

Acta Crystallographica Section D

**Biological  
Crystallography**

ISSN 0907-4449

## Density constraints and low-resolution phasing

**Alexandre G. Urzhumtsev, Natalia L. Lunina, Tatiana P. Skovoroda, Alberto D. Podjarny and Vladimir Y. Lunin**

Copyright © International Union of Crystallography

Author(s) of this paper may load this reprint on their own web site provided that this cover page is retained. Republication of this article or its storage in electronic databases or the like is not permitted without prior permission in writing from the IUCr.

## Density constraints and low-resolution phasing

Alexandre G. Urzhumtsev,<sup>a\*</sup>  
Natalia L. Lunina,<sup>b</sup> Tatiana P.  
Skovoroda,<sup>b</sup> Alberto D.  
Podjarny<sup>c</sup> and Vladimir Y.  
Lunin<sup>a,b</sup>

<sup>a</sup>LCM3B, UPRESA 7036 CNRS, Faculté des  
Sciences, Université Henri Poincaré Nancy I,  
54506 Vandœuvre-lès-Nancy, France,

<sup>b</sup>Institute of Mathematical Problems of Biology,  
Russian Academy of Sciences, Pushchino,  
Moscow Region 142292, Russia, and <sup>c</sup>Institute  
de Génétique et de Biologie Moléculaire et  
Cellulaire, CNRS, INSERM et Collège de France,  
Parc d'Innovation, BP 163, 67404 Illkirch  
CEDEX, CU de Strasbourg, France

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

Direct phasing needs additional information of a non-specific kind in order to select the correct phase set from all possible ones. This paper analyses the use of constraints which can be formulated in terms of electron-density values. One- and multi-dimensional histograms and connectivity properties are implemented as such constraints in density-modification procedures. These approaches usually cannot unambiguously select the best solution from a set of alternative phase variants. Nevertheless, they do allow the rejection of wrong solutions and the use of cluster analysis and averaging on the remaining variants provide a good starting point for further phase-refinement procedures.

Received 31 January 2000  
Accepted 27 June 2000

## 1. Introduction

Nowadays, the direct determination of atomic coordinates from X-ray diffraction data is routine work for relatively small molecules. However, the construction of a macromolecular model usually requires the calculation of the Fourier synthesis

$$\rho(\mathbf{r}) = (1/V) \sum_{\mathbf{h} \in S} F_{\mathbf{h}} \exp(i\varphi_{\mathbf{h}}) \exp[2\pi i(\mathbf{h}, \mathbf{r})] \quad (1)$$

at a limited resolution  $d = \min_{\mathbf{h} \in S} (1/|\mathbf{h}|)$  and its interpretation in terms of an atomic model. In order to calculate the distribution  $\rho(\mathbf{r})$ , phase values  $\varphi_{\mathbf{h}}$  need to be assigned to the corresponding experimental structure-factor magnitudes  $F_{\mathbf{h}}$ . In general, phasing methods use several sets of diffraction magnitudes measured under slightly different conditions: modified crystals (Perutz, 1956) or different wavelengths (for a recent review, see Hendrickson & Ogata, 1997). Otherwise, a known approximate model, usually atomic, of the whole molecule or a significant fraction of it is necessary (see, for example, the review by Rossmann, 1990). The problem of phase determination from a single set of magnitudes  $F_{\mathbf{h}}$ , also known as direct phasing, is still a challenge for macromolecular crystallography. A collection of reviews on this subject was prepared for the ECM-18 in Prague (Podjarny *et al.*, 2000).

Direct phasing needs additional information of a non-specific kind in order to select the correct phase set from all possible ones. In this paper, we restrict ourselves to the information which may be formulated directly in terms of electron-density values. To be more precise, we consider constraints applied to values of a truncated Fourier series (1) calculated at grid nodes in the unit cell. These density constraints can be conventionally divided in several major groups depending on the way in which they are imposed.

**Table 1**

Subsets of strong low-resolution reflections at 60 Å resolution for the model data in space group  $C222_1$ .

$S_0$ , 11 strongest reflections;  $S_1$ ,  $S_2$  and  $S_3$ , strongest reflections, origin-variable with respect to vectors  $\mathbf{u}_1$ ,  $\mathbf{u}_2$  or  $\mathbf{u}_3$ , respectively (for details, see §2).

<i>hkl</i>	021	023	025	110	111	112	113	114	115	116	130	131	132	133	201	203	204	Origin shift	
$S_0$			+	+	+	+				+	+				+	+	+	+	Not applied
$S_1$				+	+	+	+	+		+	+	+			+	+			$\mathbf{u}_1 = (\frac{1}{2}, 0, 0)$
$S_2$	+	+	+		+		+		+			+			+	+	+		$\mathbf{u}_2 = (0, 0, \frac{1}{2})$
$S_3$	+	+	+	+		+		+		+	+		+		+	+			$\mathbf{u}_3 = (\frac{1}{2}, 0, \frac{1}{2})$

### 1.1. Constraints on synthesis values at given points of the unit cell

Methods in this group use constraints on the density value based on the position of the grid point in the unit cell. For crystals with non-crystallographic symmetry, this can be the condition that the density values at symmetrically related points are equal (Rossmann, 1972). Another possible constraint is the equality of the density values at all points of the solvent region (Bricogne, 1974). This gives a basis for a number of solvent-flattening and density-averaging methods. In both examples, additional geometrical information is used, namely the knowledge of the molecular envelope and/or of the non-crystallographic symmetry operators.

### 1.2. Constraints on synthesis values

These methods are based on the knowledge of typical values of electron-density distributions calculated at a given resolution. For any crystal, overly high or low density values of the synthesis calculated on an absolute scale are not possible. The same argument can be reformulated for root-mean-square deviations. In a more general form, this information may be represented as a Fourier synthesis histogram. It defines both the range of possible values and the probabilities of finding them in the unit cell (see §6 for more details). Special attention should be given to the fact that such properties can vary with the synthesis resolution. For example, the property of the electron-density distribution of being non-negative everywhere was successfully used to develop direct methods (Karle & Hauptman, 1950) and some density-modification methods (Qurashi, 1953; Hoppe, 1962 and many others; reviewed, for example, by Podjarny *et al.*, 1996). Nevertheless, the truncated Fourier series do not necessarily reveal the non-negativity even when calculated with the true phases. Furthermore, the Fourier synthesis histograms are different at different resolutions.

### 1.3. Topological properties

Another way to apply constraints to a Fourier synthesis is to restrict the shape of the region containing points with specified density values. For example, at high resolution one would expect to see a continuous image for the main chain with branches corresponding to the side chains if a proper density cutoff level is chosen. At low resolution, one would expect to see a number of compact domains showing the molecular

packing. The expected shape of the molecule (if known) can be also considered as a constraint of this type. In these approaches, the criteria are purely geometric and absolute values of density distribution are not important.

This paper discusses the application of several density constraints for phasing at low resolution. Many of the tests discussed here were performed with calculated data in order to demonstrate clearly the character of the problem.

More applications, including those with experimentally obtained data, are discussed in the original papers devoted to particular methods and are referred to below.

## 2. Low-resolution crystallographic images

### 2.1. Maps and reflection sets

At the usual resolution of about 3 Å or higher at which most crystallographic macromolecular models are constructed the contrast of peaks in the electron-density maps is quite high and cannot be hidden by the absence of several low-resolution harmonics. Such a strong signal has for many years allowed crystallographers to avoid the particular problems of low-resolution data collection and phasing. In contrast, at a resolution of approximately 4 Å or lower crystallographic images do not have such strong details. Since they do not show detailed information, one could suppose that the syntheses are greatly influenced by a few of the strongest reflections and that it might be sufficient to phase these. On the other hand, data-set completeness was found to be crucial for the quality of low-resolution images (Podjarny *et al.*, 1981; Rayment, 1983; Urzhumtsev *et al.*, 1989; Urzhumtsev, 1991). Therefore, a special study was undertaken in order to check the results of phasing the strongest low-resolution reflections alone.

### 2.2. Test data

A test model was prepared simulating the position and a rough shape of the 50S particle from *Haloarcula marismortui* (H50S) phased directly at low resolution (Lunin *et al.*, unpublished work; the experimental data were provided by A. Yonath). The space group is  $C222_1$  and the unit-cell parameters are  $a = 210$ ,  $b = 300$ ,  $c = 500$  Å, with one molecule per asymmetric unit. This particular structure was chosen for tests because several different phasing methods had behaved abnormally. In order to obtain a test data set, five spheres approximating the shape and the position of the H50S particle were filled randomly by pseudo-atoms. Structure factors to 60 Å resolution (52 reflections in total) were calculated from this pseudo-atomic model and were used throughout this section to simulate experimental values (for more details, see Lunin *et al.*, 1999).

The synthesis calculated with this complete data set,  $S$ , showed eight well separated molecular envelopes in the unit cell (Fig. 1a), consistent with the eight symmetrically related molecules. Since 52 reflections is currently too many for an

exhaustive phase search (see §6.2), the 11 strongest reflections (subset  $S_0$ ) were chosen from the set  $S$ . The synthesis calculated with these reflections (using the exact phases) is shown in Fig. 1(b). The molecular envelopes lost their shape and, more importantly, the map no longer shows separated individual molecules.

### 2.3. Seminvariant study

Such deformation of molecular images can be explained in terms of seminvariant structure factors, *i.e.* those which do not change their magnitude and phase when an alternative origin permitted by the space group is used (Lunin *et al.*, 1999). Let  $\mathbf{u}$  be a permitted origin shift such that  $2\mathbf{u} = 0$  (modulo 1), which is true for most cases. Then every synthesis  $\rho(\mathbf{r})$  can be represented by a sum of two components,

$$\rho(\mathbf{r}) = \rho_{oi}(\mathbf{r}) + \rho_{ov}(\mathbf{r}) \quad (2)$$

with

$$\rho_{oi}(\mathbf{r}) = \frac{1}{2}[\rho(\mathbf{r}) + \rho(\mathbf{r} - \mathbf{u})], \quad (3)$$

$$\rho_{ov}(\mathbf{r}) = \frac{1}{2}[\rho(\mathbf{r}) - \rho(\mathbf{r} - \mathbf{u})]. \quad (4)$$

It is easy to demonstrate that these partial syntheses  $\rho_{oi}(\mathbf{r})$  and  $\rho_{ov}(\mathbf{r})$  correspond to the Fourier series over the seminvariants ('oi' stands for origin independent) and over other reflections ('ov' stands for origin variable), respectively,

$$\rho_{oi}(\mathbf{r}) = \frac{1}{V} \sum_{\mathbf{h} \in S_{oi}} F_{\mathbf{h}} \exp(i\varphi_{\mathbf{h}}) \exp[2\pi i(\mathbf{h}, \mathbf{r})], \quad \text{with } (\mathbf{h}, \mathbf{u}) = 0|_{\text{mod}1} \\ \text{for } \mathbf{h} \in S_{oi}, \quad (5)$$

$$\rho_{ov}(\mathbf{r}) = \frac{1}{V} \sum_{\mathbf{h} \in S_{ov}} F_{\mathbf{h}} \exp(i\varphi_{\mathbf{h}}) \exp[2\pi i(\mathbf{h}, \mathbf{r})], \quad \text{with } (\mathbf{h}, \mathbf{u}) \neq 0|_{\text{mod}1} \\ \text{for } \mathbf{h} \in S_{ov}. \quad (6)$$

The synthesis (3) shows the superimposition of two copies of a molecular image shifted by the vector  $\mathbf{u}$ . The addition of the extra molecular copies results in merged envelopes rather than in separated molecular images. The synthesis (4) shows the true image surrounded (and possibly distorted) by its flipped and shifted copies.

### 2.4. Application of the seminvariant decomposition

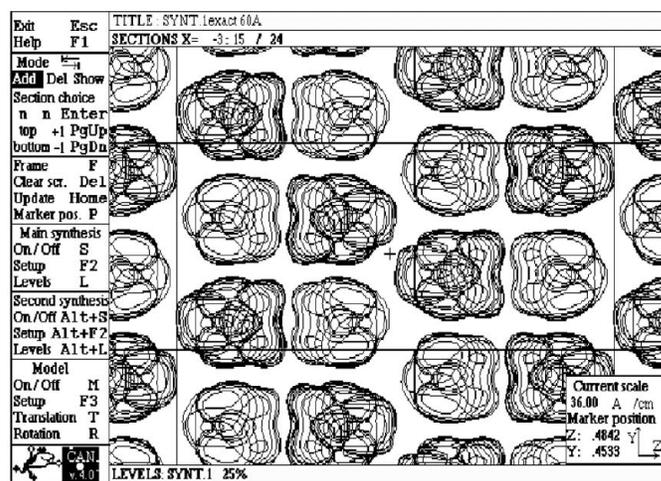
For the model data set discussed above there are three possible independent vectors  $\mathbf{u}$  for the origin shift:  $\mathbf{u}_1 = (\frac{1}{2}, 0, 0)$ ,  $\mathbf{u}_2 = (0, 0, \frac{1}{2})$  and  $\mathbf{u}_3 = (\frac{1}{2}, 0, \frac{1}{2})$ ; other origins in space group  $C222_1$  appear owing to the  $C$ -face centred cell. Three seminvariant-removed sets of reflections,  $S_1$ – $S_3$ , corresponding to  $\mathbf{u}_1$ ,  $\mathbf{u}_2$  and  $\mathbf{u}_3$ , respectively, as well as the set of 11 strongest reflections,  $S_0$ , are given in Table 1.

For each of these four data sets, a Fourier synthesis was calculated with the exact structure factors (exact in both magnitude and phase). Two extreme cases are presented in Fig. 2. If there is very little overlap of  $\rho(\mathbf{r})$  and  $\rho(\mathbf{r} - \mathbf{u})$ , as is the case for the vector  $\mathbf{u}_1$ , the  $\rho_{ov}(\mathbf{r})$  component shows eight connected molecular regions even at a quite low cutoff level. However, an overlap of  $\rho(\mathbf{r})$  and  $\rho(\mathbf{r} - \mathbf{u})$ , as with the vector  $\mathbf{u}_3$ , gives an endless continuous domain as was observed with direct phasing of the H50S particle.

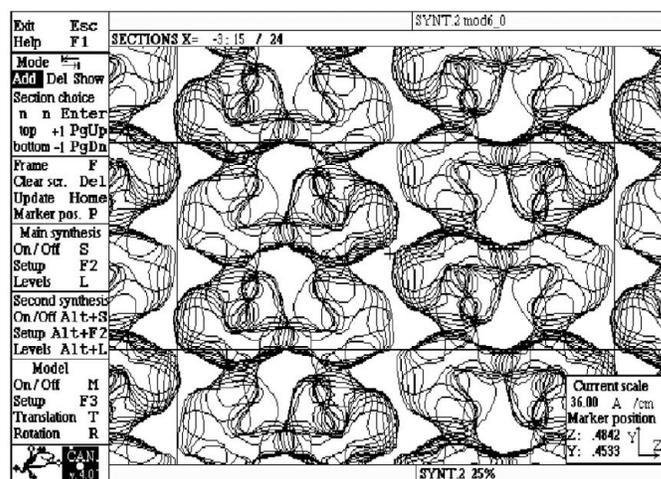
If the selected set of strong reflections used for the phasing is dominated by  $S_{oi}$  or by  $S_{ov}$  then the corresponding image may have features corresponding to one of these syntheses. Conversely, the set of structure factors for phasing can be chosen specifically to agree with a particular property and such selection will be discussed in §6. However, the complete set of reflections were used in the tests of §§3–5.

## 3. Density flattening at low resolution

For a Fourier synthesis calculated on a unit-cell grid, every grid point is characterized by its positional coordinates and by the value of the synthesis. When geometric information is also available, the position of the point can impose limitations on the synthesis value. This geometric information can be either



(a)



(b)

**Figure 1**

Fourier syntheses corresponding to the test model. (a) The synthesis calculated with all 52 low-resolution reflections of the 60 Å resolution zone reveals the expected number of globular components. An exhaustive phase search is impossible owing to too large a number of phase combinations (more than 1024). (b) The synthesis calculated with 11 strongest low-resolution reflections of the 60 Å resolution zone. Here, the envelopes are merged and envelope-based searches will fail to find the correct solution.

the position with respect to the symmetry elements (e.g. Lunin, 1989) or, more usually, whether or not a point belongs to the molecular region. The latter constraint forms the basis of the solvent-flattening procedure, which was introduced in real space by Bricogne (1974) and became very popular after an automatic procedure for molecular-envelope determination was suggested (Wang, 1985). This method of phase improvement is based on the hypothesis that the electron-density distribution is more or less uniform in the intermolecular region. At low resolution this hypothesis can be extended to the molecular region, suggesting that the whole image is essentially a binary function or a molecular mask, equal to 1 inside the molecular envelope and equal to 0 outside. As far as this second hypothesis holds, there is the possibility that envelope refinement could provide a simple phase-extension procedure with far less parameters.

These two hypothesis were tested using simulated low-resolution data for the 50S ribosomal particle from *Thermus thermophilus* (Urzhumtsev *et al.*, 1996; Podjarny *et al.*, 1998); X-ray diffraction data and an envelope obtained by electron

microscopy (Berkovitch-Yellin *et al.*, 1990) were provided by A. Yonath. The space group is  $P4_32_12$ , with unit-cell parameters  $a = b = 496$ ,  $c = 196$  Å.

The following procedure was applied: (i) a density distribution was calculated at a given resolution (60, 40 or 30 Å) and (ii) a molecular envelope was defined as a set of unit-cell points with density values above a given threshold; the threshold was chosen in such a way that the region with higher values of the density occupied a given percentage of the unit-cell volume and consisted of a single domain. Then either the density distribution was flattened in the solvent region with the density inside the molecular envelope left unchanged (soft modification) or, in order to test the hypothesis of a flat envelope, the density inside the envelope was also flattened (hard modification).

In both cases, structure factors were calculated from the modified density distribution and compared with those calculated from the original model. If, after the density modification, the phases lead to a map correlation coefficient of 0.5 or higher (Lunin & Woolfson, 1993), we consider the

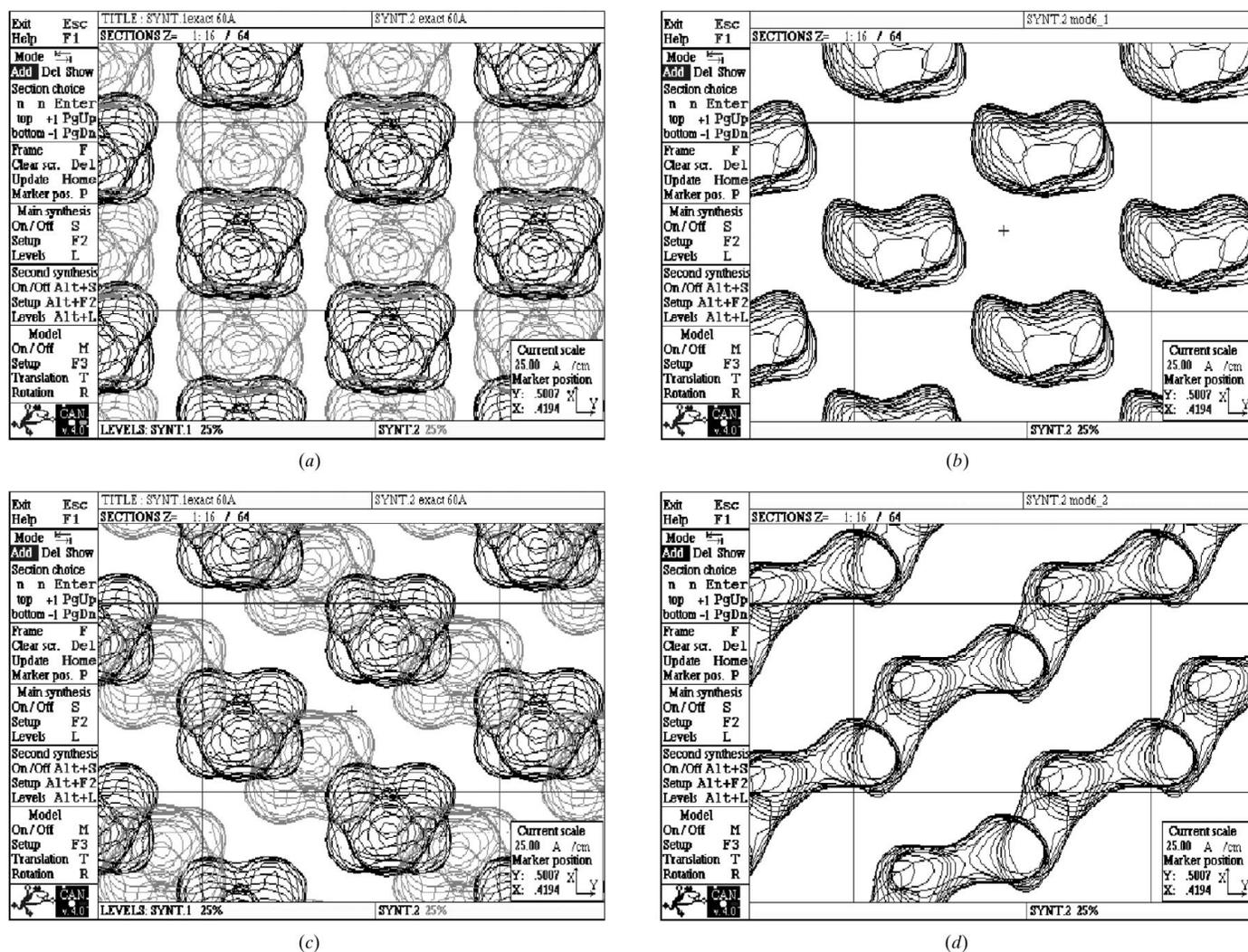


Figure 2

Superimposition of Fourier synthesis maps and their 'origin-variable' parts for two choices of possible origin (see Table 1). (a)  $\rho(\mathbf{r})$  (black) and  $\rho(\mathbf{r} - \mathbf{u}_1)$  (grey); (b)  $\rho_{ov}(\mathbf{r})$  for  $\mathbf{u}_1$ ; (c)  $\rho(\mathbf{r})$  (black) and  $\rho(\mathbf{r} - \mathbf{u}_3)$  (grey); (d)  $\rho_{ov}(\mathbf{r})$  for  $\mathbf{u}_3$ .

**Table 2**  
Solvent flattening at low resolution.

The amplitude and weighted phase correlations are given as a percentage. The columns correspond to different cutoff levels [the relative volumes  $p$  of the corresponding molecular regions, see (12), are pointed out in the column headers]. The rows correspond to resolution shells, with those shells indicating the initial resolution in italic. Cells corresponding to high correlations are shown in bold.

Density modification starting at 60 Å.

$d_{\min}$	$d_{\max}$	$N_{\text{ref}}$	Magnitude correlation					Phase correlation				
			$p = 0.1$	$p = 0.2$	$p = 0.3$	$p = 0.4$	$p = 0.5$	$p = 0.1$	$p = 0.2$	$p = 0.3$	$p = 0.4$	$p = 0.5$
<b>'Hard' modification</b>												
<i>80</i>	<i>500</i>	<i>40</i>	<b>91</b>	<b>96</b>	<b>92</b>	<b>87</b>	<b>83</b>	<b>98</b>	<b>99</b>	<b>97</b>	<b>92</b>	<b>90</b>
60	80	49	<b>83</b>	<b>88</b>	<b>89</b>	<b>84</b>	<b>74</b>	<b>96</b>	<b>97</b>	<b>96</b>	<b>93</b>	<b>85</b>
50	60	52	29	16	-2	-2	5	38	<b>55</b>	46	17	-33
45	50	48	1	-21	11	20	-3	29	3	-13	-42	-44
<b>'Soft' modification</b>												
<i>80</i>	<i>500</i>	<i>40</i>	<b>68</b>	<b>87</b>	<b>94</b>	<b>98</b>	<b>99</b>	<b>92</b>	<b>97</b>	<b>99</b>	<b>100</b>	<b>100</b>
60	80	49	<b>76</b>	<b>86</b>	<b>91</b>	<b>94</b>	<b>96</b>	<b>88</b>	<b>95</b>	<b>98</b>	<b>99</b>	<b>99</b>
50	60	52	-7	4	14	28	33	7	24	42	<b>54</b>	<b>59</b>
45	50	48	18	20	15	11	12	37	47	<b>52</b>	<b>50</b>	48

Density modification starting at 40 Å.

$d_{\min}$	$d_{\max}$	$N_{\text{ref}}$	Magnitude correlation					Phase correlation				
			$p = 0.1$	$p = 0.2$	$p = 0.3$	$p = 0.4$	$p = 0.5$	$p = 0.1$	$p = 0.2$	$p = 0.3$	$p = 0.4$	$p = 0.5$
<b>'Hard' modification</b>												
<i>80</i>	<i>500</i>	<i>40</i>	<b>91</b>	<b>98</b>	<b>92</b>	<b>89</b>		<b>99</b>	<b>99</b>	<b>96</b>	<b>94</b>	
60	80	49	<b>93</b>	<b>89</b>	<b>87</b>	<b>81</b>		<b>98</b>	<b>98</b>	<b>96</b>	<b>95</b>	
50	60	52	<b>88</b>	<b>92</b>	<b>79</b>	<b>71</b>		<b>95</b>	<b>96</b>	<b>93</b>	<b>89</b>	
45	50	48	<b>63</b>	<b>65</b>	<b>53</b>	<b>55</b>		<b>91</b>	<b>93</b>	<b>90</b>	<b>85</b>	
40	45	74	<b>59</b>	<b>72</b>	<b>60</b>	<b>64</b>		<b>86</b>	<b>96</b>	<b>90</b>	<b>85</b>	
35	40	118	16	14	5	19		<b>54</b>	44	2	-50	
30	35	211	10	3	6	17		49	16	-22	-55	
25	30	383	24	2	9	28		29	15	-22	-36	
20	25	876	8	15	15	9		9	-18	-30	-21	
<b>'Soft' modification</b>												
<i>80</i>	<i>500</i>	<i>40</i>	<b>76</b>	<b>91</b>	<b>97</b>	<b>99</b>	<b>99</b>	<b>95</b>	<b>99</b>	<b>100</b>	<b>100</b>	<b>100</b>
60	80	49	<b>85</b>	<b>95</b>	<b>98</b>	<b>99</b>	<b>99</b>	<b>95</b>	<b>98</b>	<b>100</b>	<b>100</b>	<b>100</b>
50	60	52	<b>72</b>	<b>90</b>	<b>97</b>	<b>99</b>	<b>99</b>	<b>79</b>	<b>94</b>	<b>99</b>	<b>100</b>	<b>100</b>
45	50	48	<b>51</b>	<b>77</b>	<b>90</b>	<b>95</b>	<b>96</b>	<b>78</b>	<b>94</b>	<b>98</b>	<b>99</b>	<b>100</b>
40	45	74	<b>41</b>	<b>68</b>	<b>83</b>	<b>90</b>	<b>94</b>	<b>69</b>	<b>89</b>	<b>96</b>	<b>98</b>	<b>99</b>
35	40	118	34	37	40	40	40	<b>57</b>	<b>69</b>	<b>77</b>	<b>80</b>	<b>81</b>
30	35	211	18	24	34	35	34	<b>51</b>	<b>66</b>	<b>70</b>	<b>70</b>	<b>70</b>
25	30	383	19	36	42	38	35	31	<b>50</b>	<b>52</b>	46	43
20	25	876	20	14	14	14	12	30	28	1	-18	-23

Density modification starting at 30 Å.

$d_{\min}$	$d_{\max}$	$N_{\text{ref}}$	Magnitude correlation					Phase correlation				
			$p = 0.1$	$p = 0.2$	$p = 0.3$	$p = 0.4$	$p = 0.5$	$p = 0.1$	$p = 0.2$	$p = 0.3$	$p = 0.4$	$p = 0.5$
<b>'Soft' modification</b>												
<i>80</i>	<i>500</i>	<i>40</i>	<b>85</b>	<b>96</b>	<b>99</b>	<b>100</b>	<b>100</b>	<b>97</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>
60	80	49	<b>87</b>	<b>97</b>	<b>99</b>	<b>99</b>	<b>100</b>	<b>95</b>	<b>99</b>	<b>100</b>	<b>100</b>	<b>100</b>
50	60	52	<b>76</b>	<b>92</b>	<b>97</b>	<b>98</b>	<b>99</b>	<b>90</b>	<b>98</b>	<b>100</b>	<b>100</b>	<b>100</b>
45	50	48	<b>69</b>	<b>87</b>	<b>95</b>	<b>96</b>	<b>97</b>	<b>88</b>	<b>97</b>	<b>99</b>	<b>100</b>	<b>100</b>
40	45	74	<b>68</b>	<b>89</b>	<b>96</b>	<b>97</b>	<b>98</b>	<b>89</b>	<b>96</b>	<b>99</b>	<b>99</b>	<b>100</b>
35	40	118	<b>77</b>	<b>91</b>	<b>96</b>	<b>98</b>	<b>98</b>	<b>91</b>	<b>97</b>	<b>99</b>	<b>99</b>	<b>100</b>
30	35	211	<b>76</b>	<b>87</b>	<b>92</b>	<b>95</b>	<b>97</b>	<b>93</b>	<b>97</b>	<b>98</b>	<b>99</b>	<b>100</b>
27	30	199	<b>52</b>	<b>59</b>	<b>63</b>	<b>62</b>	<b>61</b>	<b>77</b>	<b>83</b>	<b>87</b>	<b>87</b>	<b>87</b>
25	37	184	39	49	48	47	47	<b>64</b>	<b>76</b>	<b>79</b>	<b>79</b>	<b>78</b>
22	25	427	31	35	29	25	25	<b>67</b>	<b>74</b>	<b>72</b>	<b>68</b>	<b>67</b>
21	22	213	22	18	14	12	9	<b>58</b>	53	35	27	21

phase extension to be successful. Only one cycle of density modification was performed in each case. The results of these tests, presented in Table 2, can be summarized as follows.

placed in a bin  $k$  such that  $\rho_k < \rho(\mathbf{r}) < \rho_{k+1}$  and the corresponding bin counter  $\nu(k)$  is increased. After all points are treated, the normalized frequencies

(i) For a properly chosen cutoff level, the structure factors calculated after modification to the resolution of the starting synthesis are close to the correct values, thus validating the hypothesis of the flat envelope at low resolution.

(ii) For structure factors calculated in higher resolution shells, the magnitude residual is quite high, although in many cases the phases in the first resolution shell are of reasonable quality and can be used to increase the resolution.

(iii) The results of the hard modification are worse at 40 Å than at 60 Å and show that phase extension at this resolution can hardly ever be achieved by a simple refinement of the envelope, but requires knowledge of the density distribution inside the envelope.

#### 4. Constraints on the synthesis values; histograms

Constraints of this type do not depend on the position of the point in the unit cell and in particular do not depend on the character (solvent or protein) of a given point. The use of electron-density histograms provided the basis of a low-resolution *ab initio* phasing method developed by Lunin *et al.* (1990). This highlighted a number of features which were later found in many other approaches to *ab initio* phasing and it is, therefore, worth repeating the discussion here.

##### 4.1. Electron-density histograms

To define the electron-density histogram  $\nu(k)$  of a synthesis  $\rho(\mathbf{r})$  a set of density limits

$$\rho_1 < \rho_2 < \dots < \rho_k \quad (7)$$

are chosen to cover the whole range of expected values of a given synthesis class (*e.g.* of a given resolution). Then each value  $\rho(r)$  is



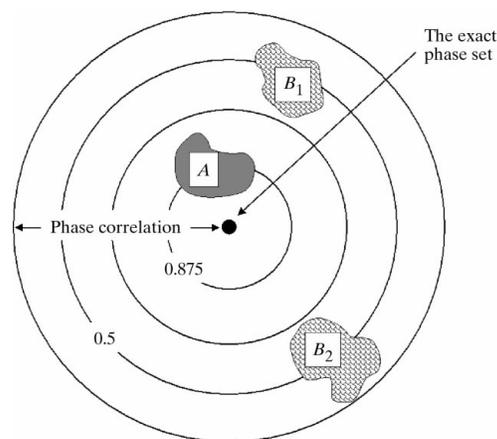
resolution phasing (see, for example, Lunin, Lunina, Petrova *et al.*, 2000). Although histogram-based phasing does have a number of difficulties such as a very high sensitivity to the set of structure factors used (§2) and the need for standard histograms, the tests show that such a method could have potential in direct phasing at low resolution.

#### 4.6. Synthesis alignment

Two phase sets, while being formally different, may present the same solution of the phase problem but corresponding to different choice of the origin and/or enantiomer. So, it is extremely important that the syntheses within a cluster are calculated with the same unit-cell origin before being averaged. The choice of the same origin and enantiomer may be performed by means of a map-alignment procedure (Lunin & Lunina, 1996).

#### 4.7. Histograms and wavelets

Recently, wavelet analysis has been introduced into the solution of the phase problem (Main, 1999; Main & Wilson, 2000; Wilson & Main, 2000; Lunin, 2000). The values of Fourier synthesis calculated at grid nodes may be considered to some extent as wavelet coefficients for a special type of wavelets (Lunin, 2000). There exist numerous links between wavelet-based and density-constraints-based approaches, which are beyond the scope of the paper, as well as some other approaches which use grid-density values as primary variables (Szöke *et al.*, 1997). Here, we mention only that the use of the histograms as restraints on wavelet coefficients has resulted in promising results in the phase extension (Main & Wilson, 2000; Wilson & Main, 2000).



**Figure 5**

A schematic presentation of a phase-variant distribution, selected on the basis of the histogram criterion. Concentric circles indicate phase sets equidistant from the correct solution. The cluster A corresponds to the group of variants marked A in Fig. 4. The group of variants marked B<sub>1</sub> and B<sub>2</sub> in Fig. 4 may actually consist of several separated clusters roughly equidistant from the correct solution.

## 5. Mixed approaches

### 5.1. Multiple histograms

The phasing method suggested by Zhang & Main (1990) combines different types of information discussed above. They modified the density values inside the molecular region in order to match the density histogram to a known one and simultaneously flattened the density values in the solvent region. Such density modification relies on knowledge of the molecular region and may be considered as a modification using two different histograms. The first one is a usual histogram linked to a molecular region. The density flattening in the solvent region may be considered as histogram matching with a singular histogram which allows only one value (the mean solvent density value) for all the points in the solvent region. A natural generalization of this procedure would be the use for the solvent region of a more sophisticated histogram which takes into account variations of solvent-density values.

Conversely, if the molecular region histogram is substituted for a singular histogram which allows only one value for all molecular region points, the procedure of density modification becomes equivalent to the flattening outside and inside molecular region, similar to the ‘hard’ modification discussed above in §3.

### 5.2. Distance-dependent histograms

Attempts to obtain a better model for the solvent distribution led Schoenborn (1988) to a model in which the solvent density  $\rho$  depends on the distance  $r$  from the molecule (see also Cheng & Schoenborn, 1990). This was further developed by Urzhumtsev & Podjarny (1995a), who supposed that the solvent-density distribution at points at a distance  $r$  from the molecular border is not a constant but may be described by a histogram  $H(\rho;r)$ . These histograms depend on the resolution  $d$  of the current Fourier synthesis. They could be used for density modification at any equidistant surface in a manner similar to the histogram matching.

To assign the solvent-density values more precisely, it is possible to use the observation that the density value could be lower for the points on a convex side of the envelope and be higher in cavities large enough to trap a solvent molecule. The obstacle for this is that at low resolution the exact molecular border is unknown. Nevertheless, these points can be discriminated on the base of their position with respect to a series of envelopes calculated at different resolutions  $d_m$ ,  $m = 1, \dots, M$  (Fig. 6). In this case, the histogram which describes the distribution of solvent-density values at the distance  $r$  from the precise molecular border may be replaced by a series of histograms  $H_{d1}(\rho;r_1), \dots, H_{dM}(\rho;r_M)$  operating with the distances  $r_1, \dots, r_M$  from the envelopes of corresponding resolution. Similarly, a set of histograms can also be calculated for points of the molecular region.

If the molecular envelope is known, for example, by electron microscopy, its position in the unit cell can be determined by molecular replacement (Urzhumtsev & Podjarny, 1995b). The molecular envelope can then be calculated at several

lower resolutions, for which the histograms  $H_{d_m}(\rho; r_m)$  are assumed to be known ( $r_m$  is the distance to the envelope calculated at the resolution  $d_m$ ). The following procedure could be used to reconstruct a density distribution at a given resolution  $d$  from this set of flat envelopes.

For each point  $r$  in the unit cell

- (i) calculate the distance  $r_m$  from the point  $\mathbf{r}$  to the molecular envelope at the resolution  $d_m$  for  $m = 1, \dots, M$ ;
- (ii) for each of the  $M$  envelopes estimate 'the probability' of different density values for this points as  $P_m(\rho) = H_{d_m}(\rho; r_m)$ ;
- (iii) calculate the combined probability distribution as

$$P(\rho) = P_1(\rho) \times \dots \times P_M(\rho); \quad (10)$$

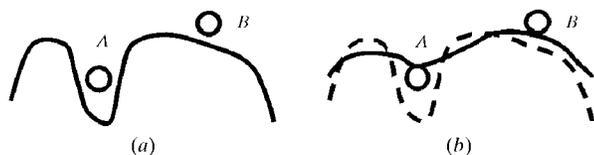
- (iv) find the most probable density value  $\rho_{\max}$   $P(\rho)$  and assign it to the point  $\mathbf{r}$ .

In this way, a set of flat envelopes and a corresponding set of histograms at a different resolutions can be used to calculate a modulated density distribution which could provide better phases, as shown in §3. In the case of a single envelope and a single histogram  $H(\rho; r)$ , the procedure will give a distance-dependent density distribution similar to that of Cheng & Schoenborn (1990).

Test calculations were performed on low-resolution data from the crystals of aldose reductase (Rondeau *et al.*, 1992). The  $H(\rho; r)$  histograms for a 6 Å density distribution were assumed to be known and were calculated for several molecular envelopes for resolutions in the range 20–6 Å. As well as using the correct position for the envelopes, tests were also carried out in which the envelopes were not positioned exactly. For the correctly placed 6 Å resolution envelope, the structure factors calculated from a flat density were reasonably good, but this was not the case when the envelope was wrongly positioned. With a single histogram at 6 Å resolution, the procedure improved both magnitudes and phases, mostly in the higher resolution zones. When all four histograms, corresponding to 6, 8, 11 and 20 Å resolution envelopes, were used, the magnitudes and the phases for the entire resolution range were well predicted (Fig. 7) improving the previous results (for further details, see Urzhumtsev & Podjarny, 1995a).

## 6. Topological features: connectivity

All mathematical methods for phasing are based on known properties of the density distribution at a given resolution. A



**Figure 6**  
An illustration of a different position of the points (atoms) with respect to molecular envelopes determined at different resolution. (a) the points  $A$  and  $B$  are outside the envelope calculated at the resolution  $d_1$  and are equidistant from it; (b) the same two points are shown with respect to a new envelope (continuous line) calculated at lower resolution  $d_2 > d_1$ ; the point  $A$  is inside and the point  $B$  is outside of this new envelope; the envelope at the resolution  $d_1$  is shown as a broken line for comparison.

Fourier synthesis can be viewed as a set of lines or surfaces joining points with the same density value. This suggests that it may be enough to recover a representative surface (or surfaces) rather than the exact density values at all points of the unit cell, providing a method to distinguish the correct solution from a number of noisy syntheses. A typical electron-density map should have the following properties.

- (i) For a high-resolution synthesis, the region selected with a very high cutoff level should show a set of atomic positions.
- (ii) For a high- or medium-resolution synthesis, the regions selected with a reasonably high cutoff level should be continuous and follow the main chain with branches showing the side chains.
- (iii) For a low-resolution synthesis, the regions selected by a reasonably high cutoff level should correspond to molecular envelopes.

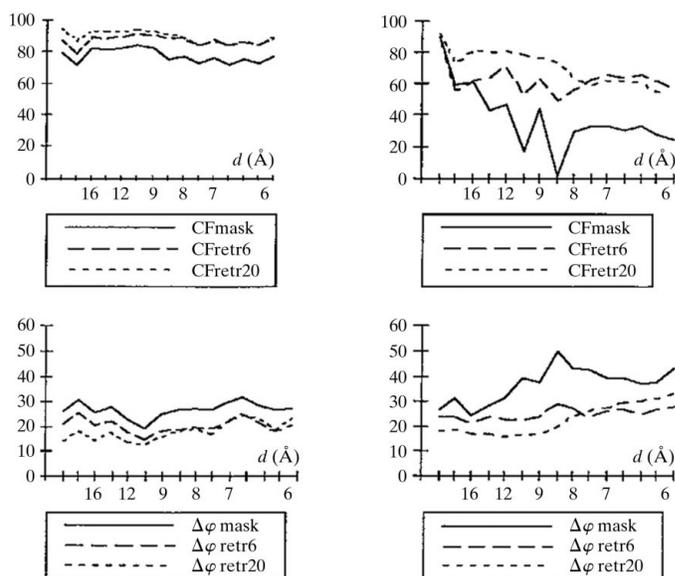
In this section, we discuss methods to employ such prior knowledge in low-resolution phasing and the search procedures and criteria used.

### 6.1. Connectivity and possible criteria

For a given cutoff level  $\kappa$  we define the set of points  $\Omega_\kappa$ , which may consist of several regions of the unit cell, by

$$\Omega_\kappa = \{\mathbf{r} : \rho(\mathbf{r}) > \kappa\}. \quad (11)$$

Some practical details on the estimation of the connectivity of such regions calculated on a periodic grid in a crystal can be



**Figure 7**  
A comparison of the structure factors retrieved from two-dimensional histograms for aldose reductase and those with the exact values: magnitude correlation (top), phase difference (bottom). The left column corresponds to the case of the exact envelope position and the right column to a shifted envelope. Continuous lines correspond to structure factors calculated from the flat envelope, broken lines to those from a single distance-dependent histogram with an envelope at 6 Å resolution and dotted lines to those from a set of histograms (envelopes calculated at resolutions from 6 to 20 Å).

found in Lunin *et al.* (1999) and Lunin, Lunina & Urzhumtsev (2000).

The function  $\rho(\mathbf{r})$  may be calculated on different scales and it is convenient to define  $\Omega_\kappa$  to be independent of the scale. In the following low-resolution studies we used two approaches, both based on the calculation of the volume for the corresponding region. Such a volume can be estimated from the number of points in the region, the unit-cell parameters and the total number of grids in the unit cell.

Firstly, for a given cutoff level, the percentage of the unit-cell volume that it defines can be calculated by

$$\frac{\text{volume of } \Omega_\kappa}{\text{total volume of the unit cell}} = p, \quad (12)$$

so that two syntheses can be analysed by comparing images corresponding to the same *relative* volume  $p$ . Secondly, it is possible to fix the *absolute* volume  $\alpha$  ( $\text{\AA}^3$  per residue) accounted for  $\Omega_\kappa$  per residue by

$$\frac{\text{volume of } \Omega_\kappa}{\text{number of residues in the unit cell}} = \alpha. \quad (13)$$

If the values of  $p$  or  $\alpha$  are fixed then a change of the  $F_{\text{obs}}$  scale changes the absolute value of  $\kappa = \kappa(\alpha)$  but does not alter the  $\Omega^\alpha = \Omega_{\kappa(\alpha)}$  region.

For a given synthesis, variation of  $\kappa$  changes the region  $\Omega_\kappa$  and its topological features. A slow decrease in the cutoff level can lead to the appearance of new domains corresponding to lower peaks which will merge into connected regions and finally give a large connected domain with a number of holes of decreasing size inside. Thus, numerical values can be assigned to the different topological characteristics of a given synthesis. For example,

(i) the cutoff value at which individual peaks merge into a connected region corresponding to a single molecule (if such an event can be observed);

(ii) the cutoff value at which a set of connected regions, one per molecule, merge into a single connected domain;

(iii) the number of connected components and their shape for a given cutoff level at which several syntheses can be compared; since shape comparison is a time-consuming procedure, it may be replaced by the weaker constraint of equality of volumes.

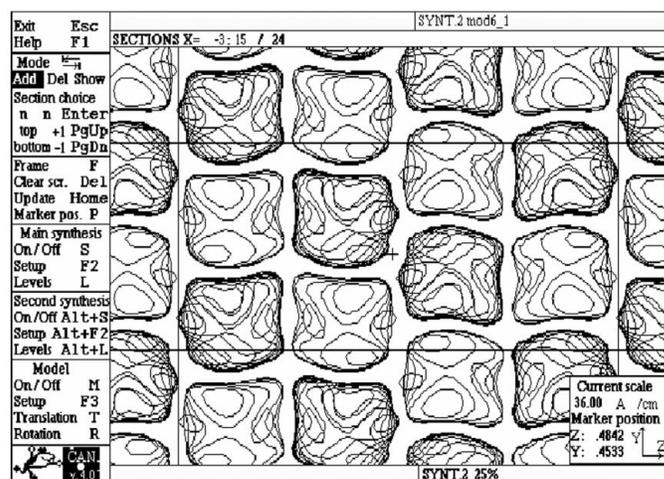
Such characteristics can be used as selection criteria for phase sets. The examples discussed below show that even in the simplest case of a single cutoff level such constraints are useful for phasing.

## 6.2. Exhaustive searches

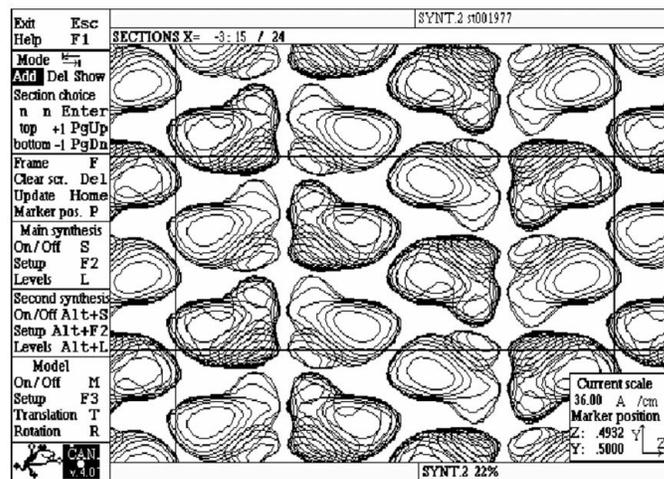
An exhaustive search in phase space can only be performed for a very small number of reflections, as the number of phase variants grows exponentially with the number of reflections. As discussed in §2, a synthesis calculated with a small number of reflections can have features which depend on the relative weight of seminvariant reflections and any selection criterion should therefore take this into account.

Conversely, for a given phasing criterion, an optimal set of structure factors can be chosen. In particular, when searching for the centre of molecules in the unit cell, it is preferable to calculate syntheses without seminvariant reflections. As an illustration, we have calculated Fourier syntheses for each of the data sets  $S_0$  to  $S_3$  (Table 1) as well as the complete data set to 60  $\text{\AA}$  resolution and the connectivity for these has been analysed at different cutoff levels. Remarkably, the size and number of connected components changed differently for the different data sets as the cutoff level was varied (Table 3).

For the same set of structure factors, a number of syntheses with wrong phases were also calculated and for a given cutoff level  $\kappa$  some of these also gave the correct number of connected domains of equal size. However, a slight variation in  $\kappa$  led either to their merging or to the appearance of noise, allowing the wrong phase sets to be identified. A numerical criterion for the selection of phase sets can therefore be



(a)



(b)

**Figure 8**

Low-resolution Fourier syntheses calculated with the  $S_1$  set of reflections (see Table 1). The exact synthesis is shown in (a) and the synthesis resulting from the systematic search with the connectivity criterion is shown in (b).

**Table 3**

Connectivity analysis for several sets of strong structure factors; model data in space group  $C222_1$  with eight molecules per unit cell.

Numbers in bold indicate the regions with designed connectivity.

$\Omega$ relative volume (%)	No. of connected components and their size (in grid points). No. of CS + No. of NCS reflections is given				
	26 + 26 reflections of the 60 Å zone	$S_0$ : 6 + 5 strongest reflections	$S_1$ : 2 + 7 $\mathbf{u}_1$ -variable reflections	$S_2$ : 5 + 5 $\mathbf{u}_2$ -variable reflections	$S_3$ : 7 + 4 $\mathbf{u}_3$ -variable reflections
5	$8 \times 148 + 8 \times 80$ + $8 \times 43 + 8 \times 35$	$8 \times 260 + 8 \times 46$	$\mathbf{8} \times 306$	$\mathbf{8} \times 306$	$\mathbf{8} \times 306$
10	$8 \times 422 + 8 \times 190$	$4 \times 1224$	$\mathbf{8} \times 612$	$4 \times 1224$	$4 \times 1224$
15	$\mathbf{8} \times 918$	$2 \times 3672$	$\mathbf{8} \times 918$	$4 \times 1836$	$4 \times 1836$
20	$\mathbf{8} \times 1224$	$2 \times 4896$	$\mathbf{8} \times 1224$	$4 \times 2448$	$4 \times 2448$
25	$\mathbf{8} \times 1530$	$2 \times 6120$	$\mathbf{8} \times 1530$	$2 \times 6120$	$1 \times 12240$
30	$4 \times 3872$	$1 \times 14688$	$1 \times 14688$	$1 \times 14688$	$1 \times 14688$

formulated in which a phase set is accepted as a possible solution for a given set of structure factors if it gives the correct number of equal connected domains at the lowest cutoff level. The examples (Table 3) show that this criterion is quite sensitive to the set of structure factors and in fact is not applicable at all for some data sets, e.g.  $S_0$ , the set of strongest reflections, where even the exact phases do not provide a synthesis with eight similar connected domains. On the other hand, the set  $S_1$  is particularly appropriate for this criterion and, in the synthesis calculated with the exact phases, the corresponding eight domains appear at high levels and do not merge until the cutoff level is quite low.

In order to check the selection power of this criterion, an exhaustive search procedure was applied for all four data sets  $S_0$ ,  $S_1$ ,  $S_2$  and  $S_3$ . For each set, all possible phases sets were checked (both values for centrosymmetric reflections and four values,  $\pm\pi/4$  and  $\pm 3\pi/4$ , for non-centrosymmetric reflections). For every phase set, the corresponding map was calculated and the lowest value for the cutoff level  $\kappa$  was found such that the image had eight connected domains (owing to crystallographic symmetry, all of them had the same volume). The synthesis with the lowest value of  $\kappa$  was accepted as the solution and is shown in Fig. 8. The map correlation with the exact synthesis (also shown in Fig. 8) is 77%.

Several important conclusions can be drawn. Firstly, the importance of the choice of structure factors must be stressed once again. More importantly, however, the tests show the usefulness of topological information in direct phasing. This could be developed further by using larger data sets which would require a more sophisticated search technique. One possible approach is discussed briefly in the following sections and the details can be found in Lunin, Lunina & Urzhumtsev (2000).

### 6.3. Random searches

As the number of phased reflections increases, a crystallographic image will show not only the molecular position, as in the previous case, but also the molecular shape. When using

calculated data, it is difficult to model accurately the influence of experimental errors or the contribution of bulk solvent and tests with experimental data are therefore more significant. We have studied the use of topological information in direct phasing with experimentally measured structure-factor magnitudes for several cases where the atomic model was known. In all the tests, structure-factor phases calculated from the corresponding refined atomic model served as a reasonably good approxima-

tion to the phases of low-resolution reflections (Podjarny & Urzhumtsev, 1997).

### 6.4. Search procedure and selection criterion

A comprehensive search becomes impractical with a large number of reflections and either a random search or some other more systematic approach such as the use of a regular grid in the space of all phase sets (Gilmore *et al.*, 1999) must be taken. Here, we have used a random search and, in order to further accelerate the search procedure, the connectivity criterion has been modified so that a single cutoff level was used in the analysis. In most of our tests, we have found that a suitable cutoff level,  $\kappa_{25}$ , corresponds to the region with a volume equal to  $25 \text{ \AA}^3$  per residue. Obviously, this does not correspond to the volume of the protein molecule but simply provides non-overlapping peaks corresponding to different molecules in a low-resolution Fourier synthesis. In general, if the cutoff level is lower, the envelopes for individual molecules begin to merge, although some exceptions will be discussed.

For each randomly generated phase set the Fourier synthesis was calculated and the number and size of connected regions for the cutoff level  $\kappa_{25}$  calculated. The phase set was selected only if the number of regions was equal to the number of molecules in the unit cell and if they were of approximately equal volume. As might be expected, a random search with such a selection criterion cannot give a single solution and statistical analysis of the selected syntheses is necessary. From a number of test applications we found that two different cases were possible, examples of which are discussed in the following sections.

### 6.5. Normal case: topologically based phasing for $\gamma$ -crystallin IIIb

$\gamma$ -Crystallin IIIb is a protein of 173 residues which crystallizes in space group  $P2_12_12_1$ , with unit-cell parameters  $a = 58.7$ ,  $b = 69.5$ ,  $c = 116.9 \text{ \AA}$  and two molecules per asymmetric unit. Among 100 000 randomly generated phase sets calculated to  $24 \text{ \AA}$  resolution (28 reflections), 576 provided a synthesis

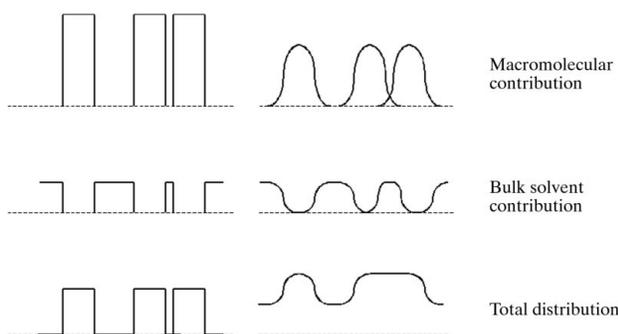
satisfying the given criterion, *i.e.* the  $\kappa_{25}$  cutoff level showed eight connected regions of very similar volume (a 10% discrepancy between the volume of the two quartets of domains was allowed because of the non-crystallographic symmetry linking them). This ensemble of selected phase sets had a higher concentration of good phase variants than the original random ensemble. The averaging of the 576 selected variants gave a map with a correlation coefficient 0.89 with the exact map at 24 Å resolution. More details of this test are given in Lunin, Lunina & Urzhumtsev (2000) and Lunin, Lunina, Petrova *et al.* (2000).

### 6.6. Special case: topologically based phasing for RNase Sa

The complete set of low-resolution data for RNase Sa (Ševčík *et al.*, 1991) was kindly provided by E. Dodson. The protein crystallizes in space group  $P2_12_12_1$ , with two molecules per asymmetric unit and unit-cell parameters  $a = 64.9$ ,  $b = 78.3$ ,  $c = 38.8$  Å. It contains 96 residues and the complete data set to 18 Å resolution consists of 29 reflections.

In contrast to the case for  $\gamma$ -crystallin IIIb, we found that the synthesis calculated with experimental magnitudes and the model phases does not show eight separated domains of approximately the same volume at any cutoff level. This is because of the dense packing of the molecules, possibly coupled with the contribution from bulk solvent (for a schematic illustration, see Fig. 9). This is confirmed by the observation that the synthesis calculated with model magnitudes and phases does show the eight separate envelopes. A study was performed to check whether the use of this idealized condition in the work with experimental data will provide the correct solution.

The calculations were performed at 18 Å resolution and a phase set was selected if at  $\kappa_{25}$  the corresponding synthesis showed eight connected domains of similar size. From 100 000 randomly generated phase sets, 558 were selected and the syntheses averaged. The correlation of the averaged map with the correct map at 18 Å resolution was 0.75 (0.91 at 24 Å). At a high cutoff level the final map showed eight separate molecular envelopes corresponding to the molecular positions. More details of this test can also be found in Lunin, Lunina & Urzhumtsev (2000).



**Figure 9**

A schematic illustration of electron-density distributions (left) and of their images at low resolution (right). Contribution of bulk solvent may lead to the merging of macromolecular envelopes.

### 6.7. Topologically based phasing: conclusions

We have found that the topological criterion expressed through the number of connected domains of similar size does allow direct phasing at low resolution in quite different cases. As with other selection criteria for direct-phasing procedures, the criterion does not provide a single solution but enriches a population of phase sets by those close to the correct solution. The topological constraints are quite weak and the selected phase sets have very different phase quality. However, a simple averaging over the selected phase sets gives a map which can be used for model positioning, for phase improvement and for some preliminary envelope analysis. As before, cluster analysis can be applied to the selected data sets in order to improve the map further.

## 7. Conclusions

A variety of density constraints have been shown to be useful for low-resolution phasing, both for phase improvement and for direct phase determination. A number of common features have been discovered and, in particular, no search criteria has been found to select unambiguously the correct phase set. Nevertheless, the selection of variants from a random population leads to a new population with a higher proportion of good phase sets. Simple averaging of these phase sets can give a reasonable macromolecular image and cluster analysis can further improve its quality. The strategy for phase searching should therefore be the statistical treatment of a relatively large number of selected variants rather than a search for the single best variant.

Since density-constraint phasing methods do not use explicit macromolecular model, they are therefore less influenced by the problem of bulk solvent. Recent results show the potential of topological criteria in a direct phasing protocol that may eventually lead to automated structure determination.

The authors thank C. Lecomte for his interest in this work. The work was supported in part by RFBR grants 97-04-48319 and 99-07-90461, by the CNRS through a fellowship (VYL) and the collaborative project RAS (VL)–CNRS (ADP, AGU), by the UHP, Nancy, by the Institut National de la Santé et de la Recherche Médicale and the Hôpital Universitaire de Strasbourg (HUS). They are also grateful to the groups headed by Yu. Chirgadze, D. Moras, E. Dodson and A. Yonath for providing the experimental diffraction data. The authors thank Dr J. Wilson for her valuable help in improving the manuscript.

## References

- Berkovitch-Yellin, Z., Wittmann, H. G. & Yonath, A. (1990). *Acta Cryst.* **B46**, 637–643.
- Bricogne, G. (1974). *Acta Cryst.* **A30**, 395–405.
- Cheng, X. & Schoenborn, B. P. (1990). *Acta Cryst.* **B46**, 195–208.
- Chirgadze, Yu. N., Brazhnikov, E. V., Garber, M. B., Nikonov, S. V., Fomenkova, N. P., Lunin, V. Yu., Urzhumtsev, A. G., Chirgadze, N. Yu. & Nekrasov, Yu. V. (1991). *Dokl. Acad. Nauk SSSR*, **320**, 488–491.

- Gilmore, C., Dong, W. & Bricogne, G. (1999). *Acta Cryst.* **A55**, 70–83.
- Hendrickson, W. A. & Ogata, C. M. (1997). *Methods Enzymol.* **276**, 494–522.
- Hoppe, W. (1962). *Acta Cryst.* **15**, 13–17.
- Karle, J. & Hauptman, H. (1950). *Acta Cryst.* **3**, 181–187.
- Lunin, V. Y. (1988). *Acta Cryst.* **A44**, 144–150.
- Lunin, V. Y. (1989). *Acta Cryst.* **A45**, 501–505.
- Lunin, V. Y. (2000). *Acta Cryst.* **A56**, 73–84.
- Lunin, V. Y. & Lunina, N. L. (1996). *Acta Cryst.* **A52**, 365–368.
- Lunin, V. Y., Lunina, N. L., Petrova, T. E., Skovoroda, T. P., Urzhumtsev, A. G. & Podjarny, A. D. (2000). *Acta Cryst.* **D56**, 1223–1232.
- Lunin, V. Y., Lunina, N. L. & Urzhumtsev, A. G. (1999). *Acta Cryst.* **A55**, 916–925.
- Lunin, V. Y., Lunina, N. L. & Urzhumtsev, A. G. (2000). *Acta Cryst.* **A56**, 375–382.
- Lunin, V. Y., Urzhumtsev, A. G. & Skovoroda, T. P. (1990). *Acta Cryst.* **A46**, 540–544.
- Lunin, V. Y. & Vernoslova, E. A. (1991). *Acta Cryst.* **A47**, 238–243.
- Lunin, V. Y. & Woolfson, M. M. (1993). *Acta Cryst.* **D49**, 530–533.
- Main, P. (1999). *Abstracts of the XVIIIth IUCr Congress and General Assembly*, p. 183. Abstract M12.BB.001.
- Main, P. & Wilson, J. (2000). *Acta Cryst.* **D56**, 618–624.
- Perutz, M. E. (1956). *Acta Cryst.* **9**, 867–873.
- Podjarny, A. D., Lunina, N., Urzhumtsev, A. G., Vernoslova, E. A., Petrova, T. & Lunin, V. (1998). *Abstracts of the American Crystallographic Association Meeting*, p. 74. Abstract 11.03.08.
- Podjarny, A. D., Rees, B. & Urzhumtsev, A. G. (1996). *Methods in Molecular Biology*, Vol. 56, *Crystallographic Methods and Protocols*, edited by C. Jones, B. Milloy & M. R. Sanderson, pp. 205–226. Totowa, New Jersey: Humana Press.
- Podjarny, A. D., Schevitz, R. W. & Sigler, P. B. (1981). *Acta Cryst.* **A37**, 662–668.
- Podjarny, A. D. & Urzhumtsev, A. G. (1997). *Methods Enzymol.* **276**, 641–658.
- Podjarny, A., Urzhumtsev, A. & Usón, I. (2000). *Reviews on Direct Phasing for the XVIIIth European Crystallography Meeting*. In the press.
- Qurashi, M. M. (1953). *Acta Cryst.* **6**, 103.
- Rayment, I. (1983). *Acta Cryst.* **A39**, 102–116.
- Rondeau, J.-M., Tête-Favier, F., Podjarny, A., Reymann, J.-M., Barth, P., Biellmann, J.-F. & Moras, D. (1992). *Nature (London)*, **355**, 469–472.
- Rossmann, M. G. (1972). *The Molecular Replacement Method*. New York, London, Paris: Gordon & Breach.
- Rossmann, M. G. (1990). *Acta Cryst.* **A46**, 73–82.
- Schoenborn, B. P. (1988). *J. Mol. Biol.* **201**, 741–749.
- Ševčík, J., Dodson, E. J. & Dodson, G. G. (1991). *Acta Cryst.* **B47**, 240–253.
- Szöke, A., Szöke, H. & Somoza, J. R. (1997). *Acta Cryst.* **A53**, 291–313.
- Urzhumtsev, A. G. (1991). *Acta Cryst.* **A47**, 794–801.
- Urzhumtsev, A. G., Lunin, V. Y. & Luzyanina, T. B. (1989). *Acta Cryst.* **A45**, 34–39.
- Urzhumtsev, A. G. & Podjarny, A. D. (1995a). *Jnt CCP4/ESF-EACBM Newslett. Protein Crystallogr.* **32**, 12–16.
- Urzhumtsev, A. G. & Podjarny, A. D. (1995b). *Acta Cryst.* **D51**, 888–895.
- Urzhumtsev, A. G., Vernoslova, E. A. & Podjarny, A. D. (1996). *Acta Cryst.* **D52**, 1092–1097.
- Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.
- Wilson, J. & Main, P. (2000). *Acta Cryst.* **D56**, 625–633.
- Zhang, K. & Main, P. (1990). *Acta Cryst.* **A46**, 41–46.