Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The use of *VLD* (*vive la difference*) in the molecular-replacement approach: a pipeline

Benedetta Carrozzini,[a] Giovanni Luca Cascarano,[a] Giuliana Comunale,[b] Carmelo Giacovazzo[a]* and Annamaria Mazzone[a]

[a]Istituto di Cristallografia, CNR, Via G. Amendola 122/O, 70126 Bari, Italy, and [b]DiCEM (Dipartimento delle Culture Europee e del Mediterraneo), Università della Basilicata, Matera, Italy

Correspondence e-mail: carmelo.giacovazzo@ic.cnr.it

Received 5 December 2012
Accepted 14 February 2013

VLD (*vive la difference*) is a novel *ab initio* phasing approach that is able to drive random phases to the correct values. It has been applied to small, medium and protein structures provided that the data resolution was atomic. It has never been used for non-*ab initio* cases in which some phase information is available but the data resolution is usually very far from 1 Å. In this paper, the potential of *VLD* is tested for the first time for a classical non-*ab initio* problem: molecular replacement. Good preliminary experimental results encouraged the construction of a pipeline for leading partial molecular-replacement models with errors to refined solutions in a fully automated way. The pipeline moduli and their interaction are described, together with applications to a wide set of test cases.

1. Notation

MR: molecular replacement.
f_j, $j = 1, \ldots, N$: atomic scattering factors for the target structure (thermal factor included).
$F = \sum_{j=1}^{N} f_j \exp(2\pi i \mathbf{hr}_j) = |F|\exp(i\varphi)$: structure factor of the target structure.
$F_p = \sum_{j=1}^{N_p} f_j \exp(2\pi i \mathbf{hr}'_j) = |F_p|\exp(i\varphi_p)$, where $\mathbf{r}'_j = \mathbf{r}_j + \Delta\mathbf{r}_j$: structure factor of the model structure, whether obtained after the orientation and location of the model molecule by an MR program or *via* Fourier inversion of a model electron-density map.
$E = R\exp(i\varphi)$, $E_p = R_p\exp(i\varphi_p)$: normalized structure factors of the target and of the model structure, respectively.
$\Sigma_N = \sum_{j=1}^{N} f_j^2$, $\Sigma_p = \sum_{j=1}^{N_p} f_j^2$.
R'_p: structure factors pseudonormalized with respect to Σ_N, the scattering power of the target structure.
$D = \langle\cos(2\pi\mathbf{h}\Delta\mathbf{r}_j)\rangle$, where the average is over resolution shells.
$\Delta\mathbf{r}_j$ is the positional misfit between the jth atomic position in the model and the corresponding atomic position in the target structure (see the definition of F_p given above).
$\sigma_A = D(\Sigma_p/\Sigma_N)^{1/2}$.
$\sigma_R^2 = (\langle|\mu|^2\rangle/\Sigma_N$; $\langle|\mu|^2\rangle$ is the measurement error.
$e = 1 + \sigma_R^2$.
$I_i(x)$: modified Bessel function of order i.
$m = \langle\cos(\varphi - \varphi_p)\rangle = I_1(X)/I_0(X) = D_1(X)$, where $X = 2\sigma_A RR_p/(e - \sigma_A^2)$.

2. Introduction

The increasing popularity of MR techniques has encouraged and is still encouraging the development of new methods for improving their robustness against insufficient quality of the models, low data resolution and automation. The classical three-dimensional approach defined by Rossmann & Blow (1962) inspired effective three-dimensional space programs

such as *AMoRe* (Navaza, 1994), *MOLREP* (Vagin & Teplyakov, 2010), *ULTIMA* (Rabinovich *et al.*, 1998), *ACORN* (Yao, 2002), *REMO* and *REMO*09 (Caliandro *et al.*, 2006, 2009*a*), and *Phaser* (McCoy *et al.*, 2007). More recently, thanks to advances in automatic computing, six-dimensional space procedures have been developed, for example *EPMR* (Kissinger *et al.*, 1999; Sheriff *et al.*, 1999), *Queen of Spades* (Glykos & Kokkinidis, 2000, 2004), *SOMoRe* (Jamrog *et al.*, 2003) and *via* genetic algorithms (Chang & Lewis, 1997); this is an expensive choice in terms of computing resources, but is sometimes able to lead to a solution in difficult cases.

Despite the spectacular advances of the last twenty years, the model molecules correctly oriented and translated by modern MR programs seldom provide phase values of sufficient quality for the calculation of electron-density maps that are immediately interpretable in terms of target structure: this is particularly frequent in cases of low sequence identity between the model and the target molecules. A very recent approach tries to overcome the present limits by combining many algorithms for crystallographic structure determination with those designed for protein structure modelling (for example, by using *Rosetta* modelling techniques; see DiMaio *et al.*, 2011). Here, we try to extend the limits of a more traditional approach, according to which the electron-density maps available at the end of the MR step are usually submitted to cycles of EDM (electron-density modification; Cowtan, 1999; Abrahams, 1997; Abrahams & Leslie, 1996; Zhang *et al.*, 2001; Refaat & Woolfson, 1993; Giacovazzo & Siliqi, 1997) in order to extend and refine the MR phases.

A noticeable advance in this field has been the EDM–DEDM procedure (in which the second acronym stands for difference electron-density modification; Caliandro *et al.*, 2009*b*): this method has been applied to refine MR phases and to *ab initio* and SAD phasing (Caliandro *et al.*, 2009*c*) and was more effective in improving phases than just EDM. Subsequently (Caliandro *et al.*, 2009*c*), it has been applied to obtain an automated protocol for protein structure refinement based on the iterative application of automated model-building programs [*ARP/wARP* (Perrakis *et al.*, 1999), *PHENIX* (Adams *et al.*, 2010), *MAID* (Levitt, 2001), *MAIN* (Turk, 1992) and *Buccaneer* (Cowtan, 2006)].

In more recent years a new *ab initio* phasing technique has been described, named *VLD* (*vive la difference*; Burla *et al.*, 2011), which is able to bring sets of random phases to solution. It succeeded in solving small- and medium-sized and protein structures at atomic resolution. The application of *VLD* to non-*ab initio* cases and at non-atomic resolution has not been attempted.

In this paper, we are interested in testing the potential of *VLD* in MR procedures, for more effective phase extension and refinement, in combination with EDM techniques. The application to a large set of test structures is described in §4. The favourable results encouraged us to implement a pipeline that, given a model structure, could automatically produce, without any user intervention, an electron-density map that is interpretable by automatic building programs. This pipeline differs from other MR pipelines developed in recent years, for example *BALBES* (Long *et al.*, 2008), *MrBUMP* (Keegan & Winn, 2008) and *NORMA* (Delarue, 2008). It skips the first step (*i.e.* exploiting the Protein Data Bank to find the optimum models for a given target protein) and only deals with the second step: protein crystal structure solution given the model. A number of new phasing tools, including *VLD*, are applied to make the solution step more straightforward and more robust. Future integration with complementary pipelines may make the full two-step pathway highly effective.

3. The pipeline moduli

The pipeline is constituted of seven moduli: *REMO*09 (Caliandro *et al.*, 2009*a*), *REFMAC* (Murshudov *et al.*, 2011), *DM* (Cowtan, 1994), *DSR* (Giacovazzo & Siliqi, 1997), *VLD* (Burla *et al.*, 2011), *free lunch* routines (Caliandro *et al.*, 2005, 2007) and *ARP/wARP* (Perrakis *et al.*, 1999). We briefly describe the role of the various moduli and their interconnection in the pipeline.

3.1. *REMO*09

*REMO*09 is an MR program based on the method of joint probability distribution functions. The MR problem is subdivided into rotation and translation steps. The implemented theory allows the program to work under different prior conditions: for example, for the rotation of a monomer it uses the conditional probability density given the rotation and the translation values of one or more other monomers. The same probabilistic approach allows the most probable translation shift for a given monomer to be found in the translation step given the orientation and/or the locations of other monomers. The program receives the necessary information on the model and on the target structure from the user, automatically takes decisions for a straightforward phasing attempt and produces the atomic coordinates of the model suitably oriented and translated as output.

3.2. *REFMAC*

REFMAC is available from *CCP*4 (Winn *et al.*, 2011). The program automatically reads the output of *REMO*09 and submits the positions and temperature factors of the model atoms to five cycles of maximum-likelihood refinement. The final phases are submitted to the *VLD* modulus.

3.3. *DM* and *DSR*

Both of the EDM programs apply real-space constraints based on known protein features to the current electron-density maps in order to extend and refine the phases. *DM* is applied when the data resolution is non-atomic (*i.e.* worse than 1.25 Å): solvent flattening and histogram mapping are the only tools used. *DSR* is preferred when the data resolution is atomic: in this case the structure is overdetermined by the data and only solvent flattening is necessary to improve the phases. *DM* or *DSR* are used at the end of *REMO*09 or just after having submitted the *REMO*09 model to *REFMAC* refinement (see §4). A maximum of 15 EDM cycles are

executed: cycling is interrupted if the average phase difference between the values in cycle j and those in cycle $j-1$ is smaller than 2° or if the average m value in cycle j is smaller than that in cycle $j-1$.

3.4. *VLD*

The phases produced by *DM* or *DSR* are submitted to the *VLD* modulus. It is based on a new type of difference Fourier synthesis (Burla *et al.*, 2010), which is particularly efficient for revealing missed or badly located atoms when the model structure is poor, with coefficients

$$\Delta E = \left[(mR - \sigma_A R_p) - R'_p(1-D)\left(\frac{e-\sigma_A^2}{1-\sigma_A^2}\right)\right]\exp(i\varphi_p). \quad (1)$$

Coefficients (1) are a combination of the classical difference term $(mR - \sigma_A R_p)\exp(i\varphi_p)$ (Read, 1986) and of the flipping term

$$-R'_p(1-D)\left(\frac{e-\sigma_A^2}{1-\sigma_A^2}\right)\exp(i\varphi_p).$$

In order to make *VLD* applicable to a large variety of cases it was assumed that $\Sigma_p = \Sigma_N$, thus giving rise to the simpler coefficients

$$\Delta E \simeq (mR - R_p)\exp(i\varphi_p). \quad (2)$$

The ΔE difference Fourier synthesis is conveniently modified and inverted, and the corresponding Fourier coefficients $E_q \equiv (R_q, \varphi_q)$ are combined with the normalized structure factors of the model structure through the tangent formula

$$\tan\varphi = \frac{R_p \sin\varphi_p + w_q R_q \sin\varphi_q}{R_p \cos\varphi_p + w_q R_q \cos\varphi_q}, \quad (3)$$

where $w_q = [2(1-\sigma_A)]^{1/2}$.

An observed Fourier synthesis using the φ phases given by (3) is then calculated and submitted to three cycles of EDM.

The difference Fourier step (2) is designed to reveal missing atomic positions and to discard wrongly positioned densities and the EDM cycles are designed to refine the new models after the application of (3).

3.5. The *free lunch* routines

The final *VLD* phases and the observed moduli are automatically used to extrapolate the moduli and phases of non-measured reflections both beyond and behind the experimental resolution. The method requires the modification of the observed electron density followed by Fourier inversion and submission of the extrapolated factors to histogram matching. The extrapolation limit is fixed to 1.2 Å, no matter what the value of the experimental resolution is; a number of extrapolated reflections are selected (for active use in the subsequent electron density) equal to 75% of the observed reflections. Their weight is fixed by extrapolating the α_A curve to 1.2 Å resolution. The *free lunch* routine is applied two times.

3.6. *ARP*/*wARP*

The current set of phases is read, dummy atoms are created in high-density regions, new atoms are added and old ones are deleted to create new models which are cyclically refined in reciprocal space by *REFMAC*. The last step requires a number of observations that is larger than the number of model parameters and this sets the resolution limit of *ARP*/*wARP* to about 2.5–2.8 Å. In our pipeline *ARP*/*wARP* is applied only at the end of the *free lunch* routine: since it is able to profit by the sequence coverage obtained in a preceding cycle, *ARP*/*wARP* is cycled up to a maximum of three times (*ARP*/*wARP* cycling is not allowed if the sequence coverage is larger than 0.90 or smaller than 0.02).

4. Applications

Any proposed new phasing procedure should be checked using a wide set of applications. To perform this, we selected four test structures with high-resolution data (conventionally better than 1.25 Å) and 41 structures with lower resolution data (1.50–2.86 Å).

The full list of structures is given in Table 1 arranged in increasing order of data resolution: in the table we give the PDB codes of the target (TARG) and of the model structure (MOD), the data resolution in Å (RES), the sequence identity between the model and the target structure (ID), the root-mean-square deviation between pairwise C^α backbone positions (RMS), the number of monomers ($n_{\rm Mon}$) and of residues in the asymmetric unit of the target ($N_{\rm resT}$) and the number of residues in the model ($N_{\rm resM}$).

Some test structures were not originally solved by MR (*e.g.* 2sar was solved by isomorphous replacement), while some others were used as test cases by three-dimensional or six-dimensional MR search programs. 1cgn, 1cgo, 1aki and 6rhn were tested by *SOMoRe*, 1cgn by *EPMR*, 1lys and 1aki by *Queen of Spades*, 1e8a by *MOLREP*, 1bxo by *ACORN* and 1lat by *ULTIMA*. It is therefore interesting to check the degree of success of our pipeline in the above cases, which were evidently considered to be interesting tests or difficult cases by the authors of the above programs.

The results of our applications are shown in Table 2. Two lines in the table are associated with each test structure. In the first line (protocol 1) we give the mean phase error after the application of *REMO*09 (REMO), *VLD* (VLD) and the *free lunch* routine (FL). The sequence coverage automatically obtained after the first application of *ARP*/*wARP* and after its second and third applications follow (COV1, COV2 and COV3, respectively).

The only difference between the first and the second line is the following: in the REMO column the molecular model suitably oriented and positioned by *REMO*09 is submitted to five *REFMAC* least-squares cycles (protocol 2). The average phase error reported in the second line under the heading REMO is that obtained after *REFMAC* refinement.

The analysis of our experimental results (obtained with the default conditions specified for protocols 1 and 2) will be made in two steps: some general conclusions will first be established

Table 1
List of test structures.

For each test structure the following parameters are given: the PDB codes for the target (TARG) and for the model structure (MOD), the data resolution (RES), the sequence identity (ID) between the model and target structures, the root-mean-square deviation between pairwise C^{α} backbone positions (RMS), the number of monomers ($n_{\rm Mon}$) and of residues in the asymmetric unit of the target ($N_{\rm resT}$) and the number of residues in the model ($N_{\rm resM}$). The model for 1tgx was kindly provided by E. Dodson.

TARG	MOD	RES (Å)	RMS	ID	$n_{\rm Mon}$	$N_{\rm resT}/N_{\rm resM}$
1dy5	1lsq	0.87	0.27	1	2	248/124
1bxo	1er8	0.9	1.15	0.55	1	323/330
1a6m	1mbc	1	0.22	1	1	151/153
1kf3	7rsa	1.05	0.11	0.97	1	124/124
1aki	2ihl	1.5	0.39	0.97	1	129/129
1tp3	1be9	1.54	0.30	1	1	115/115
2fc3	1xbi	1.54	0.82	0.58	1	124/118
1tgx	—	1.55	0.65	1	3	180/50
1zs0	1i76	1.56	0.36	1	1	165/163
2a46	1g7k	1.65	0.95	0.4	1	217/217
6ebx	3ebx	1.7	0.82	1	2	124/62
1lys	2ihl	1.72	0.60	0.96	2	258/129
1cgo	2ccy	1.79	1.73	0.3	1	127/127
2otb	1zux	1.79	0.50	0.68	2	432/226
1kqw	1opa	1.8	0.54	0.74	1	134/133
2sar	1ulc	1.85	0.32	0.98	2	192/96
2hyu	1xjl	1.86	0.50	0.99	1	308/319
1lat	1glu	1.9	0.94	0.89	2	145/81
1e8a	1mho	1.95	1.52	0.36	2	175/88
2f53	2bnr	1.99	1.03	0.98	1	811/820
2ayv	1x23	2	0.79	0.8	1	148/153
2h8q	1g7k	2	0.33	0.97	2	868/436
2omt	1o6s	2	0.37	1	1	565/564
2pby	1mki	2.07	1.48	0.46	2	1155/624
2f8m	1uj5	2.09	1.20	0.4	2	472/225
1yxa	1qlp	2.1	1.68	0.46	2	740/372
2f84	12aqw	2.1	0.99	0.64	1	323/321
2hyw	1xjl	2.1	0.40	1	2	616/319
1cgn	2ccy	2.15	1.73	0.31	1	125/127
6rhn	4rhn	2.15	0.30	1	1	115/104
1xyg	1vkn	2.19	1.26	0.45	1	1380/1360
2ah8	1ema	2.21	0.39	0.96	2	466/225
2a4k	1uls	2.3	1.08	0.64	2	439/245
2gq3	1n8i	2.3	0.50	1	2	1434/701
2i3p	1g9y	2.3	0.33	0.99	1	304/304
2o3k	1ysb	2.3	0.35	0.99	1	307/317
2p0g	2oka	2.3	0.60	0.64	1	318/336
2a03	1isc	2.33	0.89	0.53	1	394/384
1na7	1m2r	2.4	0.84	0.76	1	326/327
2b5o	1b2r	2.5	1.16	0.63	2	584/295
2oka	20bk	2.5	0.45	0.887	1	336/335
1ycn	1n00	2.51	1.16	0.72	2	619/318
2iff	1hem	2.58	0.50	0.98	1	555/129
1s31	1c8z	2.7	1.26	0.96	1	273/265
2qu5	2p2i	2.86	0.81	1	1	292/289

and a few examples will then be discussed in some detail. Under the reasonable assumption that the target structure is satisfactorily solved and interpreted if the sequence coverage by *ARP*/*wARP* is larger than 70%, the general conclusions are as follows.

(i) The four high-resolution test structures are easily solved and fully interpreted even when *REMO*09 ends with an average phase error of 74° (this is the case for 1dy5 and 1bxo). The reason is that such structures are overdetermined by the data, and it is therefore more easy to recover good phases from relatively bad starting sets. For the above structures the *free lunch* routine cannot be applied.

(ii) If protocol 1 is used, attempts to interpret the final electron density using *ARP*/*wARP* failed for six structures: 1lat, 2pby, 1yxa, 1xyg, 2iff and 2qu5. For all of them *REMO*09 correctly oriented and located the model molecules, but the subsequent steps were not able to lead *ARP*/*wARP* to a correct interpretation. The case of 2qu5, which was well solved but not interpreted, will be discussed in detail below.

(iii) If protocol 2 is used, only three structures remained unsolved: 1lat, 2pby and 2iff (however, 2pby may be solved under nondefault conditions; see below). The improvement is mainly owing to the use of *REFMAC*, which was able to lower the *REMO*09 mean phase error, with beneficial effects on the phase errors of the subsequent phasing steps.

(iv) *VLD* works particularly well when applied to phase sets with relatively high average phase error for both protocols (see the cases 2a46, 1cgo, 1e8a, 2f8m, 1cgn, 2a4k *etc.*). When the average phase error before the application of *VLD* is low, the phase improvement may be marginal or vanishing, as may be expected when EDM techniques are applied to models that are well refined using least-squares or maximum-likelihood procedures. This situation occurs when the model molecule used in the MR step has a high value of ID and/or a small value of RMS or when the *REFMAC* refinement is particularly effective: obviously, in these cases the amount of phase improvement is of minor interest for the phasing procedure. It may also be of interest to note that the *VLD* modulus involves the calculation and inversion of only a few electron-density maps: therefore, a minimum CPU time is necessary to obtain the noticeable phase improvement shown in Table 2. A brief comparison between *VLD* and its precursor, the EDM–DEDM approach, is useful: the *VLD* phase improvement is larger (by about 2° on average) and was obtained in about 1/7 of the CPU time necessary for EDM–DEDM. Our *a posteriori* analysis of the main *VLD* features indicates that the *VLD* difference electron density is responsible for an increase of the mean phase error in the first step of the algorithm, followed by its rapid decrease in the EDM section of the algorithm. Without the first variation, the second does not occur.

(v) The efficiency of the *free lunch* routine noted in other applications (*e.g.* Usón *et al.*, 2007) is weakened by the high efficiency of *VLD*: as a rule, the phases of the observed reflections improve by few degrees. It does not provide additional reduction of the phase error when the input (to the *free lunch* modulus) average phase error is very small.

(vi) In most of the cases (32 for both of the protocols) a sequence coverage of 90% or higher is obtained after the first application of *ARP*/*wARP*; in few cases a third application is necessary for a satisfactory interpretation. This favourable behaviour leads to a lower CPU time being necessary for a complete run.

(vii) The resolution is certainly one of the limiting factors of the pipeline. The efficiency of *VLD* decreases when the resolution increases, but it is still able to satisfactorily expand and refine phases up to about 2.8 Å, which is close to the *ARP*/*wARP* algorithm limit.

Table 2
For each test structure we show the average phase error (°) at the end of each pipeline modulus.

COV1, COV2 and COV3 are the protein coverages obtained after the first, the second and the third consecutive application of *ARP*/*wARP*, respectively.

CODE	REMO	VLD	FL	COV1 (%)	COV2 (%)	COV3 (%)
1dy5	74	19		99		
	31	19		99		
1bxo	74	21		99		
	51	20		99		
1a6m	43	20		99		
	27	20		99		
1kf3	21	21		99		
	12	22		99		
1aki	38	35	32	99		
	24	26	27	99		
1tp3	50	47	44	99		
	41	42	42	99		
2fc3	54	41	37	99		
	38	36	33	99		
1tgx	58	49	45	92		
	58	49	45	92		
1zs0	43	39	32	99		
	32	33	29	99		
2a46	69	40	33	99		
	57	40	32	99		
6ebx	44	38	36	99		
	30	30	31	99		
1lys	53	47	46	99		
	37	36	37	99		
1cgo	74	66	64	99		
	67	57	54	99		
2otb	58	48	44	99		
	41	39	37	99		
1kqw	59	47	43	99		
	41	38	35	99		
2sar	52	42	39	99		
	42	39	37	99		
2hyu	50	40	37	99		
	34	33	32	99		
1lat	70	68	68	0		
	62	64	64	0		
1e8a	69	56	53	99		
	54	45	42	99		
2f53	59	49	48	98		
	41	38	39	99		
2ayv	53	45	44	89	90	
	39	37	37	87	85	91
2h8q	48	44	45	76	84	94
	37	37	39	94		
2omt	38	41	41	99		
	30	33	34	99		
2pby	79	72	72	0		
	75	68	67	12	8	16
2f8m	64	53	52	99		
	54	47	45	99		
1yxa	74	68	67	0		
	69	61	60	8	52	90
2f84	55	46	47	97		
	43	40	40	92		
2hyw	49	44	43	99		
	33	35	36	99		
1cgn	73	61	59	99		
	72	60	58	99		
6rhn	32	32	32	99		
	26	27	28	99		
1xyg	63	55	54	0		
	53	47	47	49	83	95
2ah8	40	33	33	99		
	32	30	30	99		
2a4k	60	47	45	99		
	48	39	38	99		
2gq3	45	41	42	85	89	89
	34	36	37	85	88	88
2i3p	37	36	37	78	91	
	32	33	34	90		
2o3k	36	35	36	99		
	28	29	30	99		
2p0g	51	42	42	99		
	40	36	36	99		
2a03	49	37	37	96		
	37	34	34	96		
1na7	46	42	42	92		
	37	36	37	88	97	
2b5o	50	44	43	72	78	88
	42	39	39	87	79	91
2oka	36	32	33	99		
	26	26	28	96		
1ycn	56	45	45	61	84	85
	41	38	38	73	82	90
2iff	62	65	65	0		
	63	70	72	0		
1s31	49	35	35	86	94	
	37	32	33	88	99	
2qu5	44	34	35	20	0	
	27	26	28	0		

A more detailed discussion of some test cases may be useful to complete the analysis of the experimental results.

4.1. 2qu5

This is one of the structures for which the final electron density is not interpreted by *ARP*/*wARP*. This failure may be ascribed to the poor data resolution (2.86 Å, which is close to the limit of *ARP*/*wARP*) rather than to inefficiency of the pipeline. Indeed, the *free lunch* routine ends with average phase errors of 35 and 28° for protocols 1 and 2, respectively. It may also be noted that *ARP*/*wARP* succeeds in the case of 1s31, for which the data resolution is 2.70 Å and the *free lunch* routine ends with average phase errors of 35° and 33° for protocols 1 and 2, respectively.

4.2. 2pby

This structure cannot be solved and interpreted using either of the two protocols. It has been solved using a nondefault approach by extending the number of *REFMAC* cycles to ten. In this case the corresponding values in Table 2 become REMO = 72, VLD = 64, FL = 64, COV1 = 27%, COV2 = 98%. Application of the above procedure to other unsolved structures was not successful.

4.3. 2iff

The scattering power of the model is a small percentage (about 0.23) of that of the target. However, *REMO*09 was able to find the correct rotation and translation, ending with an average phase error of 62°. The subsequent application of the *VLD* and *free lunch* moduli was not able to bring the phases closer to the correct values. The reason is probably the poor quality of the model combined with the poor data resolution. This was reported as 2.58 Å in Table 1, but the effective

resolution is equal to 2.95 Å because of the high percentage of nonmeasured reflections in the highest resolution shell.

4.4. 1lat

1lat was used as an MR test structure by *ULTIMA* and by *REMO* but ended with an average phase error of 70°, despite favourable values of RMS and ID (0.94 Å and 0.89, respectively). This was mainly owing to the fact that the scattering power of the model is about 56% of that of the target. Protocol 1 was not effective (the average phase error only decreased to 68°); protocol 2 was more efficient, ending with an average phase error of 64°, but *ARP*/*wARP* was unable to interpret the final electron density. This is probably owing to the fact that about 772 atoms of DNA coexist with the 1137 protein atoms in the asymmetric unit.

5. Conclusions

The potential of *VLD* has been tested in a novel MR pipeline. Application to a wide set of test structures suggest that (i) it is quite effective even far from atomic resolution, a condition under which it had not been tested, and (ii) in combination with EDM techniques it is able to efficiently extend and reduce the phase error.

The pipeline has been tested according to two protocols: the second involves the application of *REFMAC* refinement to the model suitably oriented and positioned by *REMO*09. This second protocol is assumed to be the default of the pipeline, which thus is able to automatically bring most of the test structures to solution. The quality of the resulting phases allows quite satisfactory sequence coverage by *ARP*/*wARP*.

To draw more accurate conclusions on the usefulness of the pipeline we made two supplementary tests.

(i) To check how good models directly provided by *REMO*09 are, we excluded the *VLD* and *free lunch* steps from the pipeline: now *ARP*/*wARP* is directly applied to the output phases of *REMO*09. The results may be condensed as follows. Six more structures remained unsolved (1cgo, 1yxa, 1cgn, 1xyg, 2b5o and 1ycn) in addition to the three cases in which the default choice of the pipeline did not succeed (1lat, 2pby and 2iff; however, 2pby is solved by the pipeline under nondefault conditions). The failures of the truncated pipeline correspond to cases in which the *REMO*09 average phase errors are too large to allow good coverage by *ARP*/*wARP*.

(ii) To check the relative efficiency of the pipeline segment devoted to phase extension and refinement, we considered the ten solved test cases at non-atomic resolution in which *REMO*09 ended with an average phase error larger than 60° (*i.e.* the most interesting cases). We first extended and refined *REMO*09 phases using only *DM*: in terms of phase error the average gain was about 10°. For the same set of test structures we then stopped the pipeline at the *free lunch* routine (*i.e.* we did not use *ARP*/*wARP*): the average phase gain was about 19°.

The above tests show that the phase-extension and refinement step of the pipeline (*REFMAC* + *DM* + *VLD* + the *free lunch* routine) may increase the efficiency of the usual phase-extension and refinement approaches based only on EDM techniques and makes the success of the automated model-building programs more easier; its use is not expensive in terms of CPU time and therefore may be combined not only with *REMO*09 and/or with *ARP*/*wARP* but also with more complex and efficient algorithms such as those implemented in *PHENIX* by DiMaio *et al.* (2011).

6. Availability

The algorithms and the pipeline described in this paper have been implemented in v.2.0 of *Sir*2011, which is presently under development. Interested parties can obtain a beta version by contacting the e-mail address sirmail@ic.cnr.it.

References

Abrahams, J. P. (1997). *Acta Cryst.* D**53**, 371–376.
Abrahams, J. P. & Leslie, A. G. W. (1996). *Acta Cryst.* D**52**, 30–42.
Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.Burla, M. C., Caliandro, R., Giacovazzo, C. & Polidori, G. (2010). *Acta Cryst.* A**66**, 347–361.
Burla, M. C., Carrozzini, B., Cascarano, G. L., Giacovazzo, C. & Polidori, G. (2011). *J. Appl. Cryst.* **44**, 1143–1151.
Caliandro, R., Carrozzini, B., Cascarano, G. L., De Caro, L., Giacovazzo, C., Mazzone, A. M. & Siliqi, D. (2006). *J. Appl. Cryst.* **39**, 185–193.
Caliandro, R., Carrozzini, B., Cascarano, G. L., De Caro, L., Giacovazzo, C., Moustiakimov, M. & Siliqi, D. (2005). *Acta Cryst.* A**61**, 343–349.
Caliandro, R., Carrozzini, B., Cascarano, G. L., De Caro, L., Giacovazzo, C. & Siliqi, D. (2007). *J. Appl. Cryst.* **40**, 931–937.
Caliandro, R., Carrozzini, B., Cascarano, G. L., Giacovazzo, C., Mazzone, A. M. & Siliqi, D. (2009*a*). *Acta Cryst.* A**65**, 512–527.
Caliandro, R., Carrozzini, B., Cascarano, G. L., Giacovazzo, C., Mazzone, A. M. & Siliqi, D. (2009*b*). *Acta Cryst.* D**65**, 249–256.
Caliandro, R., Carrozzini, B., Cascarano, G. L., Giacovazzo, C., Mazzone, A. M. & Siliqi, D. (2009*c*). *Acta Cryst.* D**65**, 477–484.
Chang, G. & Lewis, M. (1997). *Acta Cryst.* D**53**, 279–289.
Cowtan, K. (1994). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **31**, 34–38.
Cowtan, K. (1999). *Acta Cryst.* D**55**, 1555–1567.
Cowtan, K. (2006). *Acta Cryst.* D**62**, 1002–1011.
Delarue, M. (2008). *Acta Cryst.* D**64**, 40–48.
DiMaio, F., Terwilliger, T. C., Read, R. J., Wlodawer, A., Oberdorfer, G., Wagner, U., Valkov, E., Alon, A., Fass, D., Axelrod, H. L., Das, D., Vorobiev, S. M., Iwaï, H., Pokkuluri, P. R. & Baker, D. (2011). *Nature (London)*, **473**, 540–543.
Giacovazzo, C. & Siliqi, D. (1997). *Acta Cryst.* A**53**, 789–798.
Glykos, N. M. & Kokkinidis, M. (2000). *Acta Cryst.* D**56**, 169–174.
Glykos, N. M. & Kokkinidis, M. (2004). *J. Appl. Cryst.* **37**, 159–161.
Jamrog, D. C., Zhang, Y. & Phillips, G. N. (2003). *Acta Cryst.* D**59**, 304–314.
Keegan, R. M. & Winn, M. D. (2008). *Acta Cryst.* D**64**, 119–124.
Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B. (1999). *Acta Cryst.* D**55**, 484–491.
Levitt, D. G. (2001). *Acta Cryst.* D**57**, 1013–1019.
Long, F., Vagin, A. A., Young, P. & Murshudov, G. N. (2008). *Acta Cryst.* D**64**, 125–132.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.

Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Rabinovich, D., Rozenberg, H. & Shakked, Z. (1998). *Acta Cryst.* D**54**, 1336–1342.
Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.
Refaat, L. S. & Woolfson, M. M. (1993). *Acta Cryst.* D**49**, 367–371.
Rossmann, M. G. & Blow, D. M. (1962). *Acta Cryst.* **15**, 24–31.
Sheriff, S., Klei, H. E. & Davis, M. E. (1999). *J. Appl. Cryst.* **32**, 98–101.
Turk, D. (1992). PhD thesis. Technische Universitat München, Germany.
Usón, I., Stevenson, C. E. M., Lawson, D. M. & Sheldrick, G. M. (2007). *Acta Cryst.* D**63**, 1069–1074.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Yao, J.-X. (2002). *Acta Cryst.* D**58**, 1941–1947.
Zhang, K. Y. J., Cowtan, K. D. & Main, P. (2001). *International Tables for Crystallography*, Vol. *F*, edited by M. G. Rossmann & E. Arnold, pp. 311–331. Dordrecht: Kluwer Academic Publishers.