

Crystal Structure of the 30 S Ribosomal Subunit from *Thermus thermophilus*: Purification, Crystallization and Structure Determination

William M. Clemons Jr^{1,2}, Ditlev E. Brodersen¹, John P. McCutcheon²
Joanna L. C. May², Andrew P. Carter¹, Robert J. Morgan-Warren¹
Brian T. Wimberly^{1,2} and V. Ramakrishnan^{1,2*}

¹MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

²Department of Biochemistry, University of Utah School of Medicine, Salt Lake City UT 84132, USA

We describe the crystallization and structure determination of the 30 S ribosomal subunit from *Thermus thermophilus*. Previous reports of crystals that diffracted to 10 Å resolution were used as a starting point to improve the quality of the diffraction. Eventually, ideas such as the addition of substrates or factors to eliminate conformational heterogeneity proved less important than attention to detail in yielding crystals that diffracted beyond 3 Å resolution. Despite improvements in technology and methodology in the last decade, the structure determination of the 30 S subunit presented some very challenging technical problems because of the size of the asymmetric unit, crystal variability and sensitivity to radiation damage. Some steps that were useful for determination of the atomic structure were: the use of anomalous scattering from the LIII edges of osmium and lutetium to obtain the necessary phasing signal; the use of tunable, third-generation synchrotron sources to obtain data of reasonable quality at high resolution; collection of derivative data precisely about a mirror plane to preserve small anomalous differences between Bijvoet mates despite extensive radiation damage and multi-crystal scaling; the pre-screening of crystals to ensure quality, isomorphism and the efficient use of scarce third-generation synchrotron time; pre-incubation of crystals in cobalt hexaammine to ensure isomorphism with other derivatives; and finally, the placement of proteins whose structures had been previously solved in isolation, in conjunction with biochemical data on protein-RNA interactions, to map out the architecture of the 30 S subunit prior to the construction of a detailed atomic-resolution model.

© 2001 Academic Press

*Corresponding author

Keywords: 30 S; ribosome; crystallography; anomalous scattering; RNA

Introduction

The ribosome is the large macromolecular enzyme that performs messenger RNA-directed protein synthesis using aminoacylated transfer RNA (aa-tRNA) as its substrate. In all organisms,

the ribosome consists of a large and a small subunit, which in bacteria are referred to as 50 S and 30 S, respectively, according to their rate of sedimentation. The ribosome has three binding sites for tRNA, designated the A (for acceptor- or aminoacyl-), P (peptidyl) and E (exit) sites which are located in part on both subunits. The aminoacyl- or acceptor stem regions of the tRNA molecules bind in the 50 S subunit, where peptide bond formation, which involves the addition of an amino acid to the growing polypeptide chain, is catalyzed. The small subunit is involved in binding mRNA and the anticodon stem loops of the tRNAs. Its role is to discriminate between cognate and non-cognate aa-tRNAs in the A site during translation of the genetic code. Both subunits act in

Present address: V. Ramakrishnan, MRC Laboratory of Molecular Biology, Hills Road, Cambridge, CB2 2QH, UK.

Abbreviations used: aa, aminoacylated; MPD, 2-methyl-2,4-pentanediol; cryo-EM, cryo-electron microscopy; MAD, multiple wavelength anomalous diffraction.

E-mail address of the corresponding author: ramak@mrc-lmb.cam.ac.uk

concert during the subsequent process of translocation, which involves the movement of mRNA and all three tRNAs by precisely one codon with respect to the ribosome. The entire process also depends on several extrinsic protein factors and the hydrolysis of GTP. An overview of the current state of protein synthesis can be found in a recent symposium volume.¹

Since the discovery of the ribosome several decades ago, virtually every aspect of its function has been studied both biochemically and genetically. However, until recently, the absence of high-resolution structures of the ribosome or its subunits has hindered progress towards arriving at detailed mechanistic explanations for the various steps in translation.

Over the last two decades, there have been different and complementary approaches to crystallographic studies of ribosomal components. We and others concentrated initially on determination of structures of individual proteins (reviews^{2,3}), or binary complexes with fragments of ribosomal RNA,⁴⁻⁶ whereas others have pioneered the approach of crystallization of whole subunits or ribosomes as described below.

Three-dimensional crystals of the 50 S subunit from *Bacillus stearothermophilus* suitable for X-ray analysis were first obtained by Yonath, Wittmann and their co-workers.⁷ Eventually, Yonath and co-workers went on to obtain crystals of 50 S subunits from *Haloarcula marismortui* that diffracted to beyond 3 Å resolution, showing that the determination of the atomic structure of the subunit was possible in principle.⁸

Crystals of the 30 S subunit^{9,10} and of entire 70 S ribosomes¹⁰ from *Thermus thermophilus* were first reported by the Puschino group using 2-methyl-2,4-pentanediol (MPD) as the precipitant. Soon afterwards, Yonath and co-workers reported similar crystals of the 30 S subunit in a mixture of ethylbutanol and ethanol,¹¹ but this crystallization condition appears not to have been pursued further. The following year, both the Puschino and Yonath groups showed that the MPD form of the 30 S crystals diffracted to around 10-12 Å.^{12,13} By 1995, the reported diffraction limit of these crystals had improved only modestly, to about 7-8 Å resolution.¹⁴ Around this time, it was becoming clear from cryo-electron microscopy (cryo-EM) that the 30 S is conformationally flexible. For example, the orientation of the upper third of the subunit (termed the "head") relative to the lower two-thirds or "body" of the subunit is different in isolated 30 S subunits and 70 S ribosomes.¹⁵⁻¹⁸ It therefore seemed likely, as has been shown for many classical enzymes, that the binding of substrates or factors to the 30 S subunit would stabilize a unique conformation and hence lead to crystals with improved diffraction.

When ribosomal subunits were first crystallized, it was not clear that standard crystallographic methods would prove sufficient to determine their structure because the size of the asymmetric unit

was over an order of magnitude larger than anything that had previously been solved, except for viruses that had significant non-crystallographic symmetry. However, the determination of structures as large as F1-ATPase with only a small number of heavy-atom sites showed that a much smaller signal is now detectable than was previously thought possible, especially if non-isomorphism could be avoided.¹⁹ In addition, a number of parallel technical improvements in crystallography made it possible to tackle such large and complex problems. These include the development of high-intensity, tunable X-ray beams from third generation synchrotron radiation sources, large and sensitive area detectors, improved software for data integration and analysis, and more powerful computing.^{20,21} Another important advance was the advent of cryocrystallography (reviews^{22,23}). In particular, Hope²⁴ showed that flash cooling of protein crystals to liquid nitrogen temperatures was generally feasible and greatly reduced radiation damage. In a modification of the original method, Hope, Yonath and co-workers showed that cryocrystallography significantly reduced radiation damage from crystals of ribosomal subunits.²⁵ The flash cooling of crystals was made much simpler by the introduction of the loop technique,²⁶ which resulted in an explosion in the development and use of cryocrystallography.

Despite these improvements, it was not clear that one could obtain a signal from heavy-atom derivatives that was strong enough to obtain accurate phases for ribosome-sized asymmetric units. For several decades, the use of clusters of heavy-atoms such as tungsten or tantalum had been proposed as possible derivatives for crystals of very large macromolecules.²⁷ At very low resolution, these clusters can effectively act as single "super-heavy" atoms with very large phasing power. They were used successfully to obtain low-resolution phases for large structures such as the nucleosome core particle, rubisco and the proteasome,²⁸⁻³⁰ and their use was also advocated for ribosomes.³¹ The first concrete evidence that heavy-atom clusters could lead to reliable phases for ribosomal subunits came from studies on the 50 S subunit from *H. marismortui*, where a W18 cluster could be seen clearly directly in difference Patterson maps.³² This led to the visualization of right-handed helices of double-stranded RNA in electron density maps at 9 Å resolution. In that work, it was also demonstrated that cryo-EM reconstructions of the 50 S subunit could be used as models in molecular replacement to obtain low-resolution phases, following an approach that had been used earlier in virus crystallography.³³ Thus, low-resolution phases obtained by either of these independent methods could be used subsequently to detect an arbitrary number of single-atom sites from difference Fourier maps of other derivatives, so that the problem of getting to high resolution could be bootstrapped from low-resolution phases.

Due to their multi-atom composition, heavy-atom clusters can no longer be treated as single point scatterers beyond 6-12 Å. It is difficult to account properly for their scattering in phasing at higher resolutions unless such clusters are so well-ordered that they can be treated as a collection of individual atoms in defined positions,³⁰ or sufficiently disordered rotationally so that they can be approximated by a spherically averaged form factor.³⁴ In any case, their phasing power drops dramatically at higher resolution due to interference between the various component parts of the cluster, which no longer scatter in phase, so it is not clear that they offer special advantages at high resolution. However, heavy-atom clusters can be invaluable for obtaining initial low-resolution phases in a straightforward manner, and recent work has shown that they can be used to obtain phase information to high resolution with appropriate treatment.³⁵

Anomalous scattering, especially in conjunction with the multiple wavelength anomalous diffraction (MAD) method³⁶⁻³⁸ has also been extensively used to obtain phase information. This method relies on the tunability of synchrotrons to optimize the anomalous scattering from specific elements, exploiting the presence of sharp peaks or "white lines" in the scattering factors near X-ray absorption edges.³⁹ Despite the fact that the differences in the real and anomalous scattering factors for selenium are only a few electrons at the wavelengths employed (compared to the addition of 78 electrons in a classical isomorphous Pt derivative), previous work by us has shown that the treatment of MAD data as a standard multiple isomorphous replacement (MIR) problem produces excellent phases.^{40,41} Thus, phasing based on weak anomalous scattering works because the elimination of systematic errors from non-isomorphism or multi-crystal scaling is relatively more important than the size of the raw scattering factor of the heavy-atom, suggesting that with care, the technique could be used for very large structures. A widely used approach in MAD is the use of crystals of proteins with biologically incorporated selenomethionine residues.⁴² The total number of methionine residues in a 30 S subunit is 20 when excluding the N-terminal methionine residues which are usually either absent or disordered. The convenient web page by E. Merritt (<http://brie.bmsc.washington.edu/scatter>) can be used to estimate the anomalous scattering from a hypothetical selenomethionyl 30 S crystal, using formulas based on the original equation by Crick and Magdoff.^{43,44} The result is that the anomalous signal would be around 2%, which is too small to measure reliably even with current technology, especially given the quality of data from such crystals. However, the LIII edges of other elements such as the lanthanides and osmium have very strong white lines that are as large as 20-30 electrons, resulting in significantly higher phasing power. Their use with both proteins^{45,46} and

RNA⁴⁷ has produced exceptionally good phases. A calculation similar to the one above for selenium atoms shows that even about ten osmium or lanthanide atoms per 30 S subunit could theoretically result in a signal of 5-10%, which is better than that of selenomethionyl-substituted proteins that are now solved routinely. The use of LIII edge anomalous scattering is therefore an alternative route to phasing reflections from crystals of very large molecules, and forms the basis of the strategy described in this work.

Here we describe our efforts to obtain crystals of the *T. thermophilus* 30 S subunit that diffract to higher resolution, and give details of our determination of an atomic structure for the subunit at 3 Å resolution. LIII edges from osmium or iridium hexaammine were also used in the determination of the atomic structure of the 50 S subunit⁴⁸ and the low-resolution structure of the 70 S subunit.⁴⁹

A brief account of the structure itself was published earlier.⁵⁰ We note that an independent structure of the 30 S to 3.3 Å resolution³⁵ was also determined in the past year, largely using phasing from heavy-atom clusters.

Results

Crystallization of the 30 S subunit

Hydrophobic interaction chromatography on Sepharose 4B has been used previously to separate *Escherichia coli* ribosomal subunits in the complete absence of magnesium ions,⁵¹ where it was shown that active 30 S subunits and partially dissociated 50 S subunits could be obtained. Subsequently, it was shown by us that this method produced 30 S subunits, that while active, were substoichiometric in some ribosomal proteins.⁵² Hydrophobic interaction chromatography was also effective in a quantitative separation of active 30 S and 50 S subunits from halobacteria.⁵³ In addition, the technique proved to be useful in the crystallization of 70 S ribosomes from *T. thermophilus* by the Puschino group (M. Garber, personal communication). In this work, hydrophobic interaction chromatography using Poros ET columns in 10 mM magnesium results in an efficient separation of excess 50 S subunits from 70 S ribosomes (Figure 1(a) and (b)). This step is not absolutely required for crystallization of 30 S subunits, but offers a rapid method for purifying ribosomes away from contaminating nucleases and proteases in the crude pellet. In the final step, one could separate away the minor fraction of the 30 S subunit pool that contained protein S1 (Figure 1(c)). The ribosomes made in this way are active in standard poly(U)-directed polyphenylalanine synthesis (data not shown).

Small crystals were obtained in a pH range of 6.0-7.5 by vapor diffusion using the hanging drop method with equilibration against reservoirs containing 13-17% MPD.

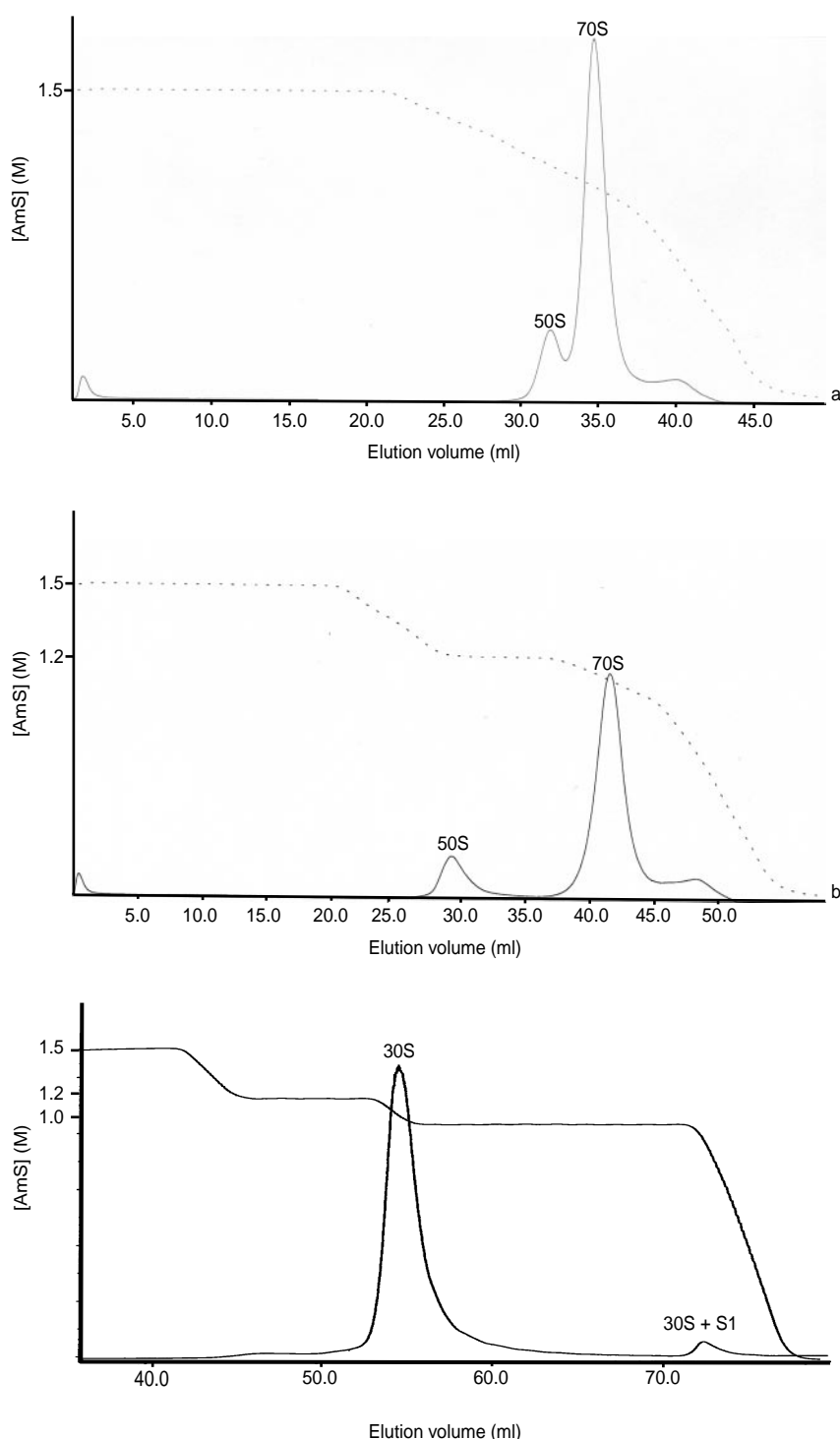


Figure 1. The use of hydrophobic interaction chromatography using the Poros ET resin in the purification of 70 S ribosomes and 30 S subunits. (a) Initial separation using a simple linear reverse gradient from 1.5 M $(NH_4)_2SO_4$ downwards; (b) optimized gradient which includes a plateau to fully isolate the 70 S from excess 50 S subunits; (c) separation of the S1-free and S1 containing 30 S pools by hydrophobic chromatography.

The largest crystals were obtained at pH 6.5. They appeared as long rods and were about $100 \mu m \times 100 \mu m \times 200\text{--}300 \mu m$ in size (Figure 2(a)), but more often smaller crystals of 20–60 μm thickness were obtained. They are morphologically similar to the ones reported originally.¹⁰ Crystals were visible in two to three weeks but usually took more than six weeks to grow to full size. Over and under-nucleation were commonly observed, and our efforts at seeding did not

result in solving the nucleation problem or in significantly larger crystals. The crystals were tetragonal with cell dimensions of approximately $401 \text{ \AA} \times 401 \text{ \AA} \times 176 \text{ \AA}$. An examination of systematic absences showed that the space group was $P4_12_12$ or $P4_32_12$; subsequently the space group was shown to be $P4_12_12$ based on the handedness of double-helical RNA in electron density maps. The best crystals diffracted to slightly beyond 3 \AA resolution (Figure 2(b)).

The crystallization was not highly reproducible. When the crystals were washed in mother liquor and then run on SDS/polyacrylamide and non-denaturing agarose/acrylamide gels, it was clear that they completely lack protein S1 (Figure 2(c)). In subsequent experiments, the S1-free population of 30 S subunits was selectively purified, either by hydrophobic interaction chromatography which resulted in a trailing edge to the main 30 S peak that was enriched in the S1-containing 30 S, or by preparative agarose/acrylamide gel electrophoresis.⁵⁴ While both methods improved crystal reproducibility, the gel electrophoresis method, while taking longer and resulting in a lower yield, gave more reproducible and better-formed crystals, presumably because it removed the S1-containing fraction more quantitatively and yielded a generally more homogeneous population of 30 S subunits. Both purified 30 S subunits as well as dissolved 30 S crystals were active in poly(U)-directed protein synthesis *in vitro* at levels comparable to that of free 70 S ribosomes (data not shown).

Phasing, model building and refinement

The first indication of a heavy-atom derivative was evident in 9 Å isomorphous and anomalous difference Patterson maps for a W17 derivative (Figure 3). The two sites identified by SOLVE⁵⁵ for this derivative were used to calculate phases to 9 Å. These phases were then used to calculate anomalous difference Fourier maps on all of the other derivative data. Data from every soak had significant peaks in difference Fourier maps. Soaks included 0.3 and 1 mM trivalent ions of the lanthanide series, Eu, Sm, Gd, Ho, Yb and Lu. Since Yb and Lu have LIII edges at the shortest wavelengths of this set, they yielded the best data due to the lower degree of X-ray absorption. Soaks were also done with tungsten clusters W30 ($K_{12.5}Na_{1.5}[NaP_5W_{30}O_{110}] \cdot 15H_2O$), W18 ($K_6[P_2W_{18}O_{62}] \cdot nH_2O$), W17 ($Li_{10}[P_2W_{17}O_{61}] \cdot nH_2O$), W12-Si ($H_4SiO_6[12WO] \cdot nH_2O$), W12 ($H_3[PW_{12}O_{40}] \cdot nH_2O$), and W11 ($K_9[PW_{11}O_{39}] \cdot 12H_2O$) and in tantalum bromide (Ta_6Br_{14}). All of these clusters showed a small number of large peaks in difference Fourier maps, and many of the tungsten clusters had overlapping sites.

For the initial phasing, we noticed that the native data set was significantly non-isomorphous with the various derivatives. We therefore decided to use the osmium hexaammine derivative, which had the best data quality and diffracted to the highest resolution,⁵⁶ as our reference or "native" data set. A scaling *R*-factor was calculated between this data set and all of the other data sets (28 in all), and a group of six data sets were chosen that were the most isomorphous with each other as judged by low scaling *R*-factor. This group consisted of the osmium hexaammine reference set (for which the anomalous differences were used in phasing), and the data from the W12Si, W17,

Ta, Yb and Lu derivatives, for which both the isomorphous and anomalous differences were used.⁵⁶ Initial phases calculated to 5.5 Å resolution in the program SOLVE⁵⁵ were solvent-flattened using SOLOMON.⁵⁷ The solvent-flattened phases yielded electron density maps in which individual strands and bumps for the phosphate positions in right-handed RNA helices were visible, as were α -helices of proteins.⁵⁶ This allowed the placement of proteins of previously known structure into the electron density maps. Using previous biochemical data, it was also possible to determine the fold of the entire central domain of 16 S RNA, the structure of the long helix 44 of the three minor domain of 16 S RNA and identify the fold and location of protein S20, whose structure and precise orientation and location in the subunit were previously unknown.⁵⁶

Subsequently, phases were calculated using only the osmium hexaammine, Lu and Yb data sets, which yielded maps that were as good as those including the cluster data, showing that the anomalous and isomorphous scattering from the single-atom derivatives was sufficiently powerful to obtain reasonable phases. Accordingly, for phasing at high resolution, we used osmium hexaammine, lutetium chloride (most of whose sites overlapped with the ytterbium sites), two pentaammines of osmium, pentaammine(dinitrogen)osmium(II)chloride and pentaammine(trifluorosulfonate) osmium(III)trifluoromethanesulfonate, and osmium bipyridine.⁵⁰

A straightforward extension of the 5.5 Å resolution structure to high resolution proved difficult for two reasons. Firstly, because the diffraction fell off very rapidly with resolution, it was necessary to expose each crystal to very high X-ray doses in order to obtain a reasonable signal-to-noise ratio at high resolution. Intrinsic variability between crystals was circumvented by extensive pre-screening of crystals, to ensure that a group of isomorphous, well-diffracting crystals of sufficiently low mosaicity were used for data collection to high resolution. Over 1000 crystals were flash-cooled; of these over 400 were prescreened. The derivative to native non-isomorphism problem seen earlier at low resolution was solved by soaking all of the non-osmium data sets in an isostructural compound, 1 mM cobalt hexaammine. This also helped to reduce the intrinsic variability in the cell dimensions of the reference crystals.

Initial phases and maps were calculated to 3.37 Å resolution using the program SOLVE. A systematic variation of the solvent content and flipping and truncation factors used in density modification with the program SOLOMON^{57,58} was necessary to determine the optimal solvent flattening procedure. Since the estimated solvent content was 70-75%, we varied the solvent content from 56-80%. For each set of these parameters, three regions of the map, which corresponded to regions of double-helical RNA or previously solved protein structures, were inspected visually to determine the optimal density modification conditions.

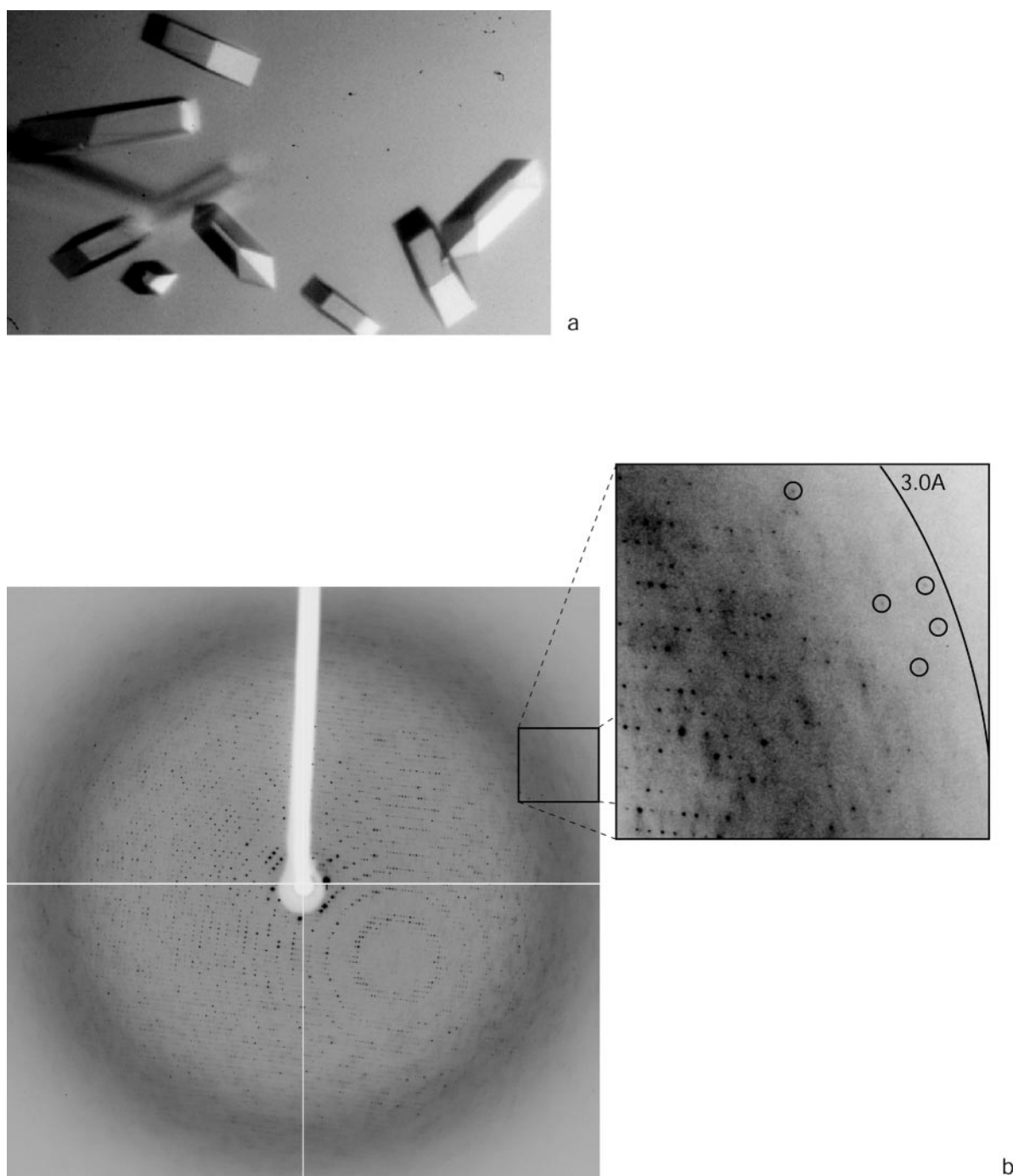


Figure 2 (legend opposite)

Initially we varied the solvent content and flipping factor independently. Visual inspection of the maps as described above suggested that the best flipping factor was the one that was related to the solvent content as described by the gamma correction formula.⁵⁸ The best parameters corresponded to a solvent content of 64%, with a flipping factor of -0.56 and a truncation factor of 0.3 . The map correlation coefficient and mean phase error as a

function of solvent content is shown in Figure 5. Density modification with SOLOMON as implemented in the SHARP package involves calculating the initial mask and modified phases at a lower resolution followed by gradual phase extension to higher resolution.⁵⁹ This procedure proved necessary to obtain interpretable maps. The averaging sphere size and the resolution of the calculated maps were initially set to 5.5 \AA , and both

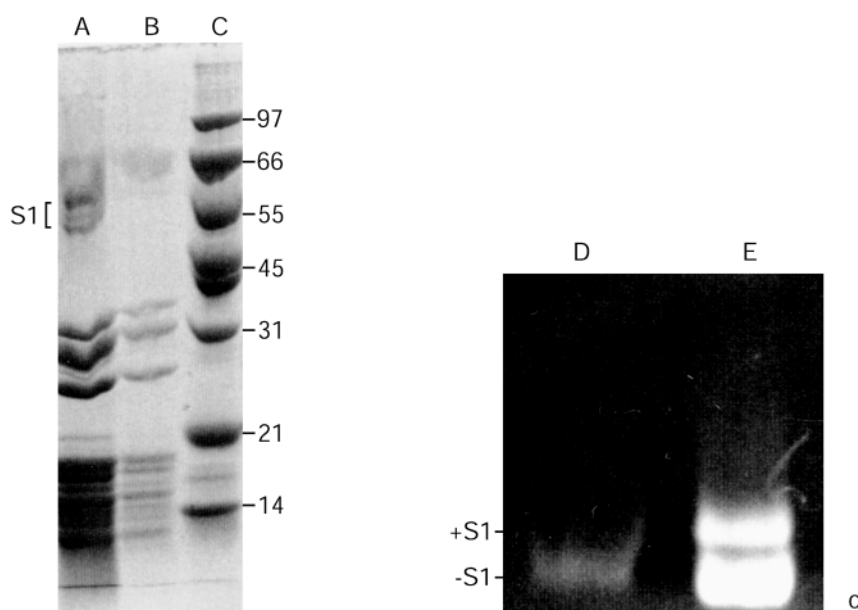


Figure 2. Crystals of the 30 S subunit. (a) Large, single crystals of *Thermus thermophilus* 30 S grown in MPD. Crystals grew to a maximum of $100 \mu\text{m} \times 100 \mu\text{m} \times 200\text{--}300 \mu\text{m}$. (b) Diffraction image collected from a Co-hexamine soaked crystal at ESRF ID14-4. The oscillation range is 0.15° and visible spots extended to slightly beyond 3.0 \AA . (c) Lanes A, B, and C, SDS-PAGE gel showing purified 30 S (A), washed and dissolved crystals (B), and molecular mass standard, numbers show molecular mass in kDa (C). Lanes D and E, a native gel showing washed and dissolved crystals (D) and purified 30 S with S1-contamination (E). The multiple bands corresponding to S1 represent different proteolytic products as judged by N-terminal sequencing.

were gradually extended in 0.1 \AA increments to 3.37 \AA resolution.

The SOLVE maps after density modification were sufficiently good to build all of the 16 S RNA. In the best regions of the maps, bases were individually discernible, and it was possible to distinguish purines from pyrimidines, thus ensuring correct registry of the model. The protein main-chain density was generally visible in the original experimental maps, and the correct topology could easily be inferred. In addition, side-chain density

was often visible for all but the smallest side-chains.⁵⁰

Maps were also calculated using the program SHARP, for which phases were calculated using only three derivatives: osmium hexaammine, pentaammine (trifluorosulfonato) osmium(III)trifluoromethanesulfonate and lutetium chloride.⁵⁰ Experimental phases were calculated to 3.2 \AA and density modification and phase extension was carried out to the limit of the native data (3.05 \AA). These maps showed significantly improved detail.

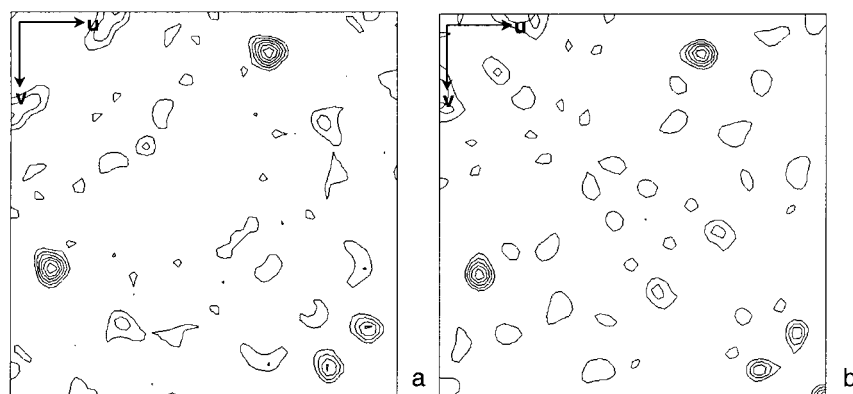


Figure 3. Peaks in the isomorphous and anomalous difference Patterson maps from a heavy-atom cluster. A Harker section of the Patterson function calculated from data collected on a crystal soaked in W17. (a) The $w = 0.5$ section of the isomorphous Patterson function, where each contour represents 1σ (b) the equivalent section of the anomalous Patterson function.

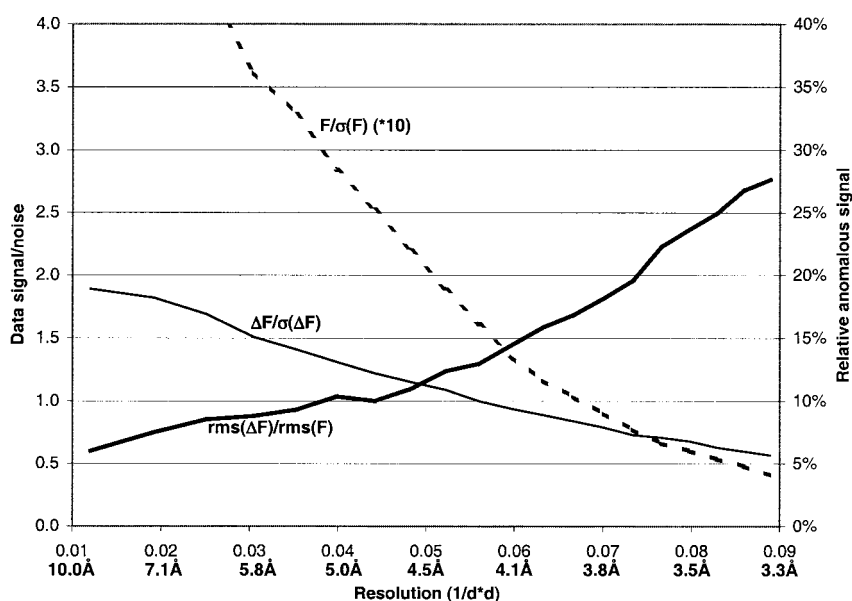


Figure 4. Anomalous signal in the osmium hexaammine derivative. The figure shows the significance of the anomalous signal present in the most important derivative, osmium hexaammine, as a function of resolution. The signal strength is represented by the average relative contribution of the anomalous signal to the structure factor amplitude as measured by the $\text{rms}(\Delta F)/\text{rms}(F)$ (thick, continuous line, measured in percentages on the right y -axis), whereas the signal/noise ratio of the anomalous data is given by $\Delta F/\sigma(\Delta F)$ (thin continuous line, absolute values on the left y -axis). Furthermore, the $F/\sigma(F)$ of the data (divided by 10 for clarity) is shown by the broken line (absolute values times 10, left y -axis).

A full description of the use of SHARP for phase determination will be published separately.

Refinement of the structure proceeded straightforwardly, in a manner that was typical for normal protein structures at this resolution, with a concomitant and gradual reduction of both the free and conventional R -factors as the structure was improved and became more complete. The final statistics are well within the acceptable criteria for correctly determined macromolecular structures.⁵⁰

The refined model could be used to estimate the relative effectiveness of the anomalous and isomorphous contributions to the overall phase errors and map correlation coefficients before and after solvent flattening. These are shown in Table 1.

Discussion

Crystallization, space group and packing

When we began this work in 1996, no crystals of the 30 S subunit had been reported that diffracted to better than 7–8 Å resolution. However, even by 1995 it was clear that the head of the 30 S had a variable conformation relative to the body of the subunit, and moreover, the various factors and RNA ligands had been mapped to the cleft between the head and body by a combination of biochemical methods and electron microscopy. It therefore seemed plausible that the 30 S subunit could be stabilized in a unique conformation by the appropriate binding of ligands, as has been done with a number of interesting enzymes, lead-

ing to improved crystals. As a starting point, we attempted to reproduce the "low resolution" crystals reported earlier in the literature.^{10–13}

Somewhat to our surprise, after optimization of the original crystallization conditions that we used as a starting point,¹⁰ we found that the apparent resolution limit of our crystals was mainly a function of crystal size and the intensity of the radiation source. Thus, our first crystals, which were about 40 μm thick, did indeed diffract only to about 9 Å at X12B at the NSLS at Brookhaven National Laboratory. However, further optimization resulted in larger crystals that diffracted beyond 6 Å on the same beamline, and beyond 4 Å at the X25 wiggler line at the NSLS. Equivalent crystals diffracted to just beyond 3 Å at the high-intensity undulator lines SBC 19ID at APS and ID14-4 at ESRF. Thus, after optimization of crystallization conditions, the crystals have intrinsic order to at least 3 Å resolution, but the fall-off in diffraction intensity means that the largest crystals and highest intensity sources are required to observe and measure data at that resolution.

With hindsight, it is clear why these crystals diffract well even without the more sophisticated co-crystallization with ligands and factors that we had planned: the P site of the 30 S subunit is filled with a tRNA mimic that consists of a stem-loop of RNA (the "spur" or helix 6) from a crystallographic symmetry-related molecule.⁶⁰ In addition, it is also apparent that the 3' end of 16 S RNA has folded back into the message-binding cleft of the 30 S subunit, so that the P site of the 30 S subunit

Table 1. Phasing statistics

	Mean phase error		Map correlation	
	Initial	After SF ^d	Initial (%)	After SF ^d (%)
	(Degrees) ^{a,b}		Coefficient ^c	
All derivatives (ano/iso)	70.9 (86.5)	32.4 (44.5)	40.8	86.6
All derivatives (ano only)	69.0 (84.4)	34.3 (45.1)	42.4	85.1
All derivatives (iso only)	77.7 (88.4)	54.0 (67.4)	27.9	66.4
Os hexaammine (ano only)	70.6 (83.9)	37.3 (48.2)	41.2	81.7

^a Mean phase difference between initial phases from SOLVE and phases calculated from the final refined model of the 30 S (Initial), in addition to similar values calculated for the solvent flattened phases (After SF). The phase difference is defined as the average difference between experimental phases and phases calculated from the final refined model, weighted by the figure of merit and measured in degrees (see the legend to Figure 5)

^b Figures in parentheses represent the outermost resolution shell used in phasing, 3.56-3.38 Å.

^c The map correlation coefficient is the correlation between a synthetic electron density map calculated based on the final model and the map corresponding to the experimental set of phases, averaged over all grid points.

^d Solvent flipping/flattening based on phases from SOLVE was carried out optimally at 68% solvent.

is filled with both message and tRNA mimics, effectively locking one subunit to the other. Thus, because of packing interactions between symmetry-related molecules, this crystal form of the 30 S subunit mimics a ligand-bound state, which is probably responsible for its conformational homogeneity and the resulting diffraction to high resolution. This also suggests that the particular length and sequence of the poorly conserved spur helix from *Thermus*, rather than the thermal stability of the molecule, is responsible for the high-resolution diffraction of these crystals.

Our crystals belong to the space group $P4_12_12$, which is different from the space group of $P4_22$ reported over several years by the Yonath group for the low-resolution form.^{12,14,25} These workers also calculated a low-resolution electron density map for the 30 S in the $P4_22$ space group.¹⁴ More recently, the Yonath group reported that both their low-resolution crystal form and a new high-resolution crystal form which was derived by soaking crystals of the low-resolution form in clusters of W18, followed by heat treatment,⁶¹ have the same space group ($P4_12_12$) as our crystals. This suggests that the earlier space group and maps are in error.

The high-resolution structures of the 30 S reported by us⁵⁰ and the Yonath group³⁵ exhibit a crystal packing that is essentially the same as that first reported by us earlier,⁵⁶ but different from the packing previously inferred by the Yonath group on the basis of molecular replacement using a cryo-EM reconstruction as a search model.⁶¹ The similarity of the overall conformation of the RNA in the two structures (PDB codes 1FJF and 1FKA) also suggests that soaking in W18 clusters and heat treatment does not represent a physiologically relevant activation, contrary to previous claims.^{35,62} Rather, it appears likely that the low-resolution crystal form represents a metastable state that can be perturbed into a form that diffracts well because the interactions in the P site of the molecule help to fix its conformation as discussed above. In any case, we have not been able to reproduce the low-resolution form of the Yonath group, in part

because the buffer and salt composition in the crystallization condition (the "H-I" buffer) is not specified explicitly either in the original crystallization report¹² or in the reference cited therein.⁷ Finally, soaking our crystals in W18 clusters does not improve diffraction, but rather, as is typical for heavy-atom derivatives, slightly worsens it. We see two large peaks in difference Fourier maps calculated from a W18 soak, but these peaks are reduced to almost the noise level upon back-soaking of the crystals (data not shown), suggesting that the clusters are not tightly bound to the 30 S subunit in our crystal lattice. This reversibility of W18 binding is different from the situation reported by the Yonath group for their high-resolution form,⁶² and once again may be a consequence of slight differences in the composition of the crystallization solutions.

Data collection

Because the crystals were variable not merely in cell dimensions, but also in diffraction limit and mosaicity, pre-screening crystals on lower intensity synchrotron sources such as the SRS or the NSLS proved to be useful to make very efficient use of limited synchrotron time at third-generation sources such as ESRF or APS. In particular, the large cell dimensions meant that we could only use crystals that had reasonable mosaic spread ($<0.35^\circ$), and even then, oscillation ranges had to be reduced to $0.10\text{--}0.15^\circ$ to avoid overlaps at high resolution. A standard strategy was used to collect non-redundant data as quickly as possible, but overlapping wedges were then collected to improve both scaling and redundancy. These overlaps meant that any particular rotation range was collected at least twice, and the resulting redundancy also allowed us to discard occasional bad wedges of data without sacrificing completeness.

Phasing and structure determination

The anomalous scattering signal, while marginal, played a crucial role in obtaining phases that were good enough to result eventually in an interpretable map. This is a tribute to the advances in synchrotron radiation facilities and crystallographic computing,^{20,21} which has made the detection of ever smaller signals possible. An analysis of the contribution of the anomalous signal to the average structure factor for the best derivative, osmium hexaammine, is shown in Figure 5. Because the anomalous signal is independent of the diffraction angle, the average relative contribution (as measured by $\text{rms}(\Delta F)/\text{rms}(F)$), increases with resolution. Much of this increase, however, is due to lower signal-to-noise at higher resolutions, as can be seen from the plot of $(\Delta F/\sigma(\Delta F))$ versus resolution. However, as judged from the maps and phase errors (Figure 6), there must be residual anomalous signal even beyond 4.5 Å resolution, where the signal-to-noise deteriorates rapidly.

In our case, we were worried about two problems that would prevent accurate measurement of a weak anomalous signal. Radiation damage would mean that anomalous pairs that were measured at different points in the data collection would have quite different scale and B factors applied to them. Secondly, scaling data from multiple crystals would also introduce systematic errors, from both absorption corrections and different degrees of radiation damage. These systematic errors in scaling can be minimized by the rotation of crystals precisely about a mirror plane. Bijvoet mates would then be measured on the same or adjacent images, thus allowing the accurate measurement of anomalous differences despite errors in scaling. Previously, it was common to measure MAD signals in this way for uncooled crystals.⁶³ However, nowadays where cryocrystallography has reduced radiation damage to negligible levels in most cases, it is possible to obtain excellent (but not necessarily optimal) MAD data without crystal orientation about a mirror plane. Similarly, precise orientation was apparently not necessary in the 50 S case, where radiation damage is apparently not as much of a problem, since each data set required just one or two crystals.⁴⁸ Orientation may be required again in future for large macromolecular structures for which radiation damage is a serious problem despite cooling to below 100 K. Careful measurements of precisely oriented and non-oriented crystals need to be done in situations where radiation damage is severe and multiple crystals are used. If it turns out that precise orientation offers significant benefits in such cases, it would speak strongly in favor of installation of kappa goniostats at those beamlines that deal with the most challenging problems in structural biology.

No particular attempt was made to specially scale the data to preserve anomalous differences. In particular, the standard procedure we used

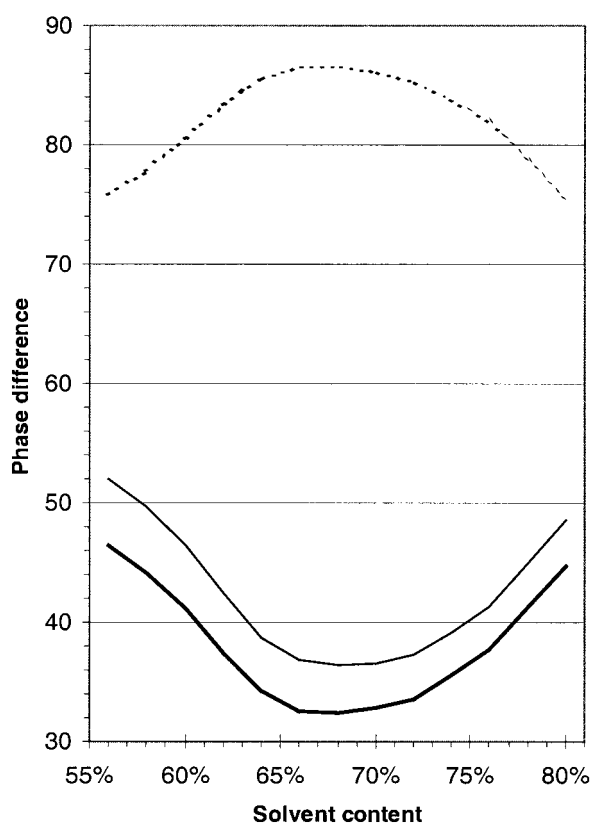


Figure 5. Determination of the optimal solvent content. Average difference between phases calculated from the final model and experimental phases determined at various levels of solvent. The graph shows the unweighted phase difference (thin continuous line) as well as the corresponding FOM-weighted difference (thick continuous line). The phase difference is calculated as:

$$\frac{1}{N} \sum_N w |\phi_{\text{mod}} - \phi_{\text{exp}}|$$

where N is the total number of reflections, w is the figure of merit (or 1 in the case of the un-weighted phase difference), ϕ_{mod} represents phases calculated from the final refined model, and ϕ_{exp} represents the experimental solvent-flattened phases calculated at a given solvent content. The Figure also shows the map correlation coefficient based on comparison between a map calculated from model phases and a given experimental electron density map (dotted line). The actual minimum of the phase difference occurs at a higher solvent content (68%) than what was found empirically to yield the most interpretable maps (64%).

treats anomalous pairs as equivalent for the purposes of scaling, but then separates them for merging and output. This has the drawback that occasionally a reflection will be rejected, while its Bijvoet mate, measured on the same or nearby image, is accepted, thereby skewing the anomalous difference for that reflection. However, this affects

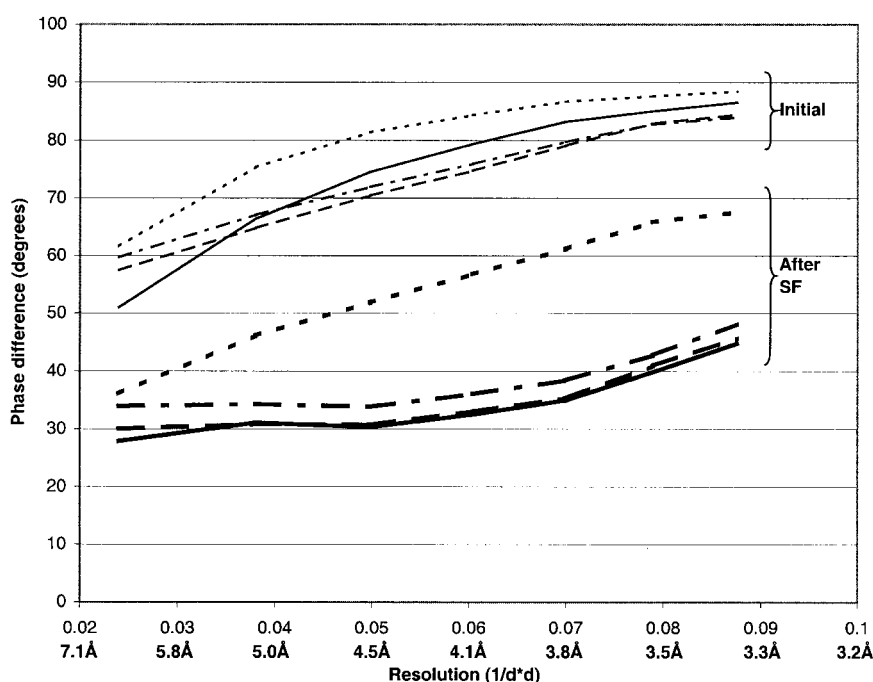


Figure 6. The importance of anomalous scattering in phase determination. The phase difference between experimental phases and model phases is shown as a function of resolution calculated as for Figure 5. The upper three graphs (thin lines, marked with Initial) show values before solvent flattening, when using both anomalous and isomorphous differences (continuous line), phasing using isomorphous differences only (dotted line), anomalous differences only (broken line), or anomalous differences from the major derivative (Os hexaammine) only (dash-dot-dash line). Below these, the four graphs show the equivalent values after solvent flattening at 68% (thick lines, marked with After SF). The lines styles are the same as above.

only a small fraction of the total number of reflections. Presumably, a more accurate anomalous signal could be extracted by more careful scaling procedures.

As was shown earlier,³² it is possible to visualize heavy-atom clusters directly in Harker sections of difference Patterson maps at low resolution even for ribosome-sized asymmetric units. In our case, we were able to see clear evidence for a W17 derivative in both isomorphous and anomalous difference Patterson maps (Figure 3). This allowed us initially to calculate 9 Å phases for the 30 S subunit by treating the W17 cluster as a point scatterer. These low-resolution maps gave us great confidence that the significant peaks in anomalous difference Fourier maps using these phases represented real heavy-atom sites. As a result, we were able to calculate a 5.5 Å map using both cluster and anomalous scattering data.⁵⁶ Subsequent calculations showed that the low-resolution phases from cluster data were not actually required to locate the sites from single-atom derivatives such as osmium and lutetium, since the automatic phasing program SOLVE was able to find a sufficient number of these sites *ab initio* using the single atom data alone. Moreover, the maps generated from these phases were comparable to those that included heavy-atom cluster data. This suggests that the careful measurement of optimized anoma-

lous scattering is capable of yielding sufficiently good phases even for very large asymmetric units.

As is often the case, density modification was crucial to obtaining an interpretable map. As shown in Figure 5, the overall phase error shows a minimum at about 68% solvent, which is closer to the estimated physical value of about 70% solvent and higher than the optimum of 64% we had found empirically from visual inspection of the maps. However, while the optimal solvent content might result in marginally improved overall maps, we found that at higher solvent contents, the automatic mask generated by the programs tended to truncate weak regions of density. These included, for example a turn of helix 44 at the interface, and helices 33, 33a and 33b that form the "beak" at the tip of the head of the subunit. These regions have weak density because they have a significantly higher *B*-factor. Thus, in practice, a compromise may have to be made between using the full power of density modification and retaining features that would otherwise be lost. In fact, it was just such an automatic mask determination and excessive truncation of the envelope that resulted in a missing beak in the head of the 30 S subunit in our lower-resolution maps at 5.5 Å.⁵⁶

Another important point is that both in our case, and in the case of the 50 S subunit, phase extension to the resolution limit of the native data was quite

useful. For example, we had phases from multiple derivatives only to just beyond 3.4 Å, with a single derivative contributing some poor phase information to 3.2 Å. Nevertheless, phase extension to the 3.05 Å limit of the native data proved useful to improve the real resolution of the maps, as judged by the separation of stacked bases and the shapes of purines and pyrimidines.⁵⁰ It is likely that in our case, phase extension was helped by the large solvent content, but it also appears to have played an important role in the phasing of the 50 S structure, where the solvent content is not as high.⁴⁸

Relative importance of anomalous scattering

To judge the relative importance of the isomorphous and anomalous scattering from the various derivatives, we have refined heavy-atom parameters and calculated phases using both isomorphous and anomalous contributions, isomorphous differences only, and anomalous differences only. In addition, we also calculated anomalous only phases from our single best derivative, osmium hexaammine. The overall map correlation coefficients and phase errors are shown in Table 1, and the phase error as a function of resolution is shown in Figure 6. Density modification was done for each case using the solvent content that gave the lowest phase error (68%).

As can be seen from both Table 1 and Figure 6, the initial phases from using anomalous scattering alone are at least as good as those obtained using both anomalous and isomorphous differences, whereas the phases from isomorphous differences alone are significantly worse and would not have led to an interpretable map. After solvent flattening, this effect is even more pronounced, and we see that using anomalous difference only results in exactly the same phase error as that of the combined signal. This shows that anomalous scattering made most of the contribution to phasing, and that the structure could have been solved with its use alone. Moreover, it is striking that just the anomalous scattering from a single derivative, osmium hexaammine, produced phases that were almost as good as all of the derivatives combined. As we have noted at the outset, the 50 S structure was also solved using primarily anomalous scattering from LIII edges,⁴⁸ as was the low-resolution structure of the 70 S ribosome.⁴⁹ In hindsight, therefore, a more appropriate strategy may have been to collect much better data on the osmium hexaammine derivative alone, rather than on several different derivatives. Precisely this strategy has worked for the 70 S ribosome.⁴⁹

The importance of anomalous scattering also lends credence to our opinion that orienting crystals about a mirror plane in cases where radiation damage and crystal scaling is a problem could be crucial to obtain good enough initial phases to solve a structure.

Model building and refinement

The model building proceeded in two stages. When 5.5 Å maps of the 30 S subunit were obtained, it was possible to place all of the proteins whose structures had been solved in isolation.⁵⁶ In combination with biochemical data on protein-RNA interactions, we could also trace the fold of the entire central domain of 16 S RNA, trace helix 44 of the 3' domain of 16 S RNA, and also identify protein S20 as a three-helix bundle near the "bottom" of the subunit. Subsequently, the incorporation of the structure of S16⁶⁴ allowed the determination of the fold of the 5' domain even at intermediate resolution. The high-resolution structures of individual ribosomal proteins were therefore extremely useful to quickly map out the architecture of the 30 S subunit, and without them, analysis of the 30 S maps would have been much more difficult.

Once high-resolution maps were obtained, model building and refinement were straightforward, and our task was made easier by the fact that a rough architecture of much of the molecule had already been determined. It is possible that our initial use of phase constraints during refinement allowed an easier and unbiased convergence. Another reason is undoubtedly the large solvent content. Finally, the fact that the structure consisted mainly of RNA (about two-thirds by mass), for which the phosphate backbone positions were very well determined, could also have been a factor in assuring straightforward refinement.

It is important to note that in a structure of this size, the experimental density maps are of utmost importance. It is possible to build an entire protein backwards (i.e. C to N terminus) without affecting the free *R*-factor significantly, since any particular protein accounts for such a small fraction of the combined scattering. On the other hand, the significant phasing power provided by the more than 1500 phosphate atoms that were placed with high certainty in the structure, resulted in a refined electron density map with very little model bias with respect to the protein structures. We generally found that, e.g. side-chains placed in the wrong position would show up in the right position in the map after a single round of refinement.

General lessons for crystallography on large asymmetric units

Some general conclusions for structure determination of large macromolecules can be drawn from this work and other structures such as the 50 S subunit,⁴⁸ an independent 30 S structure³⁵ and the structure of RNA polymerases.^{65,66}

A sufficient heavy-atom signal can be obtained even for very large structures by the careful use of anomalous scattering, in particular from the peaks of white lines at LIII edges. It is true that, e.g. the uranium M-edges have significantly greater signal per atom and that this can be used for macromol-

ecular crystallography.³⁸ However, these edges correspond to long wavelengths with tremendous technical problems in X-ray absorption, radiation damage and data collection, exacerbating an already difficult situation. Thus, LIII edges of lanthanides and elements such as osmium, tungsten, tantalum and iridium, which have anomalous scattering factors of about 20-30 electrons, may represent an optimal compromise between the much weaker K-edge of selenium and the M-edges of uranium. Careful measurement of anomalous scattering, either with oriented crystals using a kappa goniostat, or by the use of the inverse beam method, may also be required.

With small crystals of large asymmetric units, the small number of unit cells in the crystal results in weak diffraction. This can be overcome by the use of intense beams from undulator insertion device at third-generation synchrotrons, where the beam size is carefully matched to the crystal dimensions so that excess background scattering is minimized. Improvements in the size, resolution, and data readout speed of area detectors have also made collection of data on large unit cells feasible.

Problems in cell variability or quality can be solved by extensive pre-screening of crystals, on a home source if possible or on a lower-intensity synchrotron beamline. Alternatively, it may be possible to solve the cell variability problem by the addition of ligands or changes in the crystallization conditions.

The determination of a large number of heavy-atom sites is no longer a technical problem. One approach, as exemplified here, is to use low-resolution phases from heavy-atom clusters to calculate difference Fourier maps for single-atom, many-site derivatives to determine heavy-atom positions, as we did initially. However, it is now clear that programs such as SOLVE⁵⁵ or SnB, an implementation of the Shake-and-Bake methodology,⁶⁷ can now detect very large numbers of heavy-atom sites automatically, and the remaining sites can be identified from difference Fourier maps using initial phases.

This work also shows the importance of careful density modification protocols. In practice, it may be necessary to systematically vary the various parameters in order to obtain the optimal map for interpretation. Such an optimal map may not necessarily be the "best" map as judged by conventional criteria such as overall phase error or map correlation coefficient with the final refined model, but rather may be one that has the most continuous chain density or less truncation of weak density. In our work, we were not able to automate or simplify the establishment of optimal density modification parameters, but were required to manually check a large number of combinations visually. Here too, improved algorithms are on the horizon,⁶⁸ which may make manual optimization unnecessary.

In summary, improvements in crystallographic technology and methodology mean that the deter-

mination of structures as large as 1-3 MDa is now limited mainly by the availability of suitable crystals, and the achievable resolution depends mainly on the diffraction limit of the crystals and characteristics such as cell dimensions and mosaic spread.

Materials and Methods

Purification of 30 S ribosomal subunits from *Thermus thermophilus*

All buffers were at 4°C and contained 6 mM 2-mercaptoethanol, 0.1 mM phenylmethane sulfonyl fluoride (PMSF), and 0.1 mM benzamide, unless otherwise noted. *T. thermophilus* strain HB8 was grown in ATCC 697 medium supplemented with Castenholz salts⁶⁹ under vigorous aeration at 72-75°C to an A_{600} that was half the saturation value. The cells were then cooled rapidly and harvested by centrifugation. In addition, large-scale fermentation (400 l) was done at the University of Georgia fermentation facility. The cell paste was either stored at -70°C or used immediately. Cells were resuspended in an equal volume of buffer A (0.1 M NH₄Cl, 10.5 mM magnesium acetate, 0.5 mM EDTA, 20 mM Hepes, adjusted with NaOH to pH 7.5) and disrupted by two passages through a French press at 12,000 psi. Cell debris was removed by centrifugation at 15,000 rpm for 30 minutes in a Sorvall SS-34 rotor and the clear part of the supernatant was overlaid onto at least an equal volume of a cushion containing 1.0 M NH₄Cl, 1.1 M sucrose, 10.5 mM magnesium acetate, 0.5 mM EDTA, 20 mM Hepes (pH 7.5), and centrifuged at 43,000 rpm for 15 hours in a Beckman Ti70 rotor.

The pellet, containing salt-washed 70 S ribosomes, was washed briefly in buffer A to remove a surface layer of cell debris and membrane fragments, then resuspended in buffer C (1.5 M (NH₄)₂SO₄, 10 mM magnesium acetate, 20 mM Hepes (pH 7.5)), filtered through a 0.2 µm membrane, loaded onto a 25 ml Poros-ET hydrophobic interaction column equilibrated in the same buffer, and eluted with a reverse ammonium sulfate gradient made by combining buffer C and buffer D (20 mM Hepes (pH 7.5), 10 mM magnesium acetate) using a Biocad Sprint or Pharmacia ÄKTA automated chromatographic system. The 70 S peak was pooled, dialyzed into buffer E (50 mM KCl, 10 mM (NH₄)₂SO₄, 10 mM Hepes (pH 7.5), 10.25 mM magnesium acetate, 0.25 mM EDTA) and then into buffer F (the same as buffer E, but with 2.25 mM magnesium acetate) and concentrated in an Amicon pressure cell equipped with a PM30 membrane. The resulting dissociated mixture of 30 S and 50 S subunits were separated on a 10%-30% linear sucrose gradient in buffer F by centrifugation at 28,000 rpm for 16 hours in a Beckman SW28 rotor. Gradients were collected from the top using a Labconco Autodensiflow fractionator.

The 30 S peak was dialyzed into buffer C, and repurified on the PorosET column using a reverse ammonium sulfate gradient as above. A trailing edge that contained a fraction enriched in S1-containing 30 S subunits was discarded, and the remaining 30 S peak was pooled, dialyzed into buffer G (50 mM KCl, 10 mM (NH₄)₂SO₄, 5 mM Hepes (pH 7.5), 10 mM magnesium acetate, and concentrated using Millipore Ultrafree spin concentration filters (30 kDa molecular mass cutoff) to a final concentration of 20-35 mg/ml. Alternatively, the 30 S peak from sucrose gradients was dialyzed into buffer G, concen-

trated and run on a preparative non-denaturing agarose/acrylamide gel⁵⁴ using a Biorad PrepCell. The S1-free population of 30 S subunit present as a leading peak was pooled, re-dialyzed into buffer G and concentrated as described above.

Crystallization

Just prior to setting up crystallizations, the 30 S solution in buffer G was incubated at 40–50 °C for 15 minutes. Crystallization was done using the hanging drop method, with a 1:1 mixture of 30 S subunits in buffer G and the reservoir solution, which contained 14–16% MPD (v/v) in buffer H (0.2 M KCl, 75 mM NH₄Cl, 15 mM MgCl₂, 0.1 M buffer, with a pH initially ranging from 6.0–8.5). Optimization of crystallization conditions, 0.1 M Mes or cacodylate was used as the buffer at pH 6.5. The volume of each drop was 8–20 μ l.

Cryoprotection, flash cooling and screening of crystals

Entire drops containing crystals to be transferred into the final cryoprotecting buffer were equilibrated by vapor diffusion for two to three days against a reservoir containing 20% MPD in buffer H, then 10 μ l of this reservoir solution was added to the drop, followed by equilibration against a reservoir containing the final stabilizing buffer of 26% MPD in buffer H for another two to three days. Crystals were then transferred individually into fresh stabilizing buffer with or without an appropriate concentration of various heavy-atom compounds. Heavy-atom soaks were carried out for various times from overnight for the lanthanides and osmium hexaammine to a week for various heavy-atom clusters.

Crystals were flash-cooled in cryoloops (Hampton Research) by plunging into liquid nitrogen, and stored in liquid nitrogen until data collection. Crystals were screened, initially at beamline X12B at the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory, and more recently at beamlines 9.6, 14.1 or 14.2 at the Synchrotron Radiation Source (SRS) at Daresbury Laboratory. For screening, one or two exposures were taken at each of two orientations of the crystal, that differed at least 40° about the rotation axis. The crystals were returned to liquid nitrogen immediately after screening. The images from each crystal were then analyzed for the signal-to-noise as a function of resolution, estimated mosaic spread and cell dimensions. A recent version of MOSFLM was useful to estimate accurate cell dimensions, as well as mosaicity.⁷⁰

Data collection and integration

Diffraction data were collected at beamline X25 at the NSLS to obtain a medium-resolution structure at 5.5 Å. Beamlines SBC 19ID at the Advanced Photon Source (APS) at Argonne National Laboratory, and ID14-4 at the European Synchrotron Radiation Facility (ESRF) were used for the structure at 3 Å resolution.

Experimental phase information was obtained mainly at X25 and SBC 19ID. The wavelength of the X-ray beam was tuned to the peak of the white line for any given heavy-atom, as judged by the X-ray fluorescence measured at right-angles to the beam. When possible, data were collected about a mirror plane of the crystal as follows. The appropriate software, a modified version of MADNES in the case of X25^{71,72} or d*TREK in the case of

SBC 19ID,⁷³ was used to index a test shot, and then drive the kappa goniostat to a setting where the rotation axis coincided with the *c** axis of the crystal. At SBC 19ID and ID14-4, the size of the beam could be reduced to below 100 μ m to precisely match the crystal thickness and hence minimize background scattering. Since the long dimension of the crystal coincided with *c** axis, it was sometimes possible to take multiple shots from each crystal by translating along the spindle. Each part of the crystal typically only yielded a few degrees of data before the outer shell showed significant decay. It was often the case that data collection on one part of the crystal would damage the unexposed parts, presumably because of X-ray fluorescence from the absorption at the LIII edge. So quite often, only one sweep of data could be collected from a crystal.

Data were integrated using Denzo and Scalepack on X25 or HKL-2000 on SBC 19ID and ID14-4.⁷⁴

Phasing and calculation of electron density maps

Most standard crystallographic operations such as reduction of intensities to structure factors and calculation of Patterson and Fourier maps were done using the CCP4 package.⁷⁵ Phasing was carried out by two different methods. In the first, heavy-atom sites were identified automatically using SOLVE,⁵⁵ and were used to calculate initial phases at low-medium resolution. These phases could then be used to identify weaker sites using difference Fourier maps, and SOLVE was used again to refine parameters for all the heavy-atoms and calculate phases. In a second procedure, phases were calculated using SHARP.⁵⁹ Density modification was carried out using SOLOMON in both cases.^{57,58} In particular, the procedure implemented in the SHARP package for using SOLOMON to gradually extend phases to higher resolution was important for phase improvement. Moreover, the solvent content, truncation factor, flipping factor and starting resolution were all varied systematically.

Model building

The program O,⁷⁶ with the incorporation of tools for RNA, was used to build the model into electron density. For the 5.5 Å resolution maps, proteins whose structures were previously determined in isolation were fitted visually as rigid bodies into the density. Double-helical RNA was built as regular A-form helices. No attempt was made to do any minimization or refinement at this stage. When the resolution improved beyond 3.4 Å, a complete atomic model for the RNA was built, the proteins whose structures had been previously solved in isolation (but often from other species) were rebuilt into density using their corresponding *Thermus* sequence, and finally the previously unknown proteins were built. For the most poorly defined proteins (parts of S9 and S10), a complete model could only be established after a few rounds of refinement, when σ_A -weighted $2mF_o - DF_c$ maps⁷⁷ were used to aid interpretation along with the original experimental maps.

Refinement

Refinement was carried out using the program CNS.⁷⁸ The maximum likelihood function was used as the target throughout, and bulk solvent correction was used to allow inclusion of low-resolution data. To ensure proper

convergence, density-modified experimental phase constraints were included in the initial rounds of refinement. After two rounds, the phase constraints were removed, and the refinement was carried out against amplitudes only, to the full extent of the resolution of the data (59.7–3.05 Å). The “multi-endo” parameter file, which allows refinement of both 2' and 3'-endo sugars, was used for the RNA,⁷⁹ and the Engh-Huber parameters (the default in CNS) were used for proteins.⁸⁰ In the final stages, a positional minimization was followed by grouped *B*-factor refinement and refinement of the occupancies of metal ions. In spite of the large and uncommon nature of the macromolecular complex, standard procedures were used throughout the refinement. The native structure contained cobalt hexaammine, which could be seen as spherical density in both the experimental and refined maps. These were included and their occupancies refined. It is possible, however, that some of these ions in fact represent magnesium or other metal ions, which are hard to distinguish at this resolution.

Acknowledgments

This work was funded by the Medical Research Council (UK) and grant GM 44973 from the NIH (to S.W. White and V.R.). W.M.C. was the recipient of an NIH predoctoral fellowship, and D.E.B. was supported by a postdoctoral fellowship from The Human Frontier Science Programme (HFSP). We thank M. Pope for a gift of the various tungsten clusters, G. Schneider and J. Löwe for tantalum bromide, and B.S. Brunschwig and M. Choi for synthesizing osmium hexaammine; A. Joachimiak, S. Ginell, R. Ravelli, S. McSweeney, G. Leonard, M. Capel, H. Lewis, L. Berman, M. Papiz, S. Girdwood and M. MacDonald for help and advice with synchrotron data collection; J. Löwe for advice with density modification; M. Kjeldgaard and T.A. Jones for supplying O with RNA tools; and R.N. Duttall for help with determining and optimizing the chromatographic conditions used. We also thank A.G.W. Leslie, R. Henderson, P.R. Evans and R.A. Crowther for helpful comments on the manuscript.

References

- Garrett, R. A., Douthwaite, S. R., Liljas, A., Matheson, A. T., Moore, P. B. & Noller, H. F. (2000). Editors of *The Ribosome. Structure, Function, Antibiotics and Cellular Interactions*, ASM Press, Washington, DC.
- Liljas, A. & Garber, M. (1995). Ribosomal proteins and elongation factors. *Curr. Opin. Struct. Biol.* **5**, 721–727.
- Ramakrishnan, V. & White, S. W. (1998). Ribosomal protein structures: insights into the architecture, machinery and evolution of the ribosome. *Trends Biochem. Sci.* **23**, 208–212.
- Wimberly, B. T., Guymon, R., McCutcheon, J. P., White, S. W. & Ramakrishnan, V. (1999). A detailed view of a ribosomal active site: the structure of the L11- RNA complex. *Cell*, **97**, 491–502.
- Conn, G. L., Draper, D. E., Lattman, E. E. & Gittis, A. G. (1999). Crystal structure of a conserved ribosomal protein-RNA complex. *Science*, **284**, 1171–1174.
- Lu, M. & Steitz, T. A. (2000). Structure of *Escherichia coli* ribosomal protein L25 complexed with a 5S rRNA fragment at 1.8-Å resolution. *Proc. Natl Acad. Sci. USA*, **97**, 2023–2028.
- Yonath, A., Mussig, J., Tesche, B., Lorenz, S., Erdmann, V. A. & Wittmann, H. G. (1980). Crystallization of the large ribosomal subunits from *Bacillus stearothermophilus*. *Biochem. Int.* **1**, 428–435.
- von Böhlen, K., Makowski, I., Hansen, H. A. S., Bartels, H., Berkovitch-Yellin, Z. & Zaytzev-Bashan, A. *et al.* (1991). Characterization and preliminary attempts for derivatization of crystals of large ribosomal subunits from *Haloarcula marismortui* diffracting to 3 Å resolution. *J. Mol. Biol.* **222**, 11–15.
- Yusupov, M. M., Trakhanov, S. D., Barynin, V. V., Borovyagin, V. L., Garber, M. B. & Sedelnikova, O. M. *et al.* (1987). Crystallization of 30 S ribosomal subunits from *T. thermophilus*. *Dokl Akad. Nauk. USSR*, **292**, 1271–1274.
- Trakhanov, S. D., Yusupov, M. M., Agalarov, S. C., Garber, M. B., Ryazantsev, S. N., Tischenko, S. V. & Shirokov, V. A. (1987). Crystallization of 70 S ribosomes and 30 S ribosomal subunits from *Thermus thermophilus*. *FEBS Letters*, **220**, 319–322.
- Glötz, C., Müssig, J., Gewitz, H. S., Makowski, I., Arad, T., Yonath, A. & Wittmann, H. G. (1987). Three-dimensional crystals of ribosomes and their subunits from eu- and archaebacteria. *Biochem. Int.* **15**, 953–960.
- Yonath, A., Glötz, C., Gewitz, H. S., Bartels, K. S., von Böhlen, K., Makowski, L. & Wittmann, H. G. (1988). Characterization of crystals of small ribosomal subunits. *J. Mol. Biol.* **203**, 831–834.
- Yusupov, M. M., Tischenko, S. V., Trakhanov, S. D., Ryazantsev, S. N. & Garber, M. B. (1988). A new crystalline form of 30 S ribosomal subunits from *Thermus thermophilus*. *FEBS Letters*, **238**, 113–115.
- Schluzen, F., Hansen, H. A. S., Thygesen, J., Bennett, W. S., Volkmann, N. & Levin, I. *et al.* (1995). A milestone in ribosomal crystallography: the construction of preliminary electron density maps at intermediate resolution. *Biochem. Cell. Biol.* **73**, 739–749.
- Frank, J., Zhu, J., Penczek, P., Li, Y., Srivastava, S. & Verschoor, A. *et al.* (1995). A model of protein synthesis based on cryo-electron microscopy of the *E. coli* ribosome. *Nature*, **376**, 441–444.
- Frank, J., Verschoor, A., Li, Y., Zhu, J., Lata, R. K. & Radermacher, M. *et al.* (1995). A model of the translational apparatus based on a three-dimensional reconstruction of the *Escherichia coli* ribosome. *Biochem. Cell. Biol.* **73**, 757–765.
- Stark, H., Mueller, F., Orlova, E. V., Schatz, M., Dube, P. & Erdemir, T. *et al.* (1995). The 70 S *Escherichia coli* ribosome at 23 Å resolution: fitting the ribosomal RNA. *Structure*, **3**, 815–821.
- Lata, K. R., Agrawal, R. K., Penczek, P., Grassucci, R., Zhu, J. & Frank, J. (1996). Three-dimensional reconstruction of the *Escherichia coli*: 30 S ribosomal subunit in ice. *J. Mol. Biol.* **262**, 43–52.
- Abrahams, J. P., Leslie, A. G. W., Lutter, R. & Walker, J. E. (1994). Structure at 2.8 Å resolution of F1-ATPase from bovine heart mitochondria (see comments). *Nature*, **370**, 621–628.
- Helliwell, J. R. (1998). Synchrotron radiation facilities. *Nature Struct. Biol.* **5**, 614–617.
- Hendrickson, W. A. (2000). Synchrotron crystallography. *Trends Biochem. Sci.* **25**, 637–643.
- Garman, E. & Schneider, T. R. (1997). Macromolecular cryocrystallography. *J. Appl. Crystallog.* **30**, 211–237.

23. Garman, E. (1999). Cool data: quantity and quality. *Acta Crystallog. sect. D*, **55**, 1641-1653.
24. Hope, H. (1988). Cryocrystallography of biological macromolecules: a generally applicable method. *Acta Crystallog. sect. B*, **44**, 22-26.
25. Hope, H., Frolow, F., von Böhlen, K., Makowski, I., Kratky, C. & Halfon, Y. *et al.* (1989). Cryocrystallography of ribosomal particles. *Acta Crystallog. sect. B*, **45**, 190-199.
26. Teng, T.-Y. (1990). Mounting of crystals for macromolecular crystallography in a free-standing thin film. *J. Appl. Crystallog.* **23**, 387-391.
27. Blundell, T. L. & Johnson, L. N. (1976). *Protein Crystallography*, Academic Press, New York.
28. O'Halloran, T. V., Lippard, S. J., Richmond, T. J. & Klug, A. (1987). Multiple heavy-atom reagents for macromolecular X-ray structure determination. Application to the nucleosome core particle. *J. Mol. Biol.* **194**, 705-712.
29. Andersson, I., Knight, S., Schneider, G., Lindqvist, Y., Lundqvist, T., Brändén, C.-I. & Lorimer, G. H. (1989). Crystal structure of the active-site of ribulose-bisphosphate carboxylase. *Nature*, **337**, 229-234.
30. Knablein, J., Neufeind, T., Schneider, F., Bergner, A., Messerschmidt, A. & Lowe, J. *et al.* (1997). Ta6Br(2+)12, a tool for phase determination of large biological assemblies by X-ray crystallography. *J. Mol. Biol.* **270**, 1-7.
31. Thygesen, J., Weinstein, S., Franceschi, F. & Yonath, A. (1996). The suitability of multi-metal clusters for phasing in crystallography of large macromolecular assemblies. *Structure*, **4**, 513-518.
32. Ban, N., Freeborn, B., Nissen, P., Penczek, P., Grassucci, R. A. & Sweet, R. *et al.* (1998). A 9 Å resolution X-ray crystallographic map of the large ribosomal subunit. *Cell*, **93**, 1105-1115.
33. Jack, A., Harrison, S. C. & Crowther, R. A. (1975). Structure of tomato bushy stunt virus. II. Comparison of results obtained by electron microscopy and X-ray diffraction. *J. Mol. Biol.* **97**, 163-172.
34. Fu, J., Gnat, A. L., Bushnell, D. A., Jensen, G. J., Thompson, N. E. & Burgess, R. R., *et al.* (1999). Yeast RNA polymerase II at 5 Å resolution. *Cell*, **98**, 799-810.
35. Schlutzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M. & Janell, D., *et al.* (2000). Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. *Cell*, **102**, 615-623.
36. Phillips, J. C. & Hodgson, K. O. (1980). The use of anomalous scattering effects to phase diffraction patterns from macromolecules. *Acta Crystallog. sect. A*, **36**, 856-864.
37. Hendrickson, W. A. (1985). Analysis of protein structure from diffraction measurement at multiple wavelengths. *Trans. Am. Crystallog. Assoc.* **21**, 11-21.
38. Hendrickson, W. A. & Ogata, C. M. (1997). Phase determination from multiwavelength anomalous diffraction measurements. *Methods Enzymol.* **276**, 494-523.
39. Phillips, J. C., Wlodawer, A., Goodfellow, J. M., Watenpaugh, K. D., Sieker, L. C., Jensen, L. H. & Hodgson, K. O. (1977). Applications of synchrotron radiation to protein crystallography. II. Anomalous scattering, absolute intensity and polarization. *Acta Crystallog. sect. A*, **33**, 445-455.
40. Ramakrishnan, V., Finch, J. T., Graziano, V., Lee, P. L. & Sweet, R. M. (1993). Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. *Nature*, **362**, 219-223.
41. Ramakrishnan, V. & Biou, V. (1997). Treatment of multiwavelength anomalous diffraction data as a special case of multiple isomorphous replacement. In *Methods in Enzymology* (Carter, C. W., Jr & Sweet, R. M., eds), vol. 276, pp. 538-557, Academic Press, New York.
42. Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.* **9**, 1665-1672.
43. Crick, F. H. C. & Magdoff, B. S. (1956). The theory of the method of isomorphous replacement for protein crystals. I. *Acta Crystallog.* **9**, 901-908.
44. Hendrickson, W. A. (1991). Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science*, **254**, 51-58.
45. Kahn, R., Fourme, R., Bosshard, R., Chiadmi, M., Risler, J. L., Dideberg, O. & Wery, J. P. (1985). Crystal structure study of *Opsanus tau* parvalbumin by multiwavelength anomalous diffraction. *FEBS Letters*, **179**, 133-137.
46. Weis, W. I., Kahn, R., Fourme, R., Drickamer, K. & Hendrickson, W. A. (1991). Structure of the calcium-dependent lectin domain from a rat mannose-binding protein determined by MAD phasing. *Science*, **254**, 1608-1615.
47. Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L. & Kundrot, C. E. *et al.* (1996). Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science*, **273**, 1678-1685.
48. Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, **289**, 905-920.
49. Cate, J. H., Yusupov, M. M., Yusupova, G. Z., Earnest, T. N. & Noller, H. F. (1999). X-ray crystal structures of 70 S ribosome functional complexes (see comments). *Science*, **285**, 2095-2104.
50. Wimberly, B. T., Brodersen, D. E., Clemons, W. M. J., Morgan-Warren, R., Vornrhein, C., Hartsch, T. & Ramakrishnan, V. (2000). Structure of the 30 S ribosomal subunit. *Nature*, **407**, 327-339.
51. Kirillov, S. V., Makhno, V. I., Peshin, N. N. & Semenov, Y. P. (1978). Separation of ribosomal subunits of *Escherichia coli* by Sepharose chromatography using reverse salt gradient. *Nucl. Acids Res.* **5**, 4305-4315.
52. Ramakrishnan, V., Graziano, V. & Capel, M. S. (1986). A role for proteins S3 and S14 in the 30 S ribosomal subunit. *J. Biol. Chem.* **261**, 15049-15052.
53. Saruyama, H. (1986). Isolation of ribosomal subunits from an extremely halophilic archaeobacterium *Halobacterium halobium* by hydrophobic interaction chromatography. *Anal. Biochem.* **159**, 12-16.
54. Dahlberg, A. E. (1979). A gel electrophoretic separation of bacterial ribosomal subunits with and without protein S1. *Methods Enzymol.* **59**, 397-401.
55. Terwilliger, T. & Berendzen, J. (1999). Automated MAD and MIR structure determination. *Acta Crystallog. sect. D*, **55**, 849-861.
56. Clemons, W. M., Jr, May, J. L. C., Wimberly, B. T., McCutcheon, J. P., Capel, M. & Ramakrishnan, V. (1999). Structure of a bacterial 30 S ribosomal subunit at 5.5 Å resolution. *Nature*, **400**, 833-840.
57. Abrahams, J. P. & Leslie, A. G. W. (1996). Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallog. sect. D*, **52**, 30-42.

58. Abrahams, J. P. (1997). Bias reduction in phase refinement by modified interference functions: introducing the gamma correction. *Acta Crystallog. sect. D*, **53**, 371-376.
59. de la Fortelle, E. & Bricogne, G. (1997). Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. In *Methods in Enzymology* (Carter, C. W., Jr & Sweet, R. M., eds), vol. 276, pp. 472-493, Academic Press, New York.
60. Carter, A. P., Clemons, W. M. J., Brodersen, D. E., Morgan-Warren, R., Wimberly, B. T. & Ramakrishnan, V. (2000). Functional insights from the structure of the 30 S ribosomal subunit and its interaction with antibiotics. *Nature*, **407**, 340-348.
61. Harms, J., Tocilj, A., Levin, I., Agmon, I., Stark, H. & Kolln, I., *et al.* (1999). Elucidating the medium-resolution structure of ribosomal particles: an interplay between electron cryo-microscopy and X-ray crystallography. *Struct. Fold. Des.* **7**, 931-941.
62. Tocilj, A., Schlunzen, F., Janell, D., Gluhmann, M., Hansen, H. A. & Harms, J. *et al.* (1999). The small ribosomal subunit from *Thermus thermophilus* at 4.5 Å resolution: pattern fittings and the identification of a functional site. *Proc. Natl Acad. Sci. USA*, **96**, 14252-14257.
63. Hendrickson, W. A., Pahler, A., Smith, J. L., Satow, Y., Merritt, E. A. & Phizackerley, R. P. (1989). Crystal structure of core streptavidin determined from multiwavelength anomalous diffraction of synchrotron radiation. *Proc. Natl Acad. Sci. USA*, **86**, 2190-2194.
64. Allard, P., Rak, A. V., Wimberly, B. T., Clemons, W. M., Jr, Kalinin, A. & Helgstrand, M. *et al.* (2000). Another piece of the ribosome: solution structure of S16 and its location in the 30 S subunit. *Struct. Fold. Des.* **8**, 875-882.
65. Zhang, G., Campbell, E. A., Minakhin, L., Richter, C., Severinov, K. & Darst, S. A. (1999). Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell*, **98**, 811-824.
66. Cramer, P., Bushnell, D. A., Fu, J., Gnatt, A. L., Maier-Davis, B. & Thompson, N. E. *et al.* (2000). Architecture of RNA polymerase II and implications for the transcription mechanism. *Science*, **288**, 640-649.
67. Hauptman, H. (1997). Shake-and-bake: an algorithm for automated solution *ab initio* of crystal structures. *Methods Enzymol.* **277**, 3-13.
68. Terwilliger, T. C. (2000). Maximum-likelihood density modification. *Acta Crystallog. sect. D*, **56**, 965-972.
69. Williams, R. A. D. (1991). The genus *Thermus*. *Thermophilic Bacteria* (Kristjansson, J. K., ed.), CRC Press, Boca Raton.
70. Leslie, A. G. W. (1992). Recent changes to the MOSFLM package for processing film and image plate data, *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography*, No. 26.
71. Messerschmidt, A. & Pflugrath, J. W. (1987). Crystal orientation and X-ray pattern prediction routines for area-detector diffractometer systems in macromolecular crystallography. *J. Appl. Crystallog.* **20**, 306-315.
72. Skinner, J. M. & Sweet, R. M. (1998). Integrated software for a macromolecular crystallography synchrotron beamline. *Acta Crystallog. sect. D*, **54**, 718-725.
73. Pflugrath, J. W. (1999). The finer things in X-ray diffraction data collection. *Acta Crystallog. sect. D*, **55**, 1718-1725.
74. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In *Methods in Enzymology* (Carter, C. W. J. & Sweet, R. M., eds), vol. 276, pp. 307-325, Academic Press, New York.
75. Collaborative Computational Project Number 4 (1994). The CCP4 suite: Programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760-763.
76. Jones, T. A. & Kjeldgaard, M. (1997). Electron-density map interpretation. *Methods Enzymol.* **277B**, 173-207.
77. Read, R. J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallog. sect. A*, **42**, 140-149.
78. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P. & Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905-921.
79. Parkinson, G., Vojtechovsky, J., Clowney, L., Brünger, A. T. & Berman, H. M. (1996). New parameters for the refinement of nucleic acid containing structures. *Acta Crystallog. sect. D*, **52**, 57-64.
80. Engh, R. A. & Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallog. sect. A*, **47**, 392-400.

Edited by T. Richmond

(Received 16 February 2001; received in revised form 7 May 2001; accepted 9 May 2001)