	250	3.9
T-3/B-2	HCS2	1	0.5:0.5	300	6.4
T-3/B-1	Natrix	13	1:0.5	500	7
T-3/B-1	Natrix	13	1:1	900	7
T-1/B-3	Natrix	1	1:0.5	200	7
T-4/B-2	Natrix	18	0.5:0.5	165	7.5
T-2/B-3	Natrix	13	0.5:0.5	165	7.5
T-2/B-1	Natrix	19	0.5:0.5	300	9.6
T-3/B-1	Natrix	12	1:0.5	370	10
T-4/B-1	Natrix	19	0.5:1	435	11
T-4/B-1	Natrix	19	0.5:0.5	Aggregate	None
T-2/B-3	Natrix	12	0.5:1	120	None
T-3/B-1	Natrix	6	1:0.5	60	None
T-3/B-1	Natrix	19	0.5:0.5	75	None
T-1/B-3	HCS2	14	0.5:1	Small	None

† HCS2, Hampton Crystal Screen 2; Natrix, Hampton Natrix Screen. ‡ The volume of reservoir:sample solutions in the hanging drops. § Values refer to the maximum dimension. ¶ Maximum resolution of reflections observed in test exposures of $\leq$10 min and $\leq 1^\circ$ oscillation.

to be mounted directly from the drops in the combinatorial screen were evaluated using an in-house X-ray source as described in §2 and summarized in Table 2. In most cases the choice of cryoprotectant could be determined by testing with the reservoir solution, but in some cases it was necessary to use several crystals in order to confirm their diffraction characteristics.

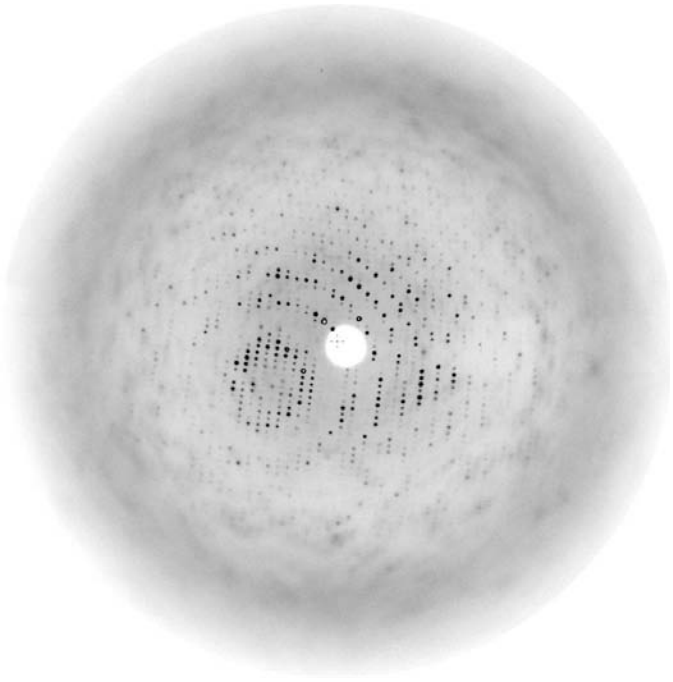

Figure 4
X-ray diffraction pattern of a T-3/B-2/MBPL30 complex crystal found in the initial screen. Data were collected at 100 K using a Rigaku FRD X-ray generator equipped with Osmic focusing mirrors and a 30 cm MAR Research image-plate scanner (Cu $K\alpha$ radiation, $\lambda = 1.54$ Å). This crystal diffracts to 3.9 Å.

Table 3
Data collection for the MBP-L30–RNA complex.

Values in parentheses are for the last resolution shell.

Crystal	T-3/B-2 complex
Space group	$P4_12_12$ ($P4_32_12$)
Unit-cell parameters (Å)	136.1 × 136.1 × 124.0
Total observations	137877
Independent reflections	36911
Resolution range (Å)	31–3.28 (3.37–3.28)
Completeness (%)	98.7 (98.6)
$R_{sym}(I)$	0.086 (0.439)
$I/\sigma(I)$	6.1 (1.8)

Sixty five drops from the combinatorial screen contained single crystals (score of '9') or many small crystals or clusters of crystals (score of '8'). Table 2 shows the diffraction limit of 17 of these crystals. Only two complexes from three conditions yielded promising diffraction, with those from the T-1/B-3 and T-3/B-2 duplexes diffracting beyond 4.0 Å resolution (Fig. 4). Based on this rate of success, the yield of diffraction-quality crystals is low. However, it should be emphasized that many drops scoring '8' or '9' were not evaluated and that further experiments could easily be performed to pursue these crystal forms. Additional fine screening of the T-3/B-2 complex yielded the best crystal form, which is being used for structure determination (Jeff A. Chao & James R. Williamson, unpublished results). For each crystal form that diffracts to high resolution, it is important to verify that the crystals contain both RNA and protein and, in fact, crystals from T-4/B-3/MBPL30 (diffracting to 2.9 Å) contained only protein.

3.5. Data collection

Optimal conditions for crystallization of the T-3/B-2 complex are 1 *M* lithium sulfate, 0.05 *M* sodium cacodylate pH 6.0, 0.01 *M* magnesium chloride and 0.5% Jeffamine. A single crystal of $\sim 0.4 \times 0.4 \times 0.4$ mm in size was frozen at 100 K using a cryoprotectant consisting of 25%(*v*/*v*) glycerol mixed with the reservoir solution and was used to collect a complete data set to 3.28 Å resolution (Table 3). It was possible to solve the structure of this crystal form using molecular replacement with the known structure of MBP. The T-3/B-2/MBPL30 complex structure was refined at 3.28 Å resolution and shown to contain RNA. In this construct, one of the helices stacks on a symmetry-related copy of itself (Jeff A. Chao & James R. Williamson, unpublished results).

4. Discussion

The precise composition of the biological macromolecule, an RNA–protein complex in this case, is critical to obtaining good crystals. Fortunately, small variations make a large difference in crystallization behavior. Therefore, it is worth the investment in time and effort to systematically explore small

variations in the molecular composition in order to obtain multiple crystal forms, increasing the probability of finding a crystal form that is amenable to freezing conditions and that diffracts well. In the case where the target molecule or complex contains a nucleic acid, it is particularly convenient to vary the sequence by chemical synthesis. Moreover, if the target nucleic acid can be assembled from more than one component, then a combinatorial approach can be used, dramatically increasing the number of unique complexes. As it is the precise composition that is critical, this approach increases the likelihood that diffraction-quality crystals can be discovered.

To our knowledge, this represents the first use of RNA in a systematic combinatorial crystallization screen of different RNA duplexes in an RNP. As has been observed in the crystallization of DNA–protein complexes (Jordan *et al.*, 1985; Aggarwal *et al.*, 1988; Schultz *et al.*, 1990), the results demonstrate that the length of the duplex and the presence of overhanging nucleotides have a dramatic effect on crystal formation. In this case, we have employed a fusion protein, initially to improve solubility of the basic RNA-binding protein L30; in practice, this provides a means of increasing the mass of protein relative to RNA, favoring protein–protein lattice contacts. Similarly, a non-specifically bound protein has been used to facilitate crystallization of DNA (Coté *et al.*, 2000). The presence of the 40 kDa maltose-binding protein, compared with the 10 kDa L30 and 10 kDa RNA, further aids in molecular-replacement calculations, while the presence of bound maltose provides an internal check of the structure determination (Chao *et al.*, 2003). In spite of the relative masses of the protein and RNA, however, the exact sequence of the RNA remains critical in obtaining good crystals.

The results show that given a sufficient diversity of complexes, diffraction-quality crystals can be obtained while conducting limited screening of conditions. It is of interest to compare the rate of success for different types of complexes. In this experiment, it was necessary to screen 16 MBP-L30–RNA complexes. Successful crystallization of the U1A protein–RNA complex required an unusual effort, entailing preparation of 23 RNA constructs and 38 protein mutants in 90 unique complexes (Oubridge *et al.*, 1995). Subsequently, the U1A protein and its mutants have been used in combination with the HDV (Ferré-D'Amaré & Doudna, 2000) and hairpin (Rupert & Ferré-D'Amaré, 2001) ribozymes, each engineered to contain the U1A protein-binding site, to generate eight and 21 unique RNPs, respectively, for crystallization. A comparable number (25) of CAP protein–DNA complexes were systematically screened to obtain diffraction-quality crystals (Schultz *et al.*, 1990). By comparison, 31 related group II intron RNA constructs were screened to obtain crystals diffracting to 3.5 Å (Ferré-D'Amaré *et al.*, 1998) and 81 DNA enzyme–RNA substrate complexes yielded two crystal forms diffracting to 2.8 Å resolution (Nowakowski *et al.*, 1999). These results indicate that the crystallization of RNP, DNA–protein, RNA and DNA–RNA complexes is of comparable difficulty, requiring that on average ∼30–40 complexes of unique sequence composition be screened. Given the necessity of screening many different sequences, a combinatorial approach as presented here is clearly advantageous.

The authors wish to thank the staff at the Stanford Synchrotron Radiation Laboratory for their generous assistance. This research was supported by ARCS and the Skaggs Institute for Chemical Biology at The Scripps Research Institute and NIH grant GM-53320 to JRW.

References

Agalarov, S. C., Prasad, G. S., Funke, P. M., Stout, C. D. & Williamson, J. R. (2000). *Science*, **288**, 107–112.
Aggarwal, A. K., Rodgers, D. W., Drottar, M., Ptashne, M. & Harrison, S. C. (1988). *Science*, **242**, 899–907.
Anderson, A. C., Earp, B. E. & Frederick, C. A. (1996). *J. Mol. Biol.* **259**, 696–703.
Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. (2000). *Science*, **289**, 905–930.
Chao, J. A., Prasad, G. S., White, S. A., Stout, C. D. & Williamson, J. R. (2003). In the press.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Conn, G. L., Draper, D. E., Lattman, E. E. & Gittis, A. G. (1999). *Science*, **284**, 1171–1174.
Conn, G. L., Gutell, R. R. & Draper, D. E. (1998). *Biochemistry*, **37**, 11980–11988.
Coté, M. L., Yohannan, S. J. & Georgiadis, M. M. (2000). *Acta Cryst.* D**56**, 1120–1131.
Cruse, W. B. T., Saludjian, P., Biala, E., Strazewski, P., Prange, T. & Kennard, O. (1994). *Proc. Natl Acad. Sci. USA*, **91**, 4160–4164.
Dabeva, M. D. & Warner, J. R. (1993). *J. Biol. Chem.* **268**, 19669–19674.
Deo, R. C., Bonanno, J. B., Sonenberg, N. & Burley, S. K. (1999). *Cell*, **98**, 835–845.
Ferré-D'Amaré, A. R. & Doudna, J. A. (2000). *J. Mol. Biol.* **295**, 541–556.
Ferré-D'Amaré, A. R., Zhou, K. & Doudna, J. A. (1998). *J. Mol. Biol.* **279**, 621–631.
Handa, N., Nureki, O., Kurimoto, K., Kim, I., Sakamoto, H., Shimura, Y., Muto, Y. & Yokoyama, S. (1999). *Nature (London)*, **398**, 579–585.
Harms, J., Schluenzen, F., Zarivach, R., Bashan, A., Gat, S., Agmon, I., Bartels, H., Franceschi, F. & Yonath, A. (2001). *Cell*, **107**, 679–688.
Jordan, S. R., Whitcombe, T. V., Berg, J. M. & Pabo, C. O. (1985). *Science*, **230**, 1383–1385.
Klein, D. J., Schmeing, T. M., Moore, P. B. & Steitz, T. A. (2001). *EMBO J.* **20**, 4214–4221.
Li, B., Vilardell, J. & Warner, J. R. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 1596–1600.
Li, H., Dalal, S., Kohler, J., Vilardell, J. & White, S. A. (1995). *J. Mol. Biol.* **250**, 447–459.
Mao, H., White, S. A. & Williamson, J. R. (1999). *Nature Struct. Biol.* **6**, 1139–1147.
Mao, H. & Williamson, J. R. (1999*a*). *Nucleic Acids Res.* **27**, 4059–4070.
Mao, H. & Williamson, J. R. (1999*b*). *J. Mol. Biol.* **292**, 345–359.
Nowakowski, J., Shim, P. J., Joyce, G. F. & Stout, C. D. (1999). *Acta Cryst.* D**55**, 1885–1892.
Oubridge, C., Nobutoshi, I., Teo, C.-H., Fearnley, I. & Nagai, K. (1995). *J. Mol. Biol.* **249**, 409–423.
Rupert, P. B. & Ferré-D'Amaré, A. R. (2001). *Nature (London)*, **410**, 780–786.
Schultz, S. C., Shields, G. C. & Steitz, T. A. (1990). *J. Mol. Biol.* **213**, 159–166.

Scott, W. G., Finch, J. T., Grenfell, R., Fogg, J., Smith, T., Gait, M. J. & Klug, A. (1995). *J. Mol. Biol.* **250**, 327–332.
Vidovic, I., Nottrott, S., Hartmuth, K., Lührmann, R. & Ficner, R. (2000). *Mol. Cell*, **6**, 1331–1342.
Vilardell, J., Chartrand, P., Singer, R. H. & Warner, J. R. (2000). *RNA*, **6**, 1773–1780.
Vilardell, J. & Warner, J. R. (1994). *Genes Dev.* **8**, 211–220.
Vilardell, J., Yu, S. J. & Warner, J. R. (2000). *Mol. Cell*, **5**, 761–766.
Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Morgan-Warren, R. J., Carter, A. P., Vonrhein, C., Hartsch, T. & Ramakrishnan, V. (2000). *Nature (London)*, **407**, 327–339.
Wimberly, B. T., Guymon, R., McCutcheon, J. P., White, S. W. & Ramakrishnan, V. (1999). *Cell*, **97**, 491–502.
Yusupov, M. M., Yusupova, G. Zh., Baucom, A., Lieberman, K., Earnest, T. N., Cate, J. H. D. & Noller, H. F. (2001). *Science*, **292**, 883–896.