Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Connectivity-based *ab initio* phasing at different solvent levels

Andrei Fokine,[a] Natalia Lunina,[b] Vladimir Lunin[b] and Alexandre Urzhumtsev[a]*

[a]LCM3B, UMR 7036 CNRS, Faculté des Sciences, Université Henry Poincaré Nancy I, 54506 Vandoeuvre-lés-Nancy, France, and [b]Institute of Mathematical Problems in Biology, Russian Academy of Sciences, Pushchino, Moscow Region, 142290, Russia

Correspondence e-mail: sacha@lcm3b.uhp-nancy.fr

Received 1 November 2002
Accepted 24 February 2003

The connectivity-based phasing method has been applied independently to three neutron diffraction data sets obtained from the same crystal of $tRNA^{Asp}$–aspartyl-tRNA synthetase complex but at different concentrations of D_2O/H_2O, thus masking different components of the crystal. The obtained low-resolution images correlate perfectly with the solvent level.

1. Introduction

Direct phasing of a single set of experimental magnitudes at low resolution can provide useful information for structure determination when conventional phasing techniques are not applicable. The direct-phasing method based on the connectivity properties of Fourier syntheses (Lunin, Lunina & Urzhumtsev, 2000) is one of the most efficient. There are several examples in which this method has allowed the determination of molecular position, molecular shape and even secondary-structure elements (Lunin, Lunina, Petrova *et al.*, 2000; Lunin *et al.*, 2002).

Macromolecular crystals contain a high percentage of bulk solvent, the contribution of which to low-resolution structure factors is very important. When using neutron diffraction, the mean solvent scattering density can be varied over a large range from -0.1×10^{-14} to 7.0×10^{-14} cm Å^{-3} by changing the ratio of D_2O to H_2O in the mother liquor. As a consequence, the solvent scattering density can be made equal to the average scattering density of protein ($\sim 2.0 \times 10^{-14}$ cm Å^{-3}) or RNA ($\sim 4.0 \times 10^{-14}$ cm Å^{-3}) (Moras *et al.*, 1983).

The connectivity properties of low-resolution Fourier syntheses for nucleoprotein complexes depend strongly on the solvent density. Therefore, the image obtained by direct phasing, when using these properties as the criterion, should also depend on the solvent scattering density level.

The *ab initio* phasing procedure (Lunin, Lunina & Urzhumtsev, 2000) has been applied independently to three sets of neutron diffraction data from the same crystal of $tRNA^{Asp}$–aspartyl-tRNA synthetase complex (Moras *et al.*, 1983) in which the D_2O/H_2O contrast was varied, masking different components of the crystal. In these tests, the obtained low-resolution images reproduce perfectly the corresponding molecule: tRNA, protein or the whole complex.

2. Phasing method

The connectivity-based phasing method (Lunin, Lunina & Urzhumtsev, 2000) is based on the observation that the

topological properties of high-density regions of Fourier syntheses are different for properly phased syntheses and for those calculated with random phases.

Let $\rho(\mathbf{r})$ be a Fourier synthesis calculated over a finite grid, with the total number of grid points in the unit cell equal to N. The region $\Omega(\alpha)$ corresponding to the relative volume α is defined as a set of αN grid points of highest density. If the cutoff level α is chosen appropriately, the corresponding region of a 'correctly phased low-resolution macromolecular synthesis' is composed of a small number of isolated 'blobs'. The number of these blobs is expected to be equal to the number of molecules in the unit cell. The blob volume (measured in number of grid points) is the same for blobs linked by crystallographic symmetry and is approximately the same for those linked by non-crystallographic symmetry. On the other hand, randomly phased syntheses are likely to show infinite merged regions or/and a large number of small 'drops'.

The phasing procedure consists of the following steps.

(i) A large number of random phase sets are generated; the phases are generated uniformly at the beginning or according to a known phase distribution if this information is already available.

(ii) For each generated phase set, the Fourier synthesis is calculated using the experimental structure-factor magnitudes.

(iii) For the chosen cutoff level α, separated connected components of a high-density region are determined; their number and volume are calculated.

(iv) If the number of components and their volumes satisfy requested conditions, the corresponding phase set is considered to be admissible and is stored, otherwise the phase set is rejected; as an example, in the simplest mode the number of components is requested to be equal to a given value.

(v) After a reasonable number M ($\sim$100) of admissible phase sets have been selected, they are averaged; this step produces the corresponding 'best' phases $\varphi_{\rm best}(\mathbf{h})$ and their figures of merit $m(\mathbf{h})$, which reflect the spread of the phase values in selected phase sets. It should be noted that the optimal alignment of the phase sets in accordance with the permitted origin shifts (Lunin & Lunina, 1996) must be performed before averaging.

The figures of merit obtained are used to weight Fourier syntheses and to estimate the phase-density distribution for further generation of phases in the subsequent phase-improvement steps.

It should be remembered that the completeness of data sets, especially at low resolution, is crucial for the success of the method. Details of the method are given in Lunin, Lunina & Urzhumtsev (2000).

3. Test object

3.1. Cubic form of tRNA$^{\rm Asp}$–aspartyl-tRNA synthetase complex

The crystals of the cubic form of the tRNA$^{\rm Asp}$–aspartyl-tRNA synthetase complex (AspRS in the following) belong to space group $I432$ (48 symmetry operations), with unit-cell parameter a = 354 Å. The asymmetric unit contains one protein homodimer and two molecules of tRNA. The crystals contain 82% bulk solvent. The protein and tRNA molecules occupy 14 and 4% of the unit cell, respectively. The structure was solved by molecular replacement (Urzhumtsev *et al.*, 1994) using 15–8 Å resolution X-ray diffraction data and a model of the complex obtained previously for orthorhombic crystals at 3 Å resolution (Ruff *et al.*, 1991).

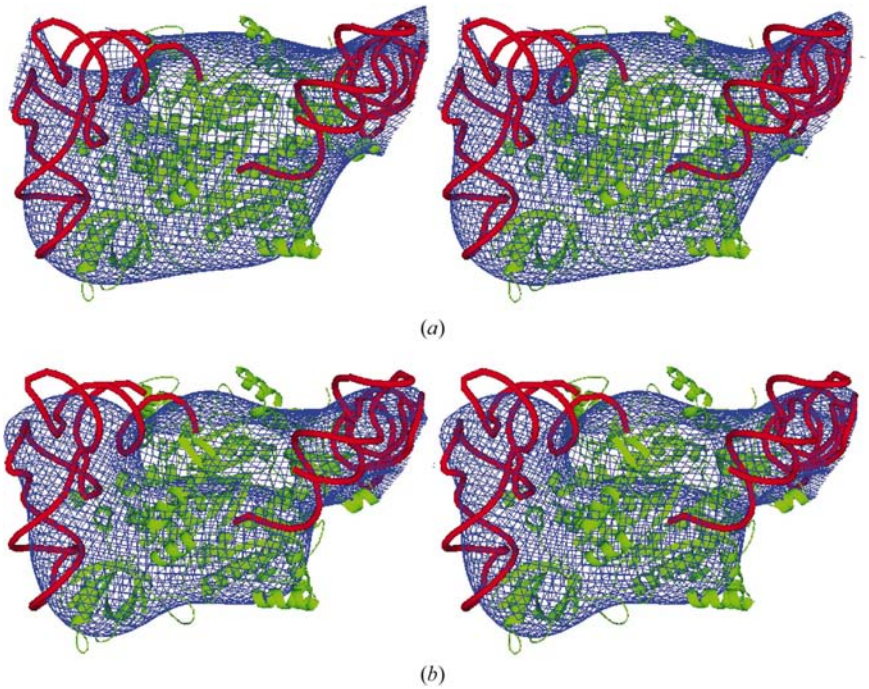

Figure 1
Synthesis calculated with exact phases $\varphi_{\rm comp}$ and experimental magnitudes $\mathbf{F}_{\rm comp}$ corresponding to the full complex molecules. The protein dimer is shown in green and tRNA molecules are shown in red. (*a*) 45 Å resolution; (*b*) 35 Å resolution.

3.2. Neutron diffraction data

Three sets of neutron diffraction data complete at low resolution (∞–24 Å) were measured for the same crystal of the complex with different concentrations of D_2O in the mother liquor (Moras *et al.*, 1983).

The first data set was measured without D_2O in the mother liquor. This corresponds to a very low solvent scattering density and a high contrast of the full complex molecule.

For the second data set, the concentration of D_2O was chosen to make the solvent scattering density equal to the average density of the protein. Thus, the protein was masked by the solvent and only the tRNA molecules could be 'seen' over the background of the solvent.

The third data set was measured with the solvent scattering density equal to the average density of tRNA. This data set corresponds to protein molecules only being 'seen' in the crystal.

These three data sets were independently used for direct phasing.

4. Analysis of low-resolution maps calculated with the 'exact' phases

To obtain a primary impression of how low-resolution syntheses look at different solvent contrasts, a series of syntheses (Figs. 1, 2 and 3) were calculated with the experimental magnitudes and the phases calculated from the known model ('exact' phases in the following). These maps were calculated as follows.

(i) The X-ray structure factors $\mathbf{F}_{\rm atom}$ were calculated from the atomic model of the AspRS complex.

(ii) The same atomic model was used to calculate the solvent mask (Jiang & Brünger, 1994) and corresponding Fourier coefficients $\mathbf{F}_{\rm mask}$.

(iii) The structure factors $\mathbf{F}_{\rm calc}$ were calculated as

$$\mathbf{F}_{\rm calc}(k_{\rm sol}) = \mathbf{F}_{\rm atom} + k_{\rm sol}\exp(-B_{\rm sol}s^2/4)\mathbf{F}_{\rm mask},$$

where $B_{\rm sol}$ = 50 Å^2 and had no practical influence on the synthesis values at such a low resolution; $k_{\rm sol}$ describes the average scattering density of the solvent and varied with the data sets as follows.

(1) To simulate a low solvent scattering density (as in the case of zero concentration of D_2O in the mother liquor), $k_{\rm sol}$ was set to 0 e Å^{-3}.

(2) To simulate masking of the protein molecules by bulk solvent, $k_{\rm sol}$ was taken as equal to the average electron density of a typical protein, which is 0.43 e Å^{-3}.

(3) To simulate masking of the tRNA molecules, $k_{\rm sol}$ was taken as 0.5 e Å^{-3}.

(iv) The experimental neutron structure-factor magnitudes associated with the phases of the structure factors $\mathbf{F}_{\rm calc}$ calculated with appropriate $k_{\rm sol}$ were used as the coefficients of the Fourier syntheses; the resulting maps are shown in Figs. 1, 2 and 3.

It should be noted that such a combination of neutron structure-factor magnitudes with simulated X-ray phases is only reasonable at low resolution, as is the case here.

Fig. 1 shows Fourier syntheses of different resolutions calculated with neutron diffraction magnitudes corresponding to the full AspRS complex and the phases calculated from the atomic model ($k_{\rm sol}$ = 0 e Å^{-3}). These syntheses correspond to the situation of low scattering density of the bulk solvent as in the case of zero concentration of D_2O in the mother liquor. The synthesis calculated at 45 Å resolution (37 independent reflections) shows the position of the molecules of the complex (Fig. 1*a*); the synthesis calculated at 35 Å resolution (74 reflections) also shows an approximate molecular envelope (Fig. 1*b*). The connectivity properties of these syntheses are summarized in Table 1. It is worthwhile to note that the regions of the highest density do not correspond to the positions of the tRNA molecules, whose scattering density is higher than that of the protein. Even at very high cutoff levels, the syntheses calculated at 45 Å resolution shows 48 isolated blobs in the unit cell located near to the centre of the AspRS complex. The synthesis calculated at 35 Å resolution and taken at high cutoff level additionally shows small drops (Table 1)

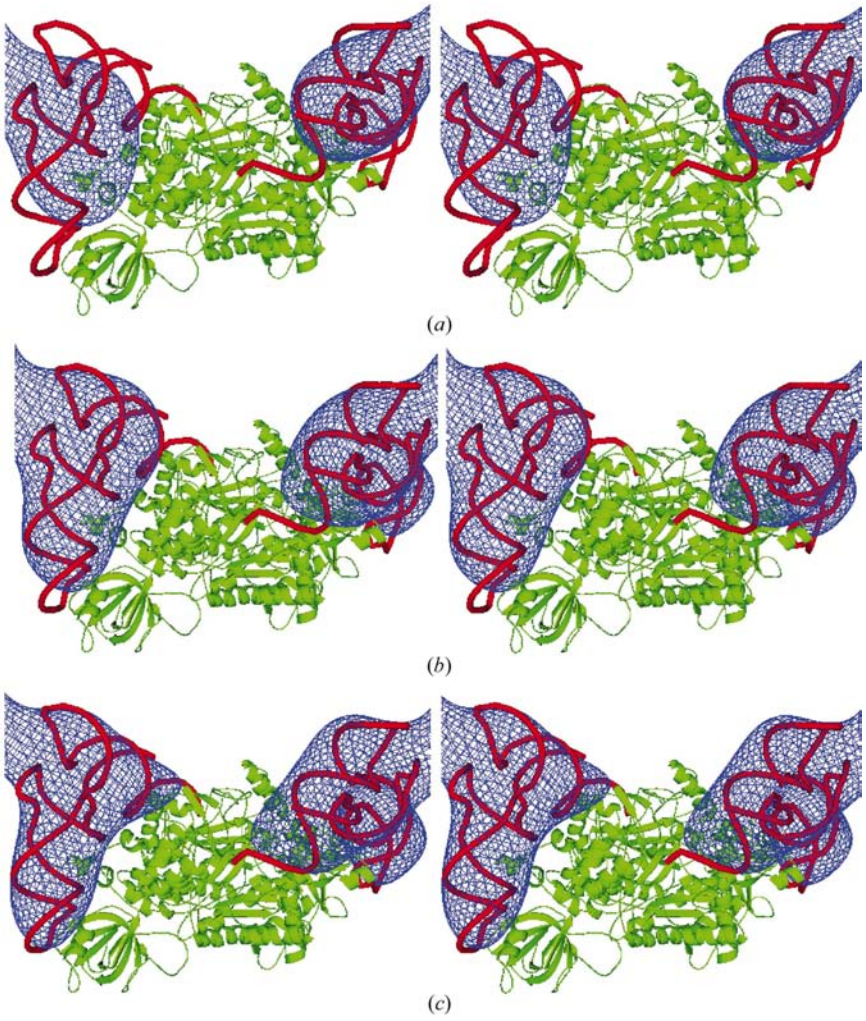

Figure 2
Synthesis calculated with exact phases $\varphi_{\rm rna}$ and experimental magnitudes $\mathbf{F}_{\rm rna}$ corresponding to the tRNA molecules. The protein dimer is shown in green and tRNA molecules are shown in red. (*a*) 45 Å resolution, (*b*) 35 Å resolution, (*c*) 30 Å resolution.

Table 1
Connectivity properties of the synthesis calculated with the exact phases φ_{comp} and experimental magnitudes $|\mathbf{F}_{comp}|$ corresponding to the AspRS complex.

The synthesis was calculated in the unit cell using a grid of dimensions 64 × 64 × 64. For details, see §4.

Synthesis resolution (Å)	Relative volume α (cutoff)	No. of connected regions	Multiplicity μ_k of connected areas and their volume N_k (measured in grid numbers)					
			μ_1	N_1	μ_2	N_2	μ_3	N_3
45	0.02	48	48	110	0	—	0	—
	0.03	48	48	163	0	—	0	—
	0.04	48	48	217	0	—	0	—
	0.05	12	12	1096	0	—	0	—
	0.07	12	12	1516	0	—	0	—
	0.10	12	12	2184	0	—	0	—
	0.20	1	1	52320	0	—	0	—
35	0.02	144	48	85	48	18	48	7
	0.03	48	48	164	0	—	0	—
	0.04	96	48	216	48	3	0	—
	0.05	96	48	267	48	6	0	—
	0.07	12	12	1524	0	—	0	—
	0.10	12	12	2184	0	—	0	—
	0.20	1	1	52320	0	—	0	—

Table 2
Connectivity properties of the synthesis calculated with the exact phases φ_{rna} and experimental magnitudes $|\mathbf{F}_{rna}|$ corresponding to the tRNA molecules.

The synthesis was calculated in the unit cell using a grid of dimensions 64 × 64 × 64. For details, see §4.

Synthesis resolution (Å)	Relative volume α (cutoff)	No. of connected regions	Multiplicity μ_k of connected areas and their volume N_k (measured in grid numbers)					
			μ_1	N_1	μ_2	N_2	μ_3	N_3
45	0.02	96	48	70	48	37	0	—
	0.03	48	48	163	0	—	0	—
	0.04	48	48	219	0	—	0	—
	0.05	48	48	273	0	—	0	—
	0.07	48	48	381	0	—	0	—
	0.10	24	24	1095	0	—	0	—
	0.20	24	24	2175	0	—	0	—
35	0.02	96	48	59	48	49	0	—
	0.03	96	48	90	48	73	0	—
	0.04	48	48	219	0	—	0	—
	0.05	48	48	274	0	—	0	—
	0.07	48	48	381	0	—	0	—
	0.10	48	48	544	0	—	0	—
	0.20	8	8	6558	0	—	0	—
30	0.02	144	48	55	48	43	48	14
	0.03	96	48	93	48	74	0	—
	0.04	48	48	220	0	—	0	—
	0.05	48	48	273	0	—	0	—
	0.07	48	48	383	0	—	0	—
	0.10	60	48	544	12	1	0	—

which also do not correspond to tRNAs and are located in the protein region. Therefore, the position of the tRNA molecules cannot be determined from these maps.

Fig. 2 represents the low-resolution syntheses calculated with experimental neutron magnitudes $|\mathbf{F}_{rna}|$ corresponding to tRNA molecules and the phases φ_{rna} calculated as described above with $k_{sol} = 0.43$ e Å^{-3}, thus masking the protein molecules. The synthesis calculated at 45 Å resolution (37 reflections) shows the positions of the tRNA molecules (Fig. 2*a*). The approximate form of the tRNA molecules can be seen (Fig. 2*b*) in the synthesis calculated at 35 Å resolution (74 reflections). The molecular envelope can be clearly determined from the synthesis calculated at 30 Å resolution using 110 reflections (Fig. 2*c*). The connectivity properties of these syntheses are summarized in Table 2. Taken at a very high cutoff level ($\alpha = 0.03$), the syntheses at 35 and 30 Å show 96 isolated blobs in the unit cell, corresponding to the tRNA molecules. At lower cutoff levels, the blobs merge together; there are 48 connected regions in the unit cell at cutoff $\alpha = 0.05$, each region now corresponding to two tRNA molecules associated with different AspRS molecules.

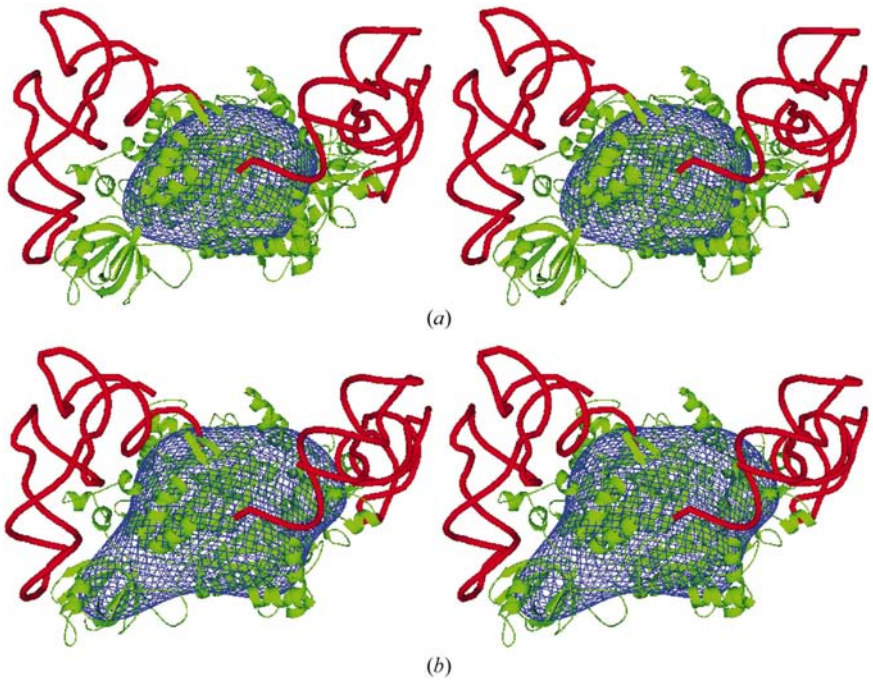

Figure 3
Synthesis calculated with coefficients $(-1)|\mathbf{F}_{prot}|\exp(\varphi_{prot})$, where φ_{prot} are exact phases and $|\mathbf{F}_{prot}|$ are experimental diffraction magnitudes corresponding to the protein molecules. The protein dimer is shown in green and tRNA molecules are shown in red. (*a*) 45 Å resolution, (*b*) 35 Å resolution.

Fig. 3 represents the syntheses calculated with coefficients $(-1)|\mathbf{F}_{prot}|\exp(i\varphi_{prot})$, where $|\mathbf{F}_{prot}|$ are experimental neutron magnitudes corresponding to the protein molecules and φ_{prot} are phases calculated with $k_{sol} = 0.5$ e Å^{-3} (equal to the average electron density of tRNA). The coefficient -1 was used because in this case the scattering density of the protein is lower than the bulk solvent. The synthesis calculated at 45 Å resolution shows the position of the protein molecules (Fig. 3*a*) and the synthesis calculated at 35 Å shows an approx-

Table 3
Connectivity properties of the synthesis calculated with the exact phases φ_{prot} and experimental magnitudes $|\mathbf{F}_{prot}|$ corresponding to the protein dimer.

The synthesis was calculated in the unit cell using a grid of dimensions 64 × 64 × 64 and flipped before studying the connectivity of the highest density regions. For details, see §4.

Synthesis resolution (Å)	Relative volume α (cutoff)	No. of connected regions	Multiplicity μ_k of connected areas and their volume N_k (measured in grid numbers)					
			μ_1	N_1	μ_2	N_2	μ_3	N_3
45	0.02	48	48	108	0	—	0	—
	0.03	48	48	165	0	—	0	—
	0.04	48	48	219	0	—	0	—
	0.05	48	48	270	0	—	0	—
	0.07	48	48	383	0	—	0	—
	0.10	12	12	2196	0	—	0	—
	0.20	12	12	4365	0	—	0	—
35	0.02	48	48	110	0	—	0	—
	0.03	48	48	165	0	—	0	—
	0.04	48	48	219	0	—	0	—
	0.05	48	48	274	0	—	0	—
	0.07	48	48	380	0	—	0	—
	0.10	48	48	549	0	—	0	—
	0.20	12	12	4384	0	—	0	—

imate corresponding envelope (Fig. 3*b*). At high cutoff levels both these syntheses show 48 isolated blobs, one per synthetase dimer (Table 3).

It should be noted that the results of this analysis were not used for further phasing.

5. *Ab initio* phasing of experimental neutron data sets

The direct-phasing procedure (Lunin, Lunina & Urzhumtsev, 2000) was applied independently to all three available experimental data sets. Only general information which is usually known *a priori*, such as the number of molecules in the unit cell and the relative unit-cell volume occupied by them, was used.

5.1. Phasing of the data set corresponding to the molecules of the full complex

The crystals contain one molecule of the AspRS complex in the asymmetric unit. Since the space group *I*432 has 48 symmetry operations, it is natural to expect that a low-resolution synthesis with correct phases would show 48 blobs of equal volume corresponding to the molecules of the complex.

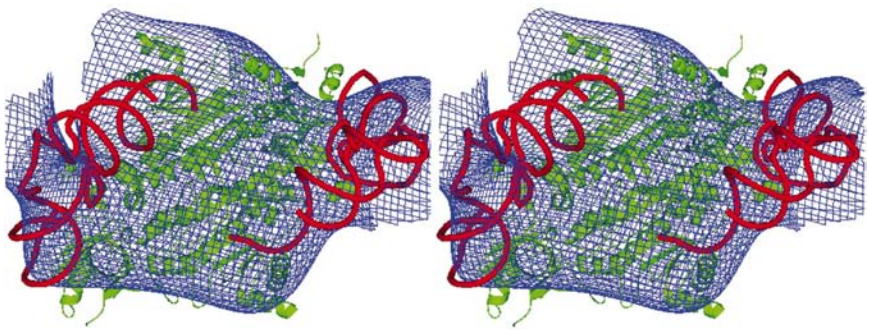

Figure 4
Synthesis calculated with experimental magnitudes $\mathbf{F}_{comp}$ corresponding to the full complex molecule and with phases obtained *ab initio* for these magnitudes. The protein dimer is shown in green and tRNA molecules are shown in red. The resolution of the synthesis is 45 Å.

Table 4
Connectivity properties of the synthesis calculated with the experimental magnitudes $|\mathbf{F}_{comp}|$ corresponding to the AspRS complex and the phases obtained *ab initio* for these magnitudes.

The synthesis was calculated in the unit cell using a grid of dimensions 64 × 64 × 64. For details, see §5.

Synthesis resolution (Å)	Relative volume α (cutoff)	No. of connected regions	Multiplicity μ_k of connected areas and their volume N_k (measured in grid numbers)					
			μ_1	N_1	μ_2	N_2	μ_3	N_3
45	0.02	48	48	108	0	—	0	—
	0.03	48	48	163	0	—	0	—
	0.04	48	48	218	0	—	0	—
	0.05	48	48	274	0	—	0	—
	0.07	48	48	383	0	—	0	—
	0.10	48	48	547	0	—	0	—
	0.20	1	1	52704	0	—	0	—

As the first step in the phasing procedure, all reflections in the resolution range ∞–45 Å were used (37 reflections) and phases were generated with uniform distribution. The selection criterion was formulated as follows: the high-density region occupying 5% of the unit cell (α = 0.05) must be composed of 48 connected regions of equal volume. The cutoff value of 5% was taken for the following reason: for macromolecular crystals with a usual macromolecular volume (about 40–50%), our experience suggests using a 10% region to study the connectivity. For AspRS crystals, the macromolecular volume is 18% and therefore a region half this size was taken.

100 appropriate phase variants were selected after about 90 000 generations. Table 4 gives the results of the connectivity analysis of the *ab initio* phased synthesis calculated at 45 Å (see Table 1 for comparison). At cutoff levels $0.03 < \alpha < 0.15$ this synthesis shows 48 isolated blobs corresponding to the position of the AspRS molecules (Fig. 4).

In the next steps, higher resolution reflections were included in phasing. For reflections used previously, phases were generated accordingly to a von Mises distribution determined by $\varphi_{j-1}^{best}(\mathbf{h})$ and $m_{j-1}(\mathbf{h})$ in the previous step (Lunin *et al.*, 2002); for new reflections, the phases were generated uniformly. The selection criterion was the same as at the first step, with the only difference being that other values of the cutoff level α ($0.03 < \alpha < 0.15$) were tried. The low-resolution images obtained in these steps did not differ greatly from the image obtained in the first step. A possible reason is that in this case the connectivity properties of the Fourier syntheses are determined by a small number of lowest resolution reflections which are very strong. Higher resolution reflections influence the syntheses only very weakly; as a consequence, the selected phase variants

contain both correct and incorrect phases for the additional higher resolution reflections. After averaging, the phases of these reflections have very low figures of merit and these reflections do not contribute to the final synthesis. The effective resolution of obtained images was approximately 45 Å. Some attempts were made to obtain better phases by changing the selection criterion. For example, we searched for maps in which 96 blobs are present at very high cutoff levels ($\alpha = 0.03$), corresponding to tRNA molecules with higher scattering density, while at lower cutoff levels ($\alpha = 0.05$) only 48 blobs corresponding to the full AspRS complex are visible. These attempts were unsuccessful; the procedure failed to find any admissible phase set among a large number (∼100 000) of generated sets. This is not surprising because the tRNA molecules are 'invisible' even in the low-resolution syntheses calculated with the exact phases (see §4 above).

5.2. Phasing of the data set corresponding to the tRNA molecules

The asymmetric unit contains two tRNA molecules linked by non-crystallographic symmetry. Since the space group *I*432 has 48 symmetry operators, it is natural to expect the low-resolution synthesis with correct phases to show 48 blobs of equal volume corresponding to the first tRNA molecule and a further 48 blobs of equal volume corresponding to the second tRNA molecule. The volumes of the blobs, which correspond to tRNA molecules linked by non-crystallographic symmetry, should be approximately equal.

In the first step of the phasing procedure, all reflections in the resolution range ∞–45Å were used (37 reflections). It was known *a priori* that the tRNA molecules occupy 4% of the unit cell; therefore, a slightly higher cutoff level of 3% was used (see the discussion above on the choice of the cutoff level). Phases were generated with uniform distribution. The selection criterion was formulated as follows.

The high-density region occupying 3% of the unit cell ($\alpha = 0.03$) must be composed of 48 connected regions of equal volume corresponding to the first tRNA molecule and a further 48 connected regions of equal volume corresponding to the second tRNA molecule; the ratio of these volumes must be between 0.7 and 1.0.

To obtain a solution, 270 000 random phase sets were generated; 100 sets from these satisfied the selection criterion. The selected variants were averaged to produce the 'best' phases $\varphi_1^{\rm best}(\mathbf{h})$ and the figures of merit $m_1(\mathbf{h})$. The *ab initio* phased synthesis at 45 Å (Fig. 5*a*) shows the approximate position of the tRNA molecules in the unit cell. The connectivity properties of this synthesis are represented in Table 5. At cutoff levels $0.02 < \alpha < 0.1$, this synthesis shows 96 blobs; the blobs corresponding to the different tRNA molecules do not merge even at cutoff $\alpha = 0.1$, while tRNA molecules occupy 0.04 of the unit cell.

In the second and further steps, the phases and their figures of merit defined previously were used as discussed above. The phasing procedure was always stopped after 100 admissible phase sets had been selected.

In the second step, reflections in the resolution range ∞–35 Å were used (74 reflections). The selection criterion was the same as at the previous step.

In the third step, the same reflections in the resolution range ∞–35 Å were used but a stronger selection criterion was used. At a high cutoff level there are 48 pairs of blobs; blobs in each pair are close to each other and an additional request was introduced that at the cutoff level of 0.05 the synthesis must show 48 merged blobs (remembering that, in fact, such pairs of blobs correspond to the tRNA molecules belonging to different and not the to same molecule of the complex).

Therefore, in the third step the selection criterion above was complemented by an

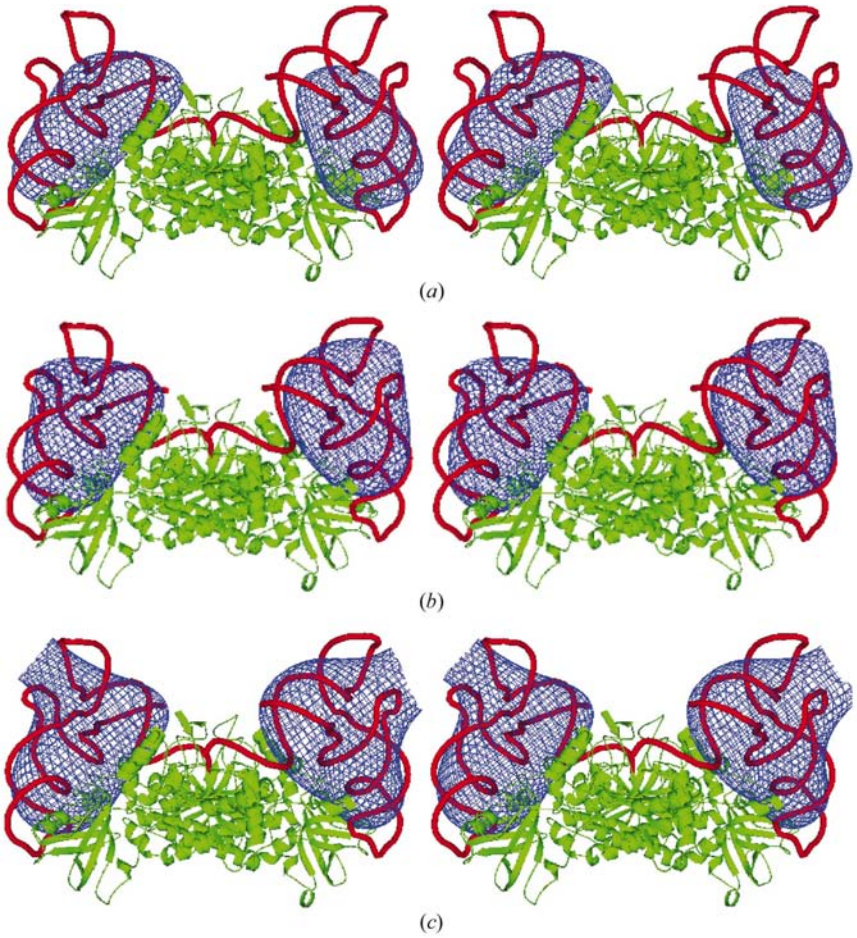

Figure 5
Synthesis calculated with experimental magnitudes $\mathbf{F}_{\rm rna}$ corresponding to the tRNA molecules and with phases obtained *ab initio* for these magnitudes. The protein dimer is shown in green and tRNA molecules are shown in red. (*a*) 45 Å resolution, (*b*) 35 Å resolution, (*c*) 30 Å resolution.

additional condition: the high-density region occupying 5% of the unit cell (α = 0.05) must be composed of 48 connected regions of equal volumes.

Fig. 5(*b*) shows the synthesis calculated at a resolution of 35 Å with the phases obtained after this step. The tRNA molecules are seen and some features of the molecular shape appear. The connectivity properties of this synthesis are summarized in Table 5. At cutoff levels $0.02 < \alpha < 0.05$, tRNA molecules are represented by separated blobs; at lower cutoff levels the blobs are merged, with one connected blob corresponding to two tRNA molecules.

The fourth and fifth steps were performed using reflections in the resolution range ∞–30 Å with the same selection criterion as before. The synthesis calculated after the fifth step is shown in Fig. 5(*c*). The approximate shape of the tRNA molecules can be determined after this synthesis.

Attempts to increase the resolution further using the same procedure were unsuccessful. The low-resolution images obtained in these steps did not show further new details compared with the image obtained in the fifth step. The effective resolution of these syntheses is close to 35 Å (compare Figs. 2 and 5 and Tables 2 and 5).

5.3. Phasing of the data set corresponding to the protein molecules

In the first step of phasing, reflections in the resolution range ∞–45 Å were used for which their phases were generated uniformly, as for two other data sets discussed above.

First of all, the same protocol was tried as previously for the corresponding tRNA data with the only difference the cutoff level was equal to 5%. The syntheses were selected if the high-resolution region was composed of two groups of regions with 48 symmetrically related regions in each and with a volume ratio between 0.7 and 1.0. This selection criterion corresponds to the situation when all protein molecules are separated from each other and each one is seen as an isolated blob. Such a search found only a few variants after 5 000 000 generations and was considered as unsuccessful. We then supposed that two independent protein molecules in the asymmetric unit are merged in a compact dimer and requested that the high-density region occupying 5% of the unit cell (α = 0.05) must be composed of 48 isolated blobs of equal volume.

100 selected variants were now found after about 50 000 generations. The *ab initio* phased synthesis at 45 Å (Fig. 6*a*) clearly shows the position of the protein homodimer.

In the second step of the phasing, all 74 reflections in the resolution range ∞–35 Å were used. The synthesis calculated at 35 Å (Fig. 6*b*) shows the approximate shape of the protein dimer. The connectivity properties of this synthesis (Table 6) are close to those of the synthesis using the exact phases (Table 3).

Table 5
Connectivity properties of the synthesis calculated with the experimental magnitudes $|\mathbf{F}_{rna}|$ corresponding to the tRNA molecules and the phases obtained *ab initio* for these magnitudes.

The synthesis was calculated in the unit cell using a grid of dimensions 64 × 64 × 64. For details, see §5.

Synthesis resolution (Å)	Relative volume α (cutoff)	No. of connected regions	Multiplicity μ_k of connected areas and their volume N_k (measured in grid numbers)					
			μ_1	N_1	μ_2	N_2	μ_3	N_3
45	0.02	96	48	70	48	40	0	—
	0.03	96	48	96	48	69	0	—
	0.04	96	48	122	48	97	0	—
	0.05	96	48	143	48	129	0	—
	0.07	96	48	203	48	180	0	—
	0.10	96	48	292	48	256	0	—
	0.20	1	1	52560	0	—	0	—
35	0.02	96	48	56	48	55	0	—
	0.03	96	48	84	48	80	0	—
	0.04	96	48	113	48	106	0	—
	0.05	96	48	143	48	130	0	—
	0.07	48	48	382	0	—	0	—
	0.10	48	48	546	0	—	0	—
	0.20	48	48	1095	0	—	0	—
30	0.02	96	48	57	48	54	0	—
	0.03	96	48	83	48	80	0	—
	0.04	48	48	218	0	—	0	—
	0.05	48	48	273	0	—	0	—
	0.07	48	48	381	0	—	0	—
	0.10	48	48	546	0	—	0	—

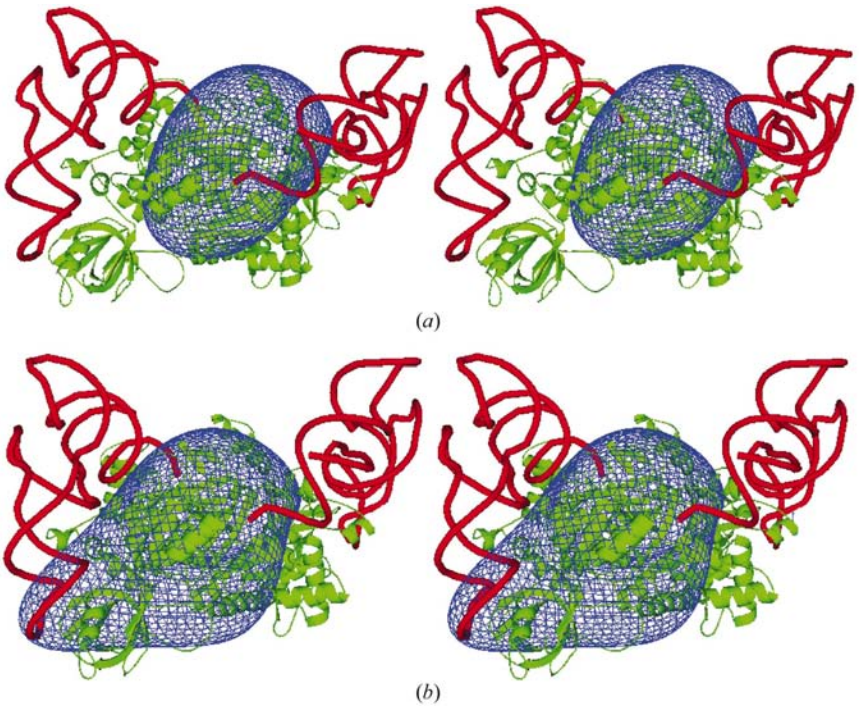

Figure 6
Synthesis calculated with experimental magnitudes $\mathbf{F}_{prot}$ corresponding to the protein molecules and with phases obtained *ab initio* for these magnitudes. The protein dimer is shown in green and tRNA molecules are shown in red. (*a*) 45 Å resolution, (*b*) 35 Å resolution.

Table 6
Connectivity properties of the synthesis calculated with the experimental magnitudes $|\mathbf{F}_{prot}|$ corresponding to the protein dimer and the phases obtained *ab initio* for these magnitudes.

The synthesis was calculated in the unit cell using a grid of dimensions 64 × 64 × 64 and was flipped before studying the connectivity of the highest density regions. For details, see §5.

Synthesis resolution (Å)	Relative volume α (cutoff)	No. of connected regions	Multiplicity μ_k of connected areas and their volume N_k (measured in grid numbers)					
			μ_1	N_1	μ_2	N_2	μ_3	N_3
45	0.02	48	48	108	0	—	0	—
	0.03	48	48	163	0	—	0	—
	0.04	48	48	218	0	—	0	—
	0.05	48	48	274	0	—	0	—
	0.07	48	48	383	0	—	0	—
	0.10	48	48	547	0	—	0	—
	0.20	12	12	4368	0	—	0	—
35	0.02	48	48	108	0	—	0	—
	0.03	48	48	163	0	—	0	—
	0.04	48	48	218	0	—	0	—
	0.05	48	48	274	0	—	0	—
	0.07	48	48	383	0	—	0	—
	0.10	48	48	547	0	—	0	—
	0.20	12	12	4368	0	—	0	—

Further attempts to increase the resolution using the same criterion but varying the cutoff level α in the range $0.05 < \alpha < 0.20$ were unsuccessful. The low-resolution images obtained in these further steps did not show further new details. The reasons for such behaviour are the same as discussed above for the phasing of the whole AspRS complex.

5.4. Map comparison

The closeness of two Fourier syntheses can be estimated by the map correlation coefficient (Lunin & Woolfson, 1993). Table 7 shows the correlation between the maps obtained *ab initio* and the maps calculated with the model structure factors simulated as explained in §4.

In general, the best correlation is obtained for pairs of corresponding synthesis. An exception is the 'protein' synthesis, which correlates equally well with the model-phased synthesis for the whole AspRS complex. This is not surprising because of the very high similarity of the corresponding synthesis with calculated structure factors (their correlation at the resolution of 35 Å is about 0.8). The same synthesis does not resemble the tRNA synthesis at all (zero correlation). As may be expected, all three *ab initio* phased syntheses at 45 Å resolution have some similarity to the 'complex' model-phased synthesis (first column). In contrast, neither of the *ab initio* phased syntheses for the protein and the complex is similar to the tRNA model-phased synthesis (second column).

Values of the map correlation show that the *ab initio* phased map with the 'protein' magnitudes at 35 Å is slightly closer to the calculated synthesis at 45 Å than to that at 35 Å. This illustrates that fact that the efficient resolution of *ab initio* phased syntheses is usually lower than the nominal resolution because of phase errors and weighting.

For the *ab initio* phased map with the 'protein' magnitudes at the resolution of 35 Å, its overall correlation with the 'complex' synthesis is still higher than the correlation with the 'protein' synthesis. A more detailed analysis was performed using the *MTRAP* program (N. Lunina, personal communication; the program estimates the percent of macromolecular atoms missed at various cutoff levels). It shows that the equipotential surfaces of the *ab initio* phased 'protein' syntheses surround the protein molecule better than the molecule of the whole complex (Fig. 7).

Table 7
Correlation between the maps phased *ab initio* and those calculated with the exact structure factors.

Exact structure factors $\mathbf{F}_{calc}(k_{sol}) = \mathbf{F}_{atom} + \mathbf{F}_{solv}(k_{sol})$ were calculated from the atomic model with the contribution of the bulk solvent determined by an appropriately chosen k_{sol}. For details, see §4.

	Calculated, resolution				
Ab initio, resolution	AspRS, k_{sol} = 0.00, 45 Å	tRNA, k_{sol} = 0.43, 45 Å	Protein, k_{sol} = 0.50, 45 Å	AspRS, k_{sol} = 0.00, 35 Å	Protein, k_{sol} = 0.50, 35 Å
AspRS, 45 Å	**0.80**	0.29	0.66		
tRNA, 45 Å	0.55	**0.75**	0.13		
Protein, 45 Å	0.72	0.05	**0.71**		
Protein, 35 Å	0.77		0.74	0.76	0.72

6. Conclusion

This study shows the influence of the density of bulk solvent on the low-resolution maps and, in particular, on the images obtained by the connectivity-based phasing method. An independent phasing with three data sets obtained from the same crystal but at different solvent levels showed three

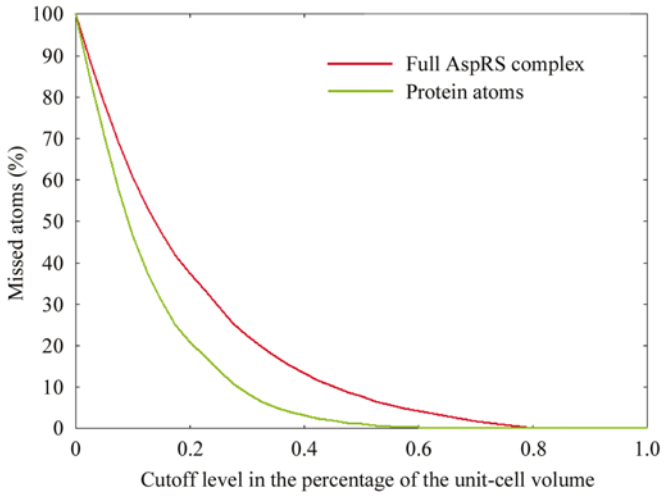

Figure 7
Percentage of atoms missed at different contours at the *ab initio* phases synthesis with experimental magnitudes $\mathbf{F}_{prot}$ corresponding to the protein molecules. The red curve corresponds to the atoms of the AspRS complex and the green curve corresponds to the protein atoms only. The relative volumes occupied by the full complex and by the protein are about 18 and 14%, respectively. These curves also reiterate that an atomic model is covered by the equipotential surface more or less completely when the volume selected by the surface is significantly larger than the formal volume of the macromolecule.

different images. The lowest resolution images provide the position of all three major crystal components: tRNA, the protein dimer and the AspRS complex. It should be emphasized that in order to obtain a crystallographic image, the phasing method needs only a single set of diffraction data and general information known *a priori*, such as the number of molecules in the unit cell and relative unit-cell volume occupied by molecules. Phase extension performed with the same connectivity-based approach quite reasonably reproduced the shape of tRNA molecules and of the protein dimer. Additional information is necessary in order to improve further the quality of the maps obtained.

The authors thank Lorraine Regional administration, Pole 'Intelligence Logicielle' CPER-Lorraine for financial support and CCH for the access to computer resources. The work of NL and VL was supported by grants 03-04-48155 and 01-07-90317, and was partially performed in the frame of collaboration CNRS-RAS. AF and AU are members of GdR 2417 CNRS.

References

Jiang, J.-S. & Brünger, A. T. (1994). *J. Mol. Biol.* **243**, 100–115.
Lunin, V. Y. & Lunina, N. L. (1996). *Acta Cryst.* A**52**, 365–368.
Lunin, V. Y., Lunina, N. L., Petrova, T. E., Skovoroda, T. P., Urzhumtsev, A. G. & Podjarny, A. D. (2000). *Acta Cryst.* D**56**, 1223–1232.
Lunin, V. Y., Lunina, N. L., Podjarny, A., Bockmayr, A. & Urzhumtsev, A. G. (2002). *Z. Kristallogr.* **217**, 668–685.
Lunin, V. Y., Lunina, N. L. & Urzhumtsev, A. G. (2000). *Acta Cryst.* A**56**, 375–382.
Lunin, V. Y. & Woolfson, M. M. (1993). *Acta Cryst.* D**49**, 530–533.
Moras, D., Lorber, B., Romby, P., Ebel, J. P., Giege, R., Lewit-Bentley, A. & Roth, M. J. (1983). *Biomol. Struct. Dyn.* **1**, 209–223.
Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J.-C. & Moras, D. (1991). *Science*, **252**, 1682–1689.
Urzhumtsev, A. G., Podjarny, A. D. & Navaza, J. (1994). *CCP4 Newsl.* **30**, 29–36.