BIOLOGICAL CRYSTALLOGRAPHY

ISSN 1399-0047

Received 24 April 2015
Accepted 4 June 2015

Edited by Z. Dauter, Argonne National Laboratory, USA

Keywords: phasing; *ARCIMBOLDO_BORGES*; *ab initio*; small fragments.

Supporting information: this article has supporting information at journals.iucr.org/d

ARCIMBOLDO_LITE: single-workstation implementation and use

Massimo Sammito,[a] Claudia Millán,[a] Dawid Frieske,[a] Eloy Rodríguez-Freire,[a] Rafael J. Borges[a] and Isabel Usón[b]*

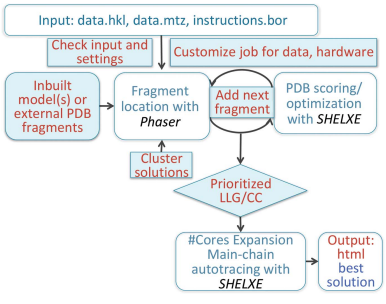
[a]Crystallographic Methods, Institute of Molecular Biology of Barcelona (IBMB–CSIC), Baldiri Reixac 15, 08028 Barcelona, Spain, and [b]Structural Biology, ICREA at IBMB–CSIC, Baldiri Reixach 13-15, 08028 Barcelona, Spain. *Correspondence e-mail: uson@ibmb.csic.es

ARCIMBOLDO solves the phase problem at resolutions of around 2 Å or better through massive combination of small fragments and density modification. For complex structures, this imposes a need for a powerful grid where calculations can be distributed, but for structures with up to 200 amino acids in the asymmetric unit a single workstation may suffice. The use and performance of the single-workstation implementation, *ARCIMBOLDO_LITE*, on a pool of test structures with 40–120 amino acids and resolutions between 0.54 and 2.2 Å is described. Inbuilt polyalanine helices and iron cofactors are used as search fragments. *ARCIMBOLDO_BORGES* can also run on a single workstation to solve structures in this test set using precomputed libraries of local folds. The results of this study have been incorporated into an automated, resolution- and hardware-dependent parameterization. *ARCIMBOLDO* has been thoroughly rewritten and three binaries are now available: *ARCIMBOLDO_LITE*, *ARCIMBOLDO_SHREDDER* and *ARCIMBOLDO_BORGES*. The programs and libraries can be downloaded from http://chango.ibmb.csic.es/ARCIMBOLDO_LITE.

1. Introduction

The phase problem is the central problem of crystallography and is a bottleneck in the determination of macromolecular crystal structures (Hendrickson, 2013). Several methods exist to provide starting phases for structure factors to approximate the correct phases, which are missed in the diffraction experiment. In the case of macromolecules, experimental phasing through heavy-atom derivatives or anomalous scattering at particular wavelengths (Hendrickson, 1991) is used if no previous structural knowledge from a related structure is available for phasing by molecular replacement (Rossman, 1972; Navaza, 1994).

Ab initio phasing from the native diffraction intensities alone is accomplished by direct methods in the case of molecules with less than 200 atoms that diffract to atomic resolution (Karle & Hauptman, 1956). Dual-space recycling has succeeded in extending the use of direct methods to small proteins of up to 1000 atoms, maintaining the atomicity restriction (Miller *et al.*, 1993; Sheldrick *et al.*, 2012). Approaches to relax this restriction have involved sophisticated use of the Patterson function (Caliandro *et al.*, 2008), data extrapolation to extend the resolution limit beyond the experimentally measured data (Caliandro *et al.*, 2005; Usón *et al.*, 2007) and density-modification algorithms tailored to the high-resolution case, such as low-density elimination (Shiono & Woolfson, 1992; Refaat & Woolfson, 1993), the sphere of influence (Sheldrick, 2002) and *VLD* (Burla *et al.*, 2010, 2011,



2012). Starting phasing from a small but highly accurate substructure in the context of *ACORN* has been reported to be remarkably effective (Yao *et al.*, 2005, 2006) and our own tests have corroborated this finding: as little as 10% of the main-chain atoms suffice to solve a structure at 2 Å resolution through density modification. This provides a way to escape the need for atomic resolution that restricts direct methods and real-space atom selection. The atomicity constraints can be substituted by enforcing secondary- or tertiary-structure stereochemistry. A related proof of principle has been established using small fragments such as α-helices (Glykos & Kokkinidis, 2003) and nucleotides (Robertson & Scott, 2008; Robertson *et al.*, 2010) as search fragments to seed phasing. This is accomplished in *ARCIMBOLDO* (Rodríguez *et al.*, 2009, 2012) by combining a search for small polyalanine model fragments with *Phaser* (McCoy *et al.*, 2007) with density modification and autotracing with *SHELXE* (Thorn & Sheldrick, 2013). Extremely successful approaches based on more complete models of lower accuracy (Rigden *et al.*, 2008) have been developed and include the enhancement of poor search models and *de novo* prediction through modelling with *Rosetta* (Qian *et al.*, 2007) or *QUARK* (Xu & Zhang, 2012) linked to molecular replacement with *Phaser*. This design underlies methods such as *MR-Rosetta* (DiMaio *et al.*, 2011) and *AMPLE* (Bibby *et al.*, 2012, 2013; Keegan *et al.*, 2015), and other implementations (Shrestha *et al.*, 2011; Shrestha & Zhang, 2015).

The original *ARCIMBOLDO* procedure relies on the assumption that the main chain of small secondary-structural elements, such as α-helices, is conserved among unrelated protein structures. Because of the generality of such partial models, which are predictable from the sequence (Cole *et al.*, 2008) or data (Caliandro *et al.*, 2012; Morris *et al.*, 2004), this method is in fact an *ab initio* approach. Recently, it has been shown how unrelated structures can share sufficiently similar local folds comprising the main chain of a few secondary-structure elements. Libraries of such folds can be extracted with *BORGES* (Sammito *et al.*, 2013) from the entire PDB (Bernstein *et al.*, 1977; Berman *et al.*, 2000) and then input into *ARCIMBOLDO_BORGES* as search models. If a the structure of a distant homologue is known, fragments from this template can also be extracted and evaluated against the rotation function, as implemented in *ARCIMBOLDO_SHREDDER* (Sammito *et al.*, 2014).

As small models account for a low fraction of the total scattering, the figures of merit characterizing a correct solution may fail to identify it as such, making it indistinguishable from a large pool of incorrect solutions. This has required the implementation of *ARCIMBOLDO* as a multisolution method, in which all possible solutions have to be explored independently up to the step of density modification and autotracing, where a correct density map is unequivocally recognizable at 2 Å resolution. Supercomputing facilities are required to manage, distribute and check this magnitude of computation, and databases to aid in storing and retrieving the amount of data generated. Improvement in the identification of correct solutions allows the pool of solutions to be narrowed in some cases (Oeffner *et al.*, 2013). The accumulated experience in using this method and advances in the software involved (*Phaser* and *SHELXE*) have made it possible to deploy a standalone version which will just require multiprocessing facilities and particular filters to reduce computation depending on the characteristics of the available hardware.

In the present work, we have developed and assessed single-machine implementations of the *ARCIMBOLDO* methods described by Millán *et al.* (2015). *ARCIMBOLDO_LITE* performs *ab initio* phasing using polyalanine helices or other single search fragments. *ARCIMBOLDO_BORGES* performs *ab initio* phasing with nonspecific libraries of small folds. For the sake of completeness, although not discussed in this paper, *ARCIMBOLDO_SHREDDER* uses fragments derived from a distant homologue template. The new software has been thoroughly rewritten and dimensioned to run on a single, multiprocessor machine, thus eliminating the limiting condition of access to a supercomputer or grid cluster. No database is required either. The binary standalone version is distributed for Linux and Mac OS at http://chango.ibmb.csic.es/ARCIMBOLDO_LITE. The performance of *ARCIMBOLDO_LITE* and *ARCIMBOLDO_BORGES* on a pool of test structures and systems is described and the conclusions of these tests have been used to set up data-dependent default parameterization.

Table 1
Performance of *ARCIMBOLDO_LITE* on a pool of 12 workstations for two test structures (PDB entries 2iu1 and 4k82).

Hardware	OS	RT, 2iu1	RT, 4k82
Mean for pool of 12 workstations		7 h 37 min	2 h 1 min
Xeon X5560 at 2.80 GHz, 8 cores sharing 24 GB	Debian 6.0	4 h 26 min	1 h 18 min
i7 2600 at 3.40 GHz, 4 cores sharing 12 GB	Debian 6.0	5 h 34 min	1 h 36 min
i7 3930K at 3.20 GHz, 6 cores sharing 16 GB	Debian 6.0	4 h 10 min	1 h 21 min
Xeon E5410 at 2.33 GHz, 8 cores sharing 32 GB	Debian 6.0	7 h 46 min	1 h 34 min
2 Duo E4500 at 2.20 GHz, 2 cores sharing 512 MB	Debian 6.0	32 h 19 min	8 h 45 min
i7 950 at 3.07 GHz, 4 cores sharing 6 GB	Ubuntu 10.04.2	6 h 41 min	2 h 3 min
Xeon E5440 at 2.83 GHz, 8 cores sharing 16 GB	Ubuntu 10.04.3	6 h 58 min	1 h 56 min
i7 X980 at 3.33 GHz, 6 cores sharing 12 GB	Ubuntu 10.04.4	5 h 1 min	1 h 21 min
Xeon E5-2650 v2 at 2.66–4.00 GHz, 16 cores sharing 125 GB	Centos 7.1.1503	1 h 55 min	32 min
Xeon E5-2680 v3 at 2.50 GHz, 24 cores sharing 16 GB	openSUSE 13.2	4 h 34 min	1 h
I7-3720QM at 2.60 GHz, 8 cores sharing 8 GB	Mac OSX Yosemite 10.10	3 h 16 min	1 h 10 min
I7-3720QM at 2.80 GHz, 8 cores sharing 4 GB	Mac OSX Mavericks 10.9	8 h 54 min	1 h 40 min

2. Materials and methods

2.1. Computing setup

Structure solution was run on the eight identical eight-core machines of an HP ProLiant BL460c blade system, using them as single workstations with dual quad-core Xeon E5440

processors (2.83 GHz and 16 GB RAM). The Linux distribution installed was Ubuntu Server 10.04 LTS.

2.2. Software versions

The *ARCIMBOLDO* binary is deployed for Linux and Macintosh (some examples of compatible distributions and versions can be found in Table 1). It was generated with *PyInstaller* 2.1 and Python 2.6, with version 2014 upwards of *SHELXE* and version 2.5.6 upwards of *Phaser* from the *PHENIX* (Adams *et al.*, 2010) or *CCP*4 (Winn *et al.*, 2011) distributions. Model and maps were examined with *Coot* (Emsley *et al.*, 2010). Figures were prepared with *PyMOL* (v.1.5.0.4; Schrödinger) and *gnuplot* 4.4. *F*-weighted mean phase errors were calculated with *SHELXE* (Sheldrick, 2002) against the final models deposited in the PDB (Bernstein *et al.*, 1977; Berman *et al.*, 2000).

2.3. Test set used

In this study, the structure of EIF5 (Bieniossek *et al.*, 2006; PDB entry 2iu1) at 1.7 Å resolution and the structure of Lv-ranaspumin (Hissa *et al.*, 2014; PDB entry 4k82) at 1.7 Å resolution previously solved with *ARCIMBOLDO* were used to benchmark performance on different hardware setups. Furthermore, we used a pool of 294 structures with resolutions between 2.2 and 0.54 Å, sizes of between 44 and 120 residues in the asymmetric unit and 47 different space groups, approximately following the frequency distribution seen in the PDB. In particular, the most frequent 36 space groups are represented, with $P2_12_12_1$ predominating. The PDB codes for these 294 structures are as follows: 1ejg, 1en2, 1ew4, 1ezj, 1f94, 1fk5, 1g2r, 1g6u, 1gk6, 1gmx, 1gvd, 1gxu, 1i2t, 1i71, 1iqz, 1j2l, 1j8b, 1j8e, 1kth, 1kwi, 1l9l, 1lsl, 1lxj, 1m1q, 1mg4, 1mk0, 1ne8, 1njh, 1nnx, 1oap, 1oks, 1ox3, 1p9g, 1pz4, 1q8d, 1r6j, 1r7j, 1riy, 1rlk, 1rw1, 1rwj, 1sbx, 1t07, 1tg0, 1tgr, 1ts9, 1ttz, 1tuk, 1tuw, 1u2h, 1u3y, 1u84, 1u9p, 1ub9, 1ucs, 1uj8, 1uoy, 1use, 1usm, 1v05, 1v2z, 1v70, 1vbw, 1vjk, 1vyi, 1whz, 1wpa, 1wri, 1xak, 1xbi, 1xe1, 1xg8, 1xw3, 1y0n, 1y6x, 1y9l, 1ygt, 1yib, 1yu5, 1yzm, 1z0p, 1z21, 1z96, 1zld, 1zt3, 1zva, 1zzk, 2asc, 2b1y, 2b8i, 2bkf, 2brf, 2c60, 2cbo, 2cg7, 2cmp, 2cwr, 2cwy, 2cyj, 2d3d, 2d9r, 2e3h, 2ea9, 2efv, 2es9, 2ewh, 2ewk, 2ewt, 2f60, 2fb6, 2fht, 2fi0, 2fq3, 2fu2, 2g7o, 2gkr, 2gpi, 2gyz, 2h8e, 2h9u, 2haz, 2hc8, 2hdz, 2hl7, 2hlr, 2hpj, 2i4a, 2i5f, 2i6v, 2iay, 2igp, 2ip6, 2ivy, 2j6b, 2j8b, 2j97, 2jku, 2nml, 2nqw, 2ns0, 2nsc, 2nuh, 2o0q, 2o1k, 2o37, 2o4t, 2o9u, 2od5, 2ooa, 2oqk, 2oqq, 2ouf, 2ovg, 2oxo, 2p5k, 2p6v, 2pk8, 2pnd, 2ppn, 2pst, 2py0, 2q2f, 2q79, 2qff, 2qmt, 2qsb, 2qsk, 2qtd, 2qvo, 2qyw, 2r39, 2r4q, 2rff, 2rh2, 2rhf, 2ril, 2uux, 2v75, 2vc8, 2vkl, 2vq4, 2vsd, 2wbx, 2wj5, 2wkd, 2wuj, 2x3g, 2xfd, 2xnq, 2yv4, 2yvi, 2yxf, 2yzt, 2zqe, 2zqm, 2zxy, 3a0s, 3a38, 3a4c, 3adg, 3agn, 3b64, 3bjo, 3bn0, 3bn7, 3bri, 3bt4, 3bv8, 3c0f, 3ca7, 3ce7, 3cec, 3ci9, 3cq1, 3ctr, 3cw3, 3df8, 3dml, 3dqy, 3e0e, 3e21, 3e56, 3e9v, 3eaz, 3efg, 3emi, 3enu, 3ewg, 3exy, 3f14, 3f2e, 3f40, 3fbl, 3fdr, 3ff2, 3ff5, 3ffy, 3fkc, 3fmy, 3ft7, 3g21, 3g2b, 3ghf, 3goe, 3gv3, 3h01, 3h36, 3h8h, 3h9w, 3hgl, 3hms, 3hnx, 3hqx, 3hrl, 3hro, 3hz7, 3i8z, 3idw, 3im3, 3iv4, 3jsc, 3jsr, 3jtz, 3ju3, 3jvl, 3k0x, 3k3v, 3kkf, 3kp8, 3kw6, 3kzd, 3l32, 3l4h, 3l9a, 3lax, 3lbj, 3ldc, 3le4, 3lwc, 3lyg, 3lyw, 3msh, 3mwz, 3mxz, 3n3f, 3nbm, 3npd, 3nrw, 3nx6, 3nzl, 3oiz, 3oou, 3osh and 3phn.

The program described has also been tested on a pool of 15 structures (Millán *et al.*, 2015) first solved with previous *ARCIMBOLDO* implementations.

2.4. Libraries of local folds

For the *ARCIMBOLDO_BORGES* tests, six precomputed libraries containing geometrically clustered variations of a particular fold were used. Two of them contain pairs of contiguous helices: parallel and antiparallel, respectively. Three correspond to 20 amino acids arranged in three-stranded β-sheets in antiparallel, parallel and parallel–antiparallel dispositions. The last contains a sample of disulfide bridges linking two tetrapeptides in all possible conformations. These libraries of models superimposed to match the generating template were computed with *BORGES* (Sammito *et al.*, 2013) and are available for download from our website, along with libraries of 24 amino acids arranged in four strands with different relative arrangements (http://chango.ibmb.csic.es/download). A library with four antiparallel β-strands has been used redundantly on six cases previously solved with the described library of three-stranded antiparallel sheets to confirm its validity. This and other new libraries will be incorporated into our public repository. Besides the libraries used in this study, a library of DNA-binding motifs is also downloadable (Pröpper *et al.*, 2014).

3. Results and discussion

3.1. Implementation

ARCIMBOLDO_LITE, *ARCIMBOLDO_SHREDDER* and *ARCIMBOLDO_BORGES* have been thoroughly rewritten in the course of the last year. The *ARCIMBOLDO_LITE* implementation is deployed as a binary file and has been designed to run on a single multicore machine. The need for a MySQL database present in previous implementations has been removed, as it is intended to solve easier target structures ranging from 90 to 200 amino acids with complete diffraction data reaching 2.0 Å resolution or beyond. Moreover, particular stress has been placed on simplifying the input, reducing it to minimal requirements and producing helpful output, in particular direct troubleshooting instructions and clear identification of solution files. The program provides default values for all parameters except for the basic specification of experimental data: paths to the reflection data files in `.hkl` (Sheldrick, 2008) and `.mtz` (Winn *et al.*, 2011) format have to be input, as well as the LABELS identifying the data in the `.mtz` file and the contents of the asymmetric unit (molecular weight and number of copies). Paths to *Phaser* and *SHELXE* have to be set unless they are already defined for the user within the system. The default search is exemplified as two copies of a model helix composed of 14 alanines, but the user should select the most appropriate hypothesis according to secondary-structure prediction or previous knowledge about the target structure. For model polyalanine helices, an inbuilt

library is provided and only the length needs to be specified. Model fragments for some cofactors and ligands with high scattering power are also provided internally, in particular for iron–sulfur clusters and the haem group. Information on defaults and usage for all mandatory and optional parameters can be obtained by typing `ARCIMBOLDO_LITE -b` in the terminal window.

Fig. 1 displays a diagram summarizing the flow. The program starts by reading and validating the instruction file. It will substitute default values by user-input choices while checking for formal correctness and an acceptable parameter range. Next, checks are performed to customize the job, adapting it to the experimental data and available hardware, as well as to prevent common errors that cause run termination at a more advanced stage.

(i) The number of physical cores in the computer is read from `/proc/cpuinfo` (in Mac OS this information is extracted with `sysctl -n hw.ncpu`).

(ii) External search models are confirmed to be accessible files in valid PDB format containing at least one atom. Format problems are corrected if possible.

(iii) Paths to files and work directories are verified in terms of existence, accessibility and user permissions.

(iv) If a sequence is provided to compute the molecular weight, its validity is checked.

(v) Python, *Phaser* and *SHELXE* versions are checked for compatibility with the deployed binary.

(vi) A *Phaser* job is launched to validate configuration and data compatibility while generating a corrected data file with anisotropically scaled amplitudes for use in subsequent jobs.

(vii) A *SHELXE* job is launched to test the input line and data format, identifying amplitudes or intensities.

(viii) The space group, if specified, is tested to be a Sohncke space group in a standard setting.

Failure of any of these tests will cause the program to stop, with an error message printed to the standard output indicating the cause. Furthermore, if the resolution of the experimental data does not reach 2 Å a warning will be issued, while below 2.5 Å resolution the program will terminate.

Results are output in xml and html format and are constantly updated for the user to monitor the program flow. The html file (Fig. 2) echoes all of the parameters used in the run and a table of results and figures of merit characterizing final and partial solutions. The table can be sorted by any of its entries. The html page provides links to the files containing the best map and trace.

ARCIMBOLDO_LITE runs in two macrocycles. During the first one *Phaser* sequentially places all requested copies of the search model(s) specified through the configuration file and partial structures are ranked and possibly optimized by trimming against the CC (Fujinaga & Read, 1987) with *SHELXE* (Sheldrick & Gould, 1995). In the second macrocycle, *ARCIMBOLDO* selects from all of the partial solutions produced a number of solutions one less than the number of physical cores in the machine to be expanded through density modification and autotracing with *SHELXE*. To render calculations manageable, hard limits are imposed on the maximum number of solutions after each step. These limits are set in relation to the number of cores available. Therefore, a more powerful machine will by default pursue more partial solutions and may solve a structure that a more limited machine would fail to phase unless a longer job is customized. By default, negative LLG values are discarded and, even though this parameter can be overridden through the `.bor` file, the alternative of revising the expected asymmetric unit composition and/or data resolution appears to be preferable.

Phaser jobs are executed as independent, monitored processes and input is given through keyword files for compatibility with all environments and middleware. Firstly, a rotation search (Storoni *et al.*, 2004) is computed for the input model and solutions within 75% of the top LLG are selected. These are clustered within an angular difference of 15°, taking symmetry into account, after converting the Euler angles to quaternions or rotation matrices. If, exceptionally, the number of rotation peaks exceeds 10 000, numerical *k*-means clustering is used as a filter prior to geometrical clustering. To save time, an ancillary Cantor pairing value is used. The rotation characterized by the highest LLG is used as a reference to determine similar rotations. The process is iterated over the nonclustered rotations until the list has been scanned.

Translation performance (McCoy *et al.*, 2005) is evaluated jointly for each cluster of rotation solutions, limiting the maximum number of solutions to be sent to packing and refinement. In this way, variability is preserved if a given fragment can be placed in several parts of a structure. For the placed fragments, an initial CC is computed with *SHELXE*, optionally eliminating residues to increase this figure of merit. PDB optimization (Sheldrick & Gould, 1995) is usually not

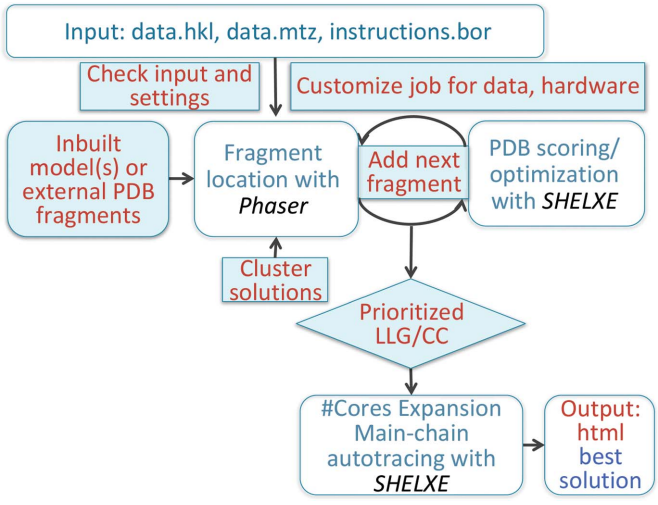


Figure 1
ARCIMBOLDO_LITE flow scheme. *ARCIMBOLDO* operations are labelled in red. Input data and program settings are validated and the run is configured with parameterization defaults depending on data resolution and limiting calculations by the number of cores in the workstation. In a first macrocycle all copies of the search fragments are placed with *Phaser* and optimized with *SHELXE*. Similar solutions are clustered. A number of substructure solutions equal to the number of physical cores minus one is subject to density modification and autotracing with *SHELXE*, prioritizing according to top LLG and CC.

==========Summary of your data==========

SPACEGROUP: P 1 21 1
CELL DIMENSIONS: 23.26, 60.44, 23.59, 90.00, 90.99, 90.00
RESOLUTION: 0.92
NUMBER OF UNIQUE REFLECTIONS: 35634.00

2. STEP: Locating Sequentially the Fragments

Fragment 1																						
Cluster	Rotation Function					Translation Function					Packing					Rigid Body Refinement			Initial CC	Best Trace CC/aa		
	#Rots.	Top LLG	Mean LLG	Top Zscore	Mean Zscore	#Trans.	Top LLG	Mean LLG	Top Zscore	Mean Zscore	#Sol.	Top LLG	Mean LLG	Top Zscore	Mean Zscore	#Sol.	Top LLG	Mean LLG	After Refinement CC	Cycle	CC	#Res. traced
0	20	18.99	14.21	4.62	3.73	2	35.51	32.35	6.90	5.71	2	35.51	32.35	6.90	5.71	2	39.20	37.75	7.56			
1	1	16.68	16.68	4.19	4.19	4	25.36	23.54	4.01	3.50	4	25.36	23.54	4.01	3.50	4	34.80	28.17	7.08			
2	3	16.67	14.46	4.19	3.78	7	22.00	21.12	3.62	3.00	7	22.00	21.12	3.62	3.00	7	24.00	22.34	6.20			
3	1	16.33	16.33	4.12	4.12	4	22.49	21.67	3.46	3.18	4	22.49	21.67	3.46	3.18	4	24.60	23.20	6.56			
4	3	13.79	13.54	3.65	3.61	2	26.09	25.99	5.18	5.15	2	26.09	25.99	5.18	5.15	2	28.90	28.80	3.84			
5	1	13.71	13.71	3.64	3.64	1	21.86	21.86	4.04	4.04	1	21.86	21.86	4.04	4.04	1	22.90	22.90	7.59			

Fragment 2																						
Cluster	Rotation Function					Translation Function					Packing					Rigid Body Refinement			Initial CC	Best Trace CC/aa		
	#Rots.	Top LLG	Mean LLG	Top Zscore	Mean Zscore	#Trans.	Top LLG	Mean LLG	Top Zscore	Mean Zscore	#Sol.	Top LLG	Mean LLG	Top Zscore	Mean Zscore	#Sol.	Top LLG	Mean LLG	After Refinement CC	Cycle	CC	#Res. traced
(0, 6)	11	56.31	50.27	4.43	3.58	7	108.16	93.75	10.25	9.18	7	108.16	93.75	10.25	9.18	4	124.90	118.85	13.38			
(0, 9)	33	52.94	49.95	3.79	3.51	21	117.06	98.86	12.30	10.05	21	117.06	98.86	12.30	10.05	14	125.00	106.64	12.16			
(1, 9)	18	48.93	40.73	3.85	3.73	92	99.10	52.27	10.29	5.14	3	99.10	91.80	10.29	9.31	3	103.90	98.40	9.99			
(1, 11)	96	54.19	42.45	4.86	3.88	163	119.42	59.00	12.64	5.70	13	119.42	94.96	12.64	9.67	5	124.60	111.36	13.08	4	45.79	81
(4, 6)	2	45.49	45.49	4.42	4.41	2	91.19	88.89	9.53	9.15	2	91.19	88.89	9.53	9.15	2	103.90	101.30	10.00			
(4, 11)	42	47.61	42.52	4.80	3.85	28	99.58	83.12	10.81	8.91	28	99.58	83.12	10.81	8.91	12	105.70	95.93	10.30			

Show All Hide Not Relevant

The current best solution is: th10_0_0xx2FR1_2-1.pdb with FINALCC: 45.79 and n. residues traced 81
file is: /home/dawid/1l9l/8_EXP_LIBRARY/4/0/th10_0_0xx2FR1_2-1.pdb

- FRF: Pos. in Rank: **48** LLG: **52.27** ZSCORE: **4.46** Top LLG in Cluster (**1, 11**): **56.31** Top ZSCORE in Cluster (**1, 11**): **4.94**
- REFINEMENT ROTATION AND MODEL

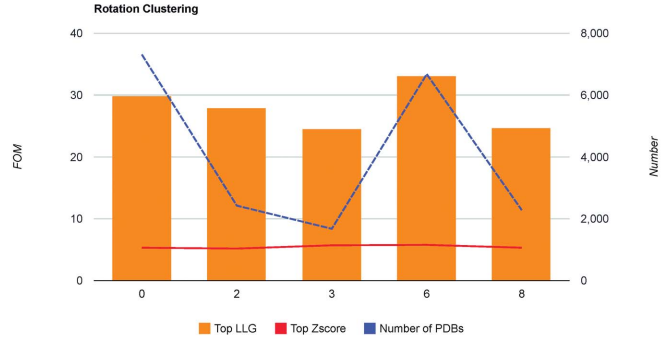
- FTF: Pos. in Rank: **1** LLG: **119.42** ZSCORE: **12.64** Top LLG in Cluster (**1, 11**): **119.42** Top ZSCORE in Cluster (**1, 11**): **12.64**
- PACK: Pos. in Rank: **1** LLG: **119.42** ZSCORE: **12.64** Top LLG in Cluster (**1, 11**): **119.42** Top ZSCORE in Cluster (**1, 11**): **12.64**
- RNP: Pos. in Rank: **19** LLG: **124.60** ZSCORE: **0.00** Top LLG in Cluster (**1, 11**): **125.00** Top ZSCORE in Cluster (**1, 11**): **0.00**
- INITIAL CC
  - After Refinement: Pos. in Rank: **5** INITCC: **12.67** Top INITCC in Cluster (**1, 11**): **13.38**
- EXPANSION

Cycle 4:
Final CC: **45.79%** N. Residues Traced: **81.00**

(*a*)

#Cluster	#Rotations	#Distinct Pdb	Top LLG	LLG Mean	Top Zscore	Zscore Mean
0	7310	4018	29.81	17.99	5.33	3.64
2	2428	1918	27.95	15.44	5.19	3.38
3	1679	1368	24.50	13.73	5.72	3.19

Show All Hide Not Relevant



#Cluster	Rotation and Model Refinement					Translation Function					Packing					Rigid Body Refinement			Initial CC		Best Trace CC/aa			
	#Rots.	Top LLG	Mean LLG	Top Zscore	Mean Zscore	#Trans.	Top LLG	Mean LLG	Top Zscore	Mean Zscore	#Sol.	Top LLG	Mean LLG	Top Zscore	Mean Zscore	#Sol.	Top LLG	Mean LLG	Before Refinement CC	After Refinement CC	MODE	Cycle	CC	#Res. traced
0	7310	29.81	17.99	5.33	3.64	700	57.20	39.86	8.07	5.54	267	57.20	40.11	8.07	5.43	267	63.60	42.44	7.49	7.85	FAST	1	45.4	103

Show All Hide Not Relevant

The current best solution is: 1les_0_115.pdb with FINALCC: 45.40 and n. residues traced 103
file is: /home/alfotest/1v70/11_EXP/0/1/0/1les_0_115_rottra.pdb

- FRF: Pos. in Rank: **6758** LLG: **26.50** ZSCORE: **4.04** Top LLG in Cluster **0**: **33.11** Top ZSCORE in Cluster **0**: **5.79**
- REFINEMENT ROTATION AND MODEL

- FTF: Pos. in Rank: **1** LLG: **57.20** ZSCORE: **7.94** Top LLG in Cluster **0**: **57.20** Top ZSCORE in Cluster **0**: **8.07**
- PACK: Pos. in Rank: **1** LLG: **57.20** ZSCORE: **7.94** Top LLG in Cluster **0**: **57.20** Top ZSCORE in Cluster **0**: **8.07**
- RNP: Pos. in Rank: **5** LLG: **58.50** ZSCORE: **0.00** Top LLG in Cluster **0**: **63.60** Top ZSCORE in Cluster **0**: **0.00**
- INITIAL CC
  - Before Refinement: Pos. in Rank: **1** INITCC: **7.49** Top INITCC in Cluster **0**: **7.49**
  - After Refinement: Pos. in Rank: **5** INITCC: **7.25** Top INITCC in Cluster **0**: **7.85**
- EXPANSION

Cycle 1:
Final CC: **45.40%** N. Residues Traced: **103.00**

It seems you have a good solution!
Here you can find the best solution and map for further refinement.
BORGES will end now.

(*b*)

Figure 2
Html output for (*a*) the *ARCIMBOLDO_LITE* solution of PDB entry 1l9l and (*b*) the *ARCIMBOLDO_BORGES* solution of PDB entry 1v70.

Table 2
Performance of *ARCIMBOLDO_LITE* and *ARCIMBOLDO_BORGES* on a pool of 294 test structures.

A detailed table with PDB codes and parameterizations is available as Supporting Information.

Test	Pool of structures	Solved by *ARCIMBOLDO_LITE*	Solved by *ARCIMBOLDO_BORGES*	Total solved
Total	294	143	38	181
Resolution				
1–0.54 Å	24	14	5	19
1.3–1.0 Å	43	25	8	33
1.6–1.3 Å	78	38	12	50
2.2–1.6 Å	149	66	13	79
Fragments				
Helices	230†	139	3	142
Fe clusters, haem	5	4	0	4
β-Sheets	147	0	34	34
Antiparallel (udu)	131	—	31	31
Parallel–antiparallel (uud)	24	—	2	2
Parallel (uuu)	13	—	1	1
Antiparallel (udud)	6 (70)	—	6	6
Disulfide bridges	20	0	1	1

† 1 curved; 2 antiparallel (ud).

critical if the search fragment is a helix, but is the default for libraries of folds.

Subsequent searches for additional copies of the same or different fragments are also subjected to clustering of rotation solutions. Clusters are identified by combinations of numbers corresponding to the original and the new rotation peaks. Equivalent combinations are kept apart for simplicity when calculating the translation search but are joined for the refinement stage so that they may be fused if equal. Clustering and evaluation results are written into formatted text files, which can be used by the program to continue an interrupted run. Relaunching an interrupted program will read the present folders and continue from the latest stage completed, which implies that a change in parameterization will only be effective if all previous output for this and subsequent steps has been deleted.

After all fragment-placement operations have been performed, selected solutions are subjected to density modification and autotracing with *SHELXE* (Sheldrick, 2010). Selected solutions may correspond to intermediate fragment-placement rounds as they are chosen to produce the highest CC and LLG. The default number of solutions probed is equal to the number of physical cores in the machine minus one.

If the user specifies a PDB format file with the extension `.ent` in the configuration file, analysis of results compared with this reference is performed. Goodness of partial solutions is evaluated through the r.m.s.d. of rotations after optimal translation and mean phase errors *versus* the reference for placed fragments and final solutions. This information is not output to the standard `.html` file, but has been used in the present study of test structures to evaluate results and to guide

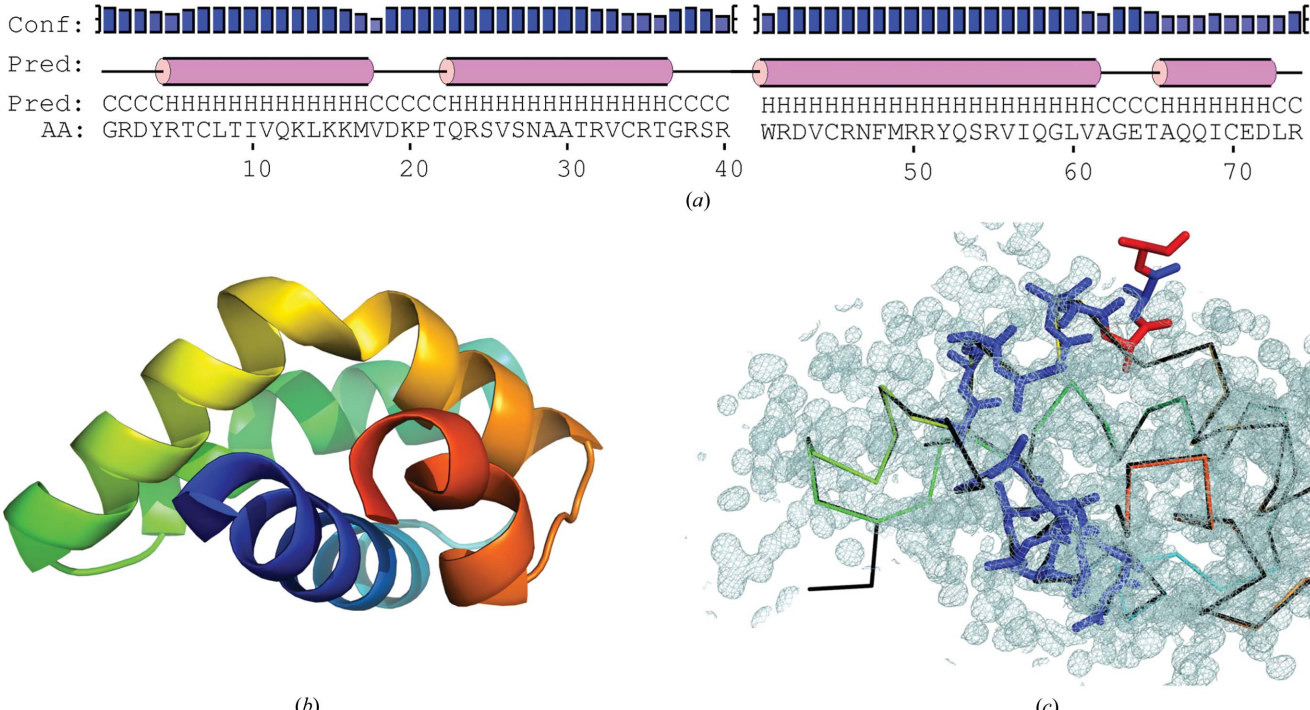


Figure 3
ARCIMBOLDO_LITE solution of PDB entry 1191. (*a*) Secondary-structure prediction. (*b*) Structure in rainbow-coloured cartoon representation. (*c*) *SHELXE* electron-density map of a solution; the trace is shown as a coil and both placed helical fragments are shown as sticks, with residues eliminated during PDB optimization displayed in red.

default parameterization. The procedure starts by extracting all of the helices (or any other search fragment given as input) found in the `.ent` file. For each one of them a rotation LLG at their current position is computed in *Phaser* through the GYRE mode to provide an estimate of the value range for a correct solution. For the placed fragments, the *F*-weighted mean phase difference to the reference structure (Lunin & Woolfson, 1993) is computed with *SHELXE*.

3.2. Timing benchmarks on various hardware

The structure of the C-terminal domain of human anti-termination factor 5 (PDB entry 2iu1; Bieniossek *et al.*, 2006) was used on 12 machines with hardware ranging from four to 24 physical cores and a minimum of 1 GB of RAM per core. A case is included in which 512 MB of RAM was shared by two cores, showing that the associated increase in run time (RT) makes such a configuration unsuitable. Conversely, a machine endowed with CPU overclocking up to 4 GHz and almost 8 GB of RAM per core rendered an RT that was half that of a typical workstation. Table 1 shows the time taken for structure solution on the various machines. It ranged from one to several hours in the default mode that runs as many processes as there are physical cores detected, reserving one thread for program calculations, mainly clustering and combination of solutions, and running as many *Phaser* and *SHELXE* jobs as there are cores minus one. It is worth remarking that the job computed is tailored to the hardware and thus a greater or lesser number of solutions are pursued depending on the number of cores. Accordingly, the time required by a more powerful machine may not be much less, but it may succeed in solving a structure where a machine with fewer cores fails.

3.3. Performance of *ARCIMBOLDO* tests

Table 2 summarizes the single-workstation performance of *ARCIMBOLDO_LITE* and *ARCIMBOLDO_BORGES* on a set of test structures. An initial baseline was set by running *ARCIMBOLDO_LITE* blindly on the pool of 294 structures with a fragment search configured to find two polyalanine helices of 14 residues and using a common *SHELXE* parameterization. This resulted in the solution of 100 of the 294 structures. *SHELXE* was set to perform five iterations involving 30 cycles of density modification and autotracing, increasing the time for random search 20-fold and taking into account the presence of helices, a solvent content of 55% and fragment optimization against the CC. In many cases, three iterations and a tenfold time increase to seed autotracing led to structure solution. Accounting for the particularities of individual structures opens a way to improve the success rate with minimum intervention.

3.3.1. Ultrahigh resolution (beyond 1 Å). Within the pool of 294 test cases, 24 are atomic resolution structures diffracting to 1 Å resolution or beyond and, of these, 11 contain at least one helix of six amino acids or longer. Such short helices are useful only in this resolution context, where some of the structures are very small. All such helical structures in this group were eventually solved. The behaviour of this group will be illustrated by the case of PDB entry 1l9l in space group $P2_1$ with diffraction to 0.92 Å resolution. Fig. 2(*a*) displays the html output from the *ARCIMBOLDO_LITE* run. The all-helical structure displays three long helices of 18, 15 and 25 residues, as expected from the secondary-structure prediction displayed in Fig. 3(*a*). The long third helix is divided into two stretches by a sharp kink. This scenario would appear to be deceptively simple, as the data quality is excellent, but the blind parameterization that was initially used failed to solve it. The same happened with a number of other structures in this group during the first, blind test. Instead, search trials with different helix lengths and parameterization tests led to a solution when searching for two helices of ten residues and increasing the number of density-modification cycles, turning on low-density elimination (Yao *et al.*, 2005), decreasing the solvent content and increasing the rounds of autotracing but lowering the time seed (*SHELXE* command line `-m200 -v0.5 -a10 -t1 -q -f -s0.25 -o`).

This result suggested the convenience of taking into account particularities anticipated from data resolution to automate default parameterization. Some common settings are adopted, for instance the presence of helices is always assumed for tracing whenever a helical search model is used and default memory allocation is increased to fit the hardware. The model helix of 14 amino acids fits most helical stretches within an r.m.s.d. of 0.3 Å for the C_α atoms. The availability of extremely high resolution data allows smaller fragments to be extended, but suffers more severely from the lack of accuracy in the model. As can be seen in Figs. 3(*b*) and 3(*c*), they contained errors as one of the helices is misplaced. A solution from an imperfect structure is frequently found, and wrong residues may be initially eliminated to increase the CC. Thus, PDB optimization is switched on and, given sufficient cycles and high resolution, the remaining errors are slowly erased. Starting from shorter helices may be more effective as accuracy is enhanced. More density-modification cycles are beneficial and low-density elimination should be turned on, whereas tracing of the atomic map is simpler and can be accelerated with a smaller time parameter. The solvent content is typically extremely low for this resolution group.

3.3.2. High resolution (1.3–1 Å). In the group of 33 structures with helices and resolutions between 1 and 1.3 Å, 25 could be solved when searching for predicted helices and running eight iterations of density modification and autotracing; the number of density-modification cycles was not increased so much this time, setting it to 100, and low-density elimination was given a lower weight of 0.25. A solvent content of 35% was assumed and a time seed of 10 was used. As the resolution becomes lower, more starting information is needed for successful expansion; thus, longer helices of 14 amino acids rather than the extremely small helices of less than ten residues were used as search fragments.

3.3.3. Medium resolution (1.6–1.3 Å). In the pool of 62 structures with helices and resolutions between 1.3 and 1.6 Å, 38 could be solved when searching for predicted helices and running eight autotracing cycles interspersed with 50 density-modification cycles, with low-density elimination set to

0.1 and a time seed of 10 and assuming 45% solvent content.

3.3.4. Low resolution (2.2–1.6 Å). At the lower resolution end, 66 of 149 structures were solved. 124 of the structures contained at least 8% helices, which is the lowest limit found to be accessible at ultrahigh resolution, but in this context such a low percentage is clearly not sufficient to extend helical fragments. More fragments of longer size but tolerating higher deviations provide successful starting substructures. The lower the resolution, the more difficult it becomes to successfully extend the partial structures, even if they are located. The number of density-modification cycles is gradually decreased from 15 at the higher resolution end of this span to ten below 2 Å, a solvent content of 50–60% is assumed and low-density elimination is entirely switched off. The time seed for autotracing is increased.

3.3.5. Alternative search fragments. Five of the structures in the test pool contained cofactors, in particular Fe_4S_4, Fe_2S_2 and a haem group. Despite the presence of helices, solving such structures without first accounting for the electron-rich group failed at the stage of fragment location. As the presence of cofactors is easily predictable, their use as search fragments appeared to be indicated. The search for them may be combined with that for other fragments, but in the cases tested the location of model fragments from the *REFMAC*5 dictionary (Murshudov *et al.*, 2011; Vagin *et al.*, 2004) was sufficient for a successful expansion to the full structure in four of the five cases. The parameterization was set to match the resolution group, but the initial fragment was kept throughout the first autotracing cycles (*SHELXE* `-K` flag) as only polypeptide main chain is built into the electron-density map (Sheldrick, 2010). These fragments may be specified by their three-letter library name in the instruction `.bor` file as they have been included along with the model helix in the internal library.

The straight model helix built into *ARCIMBOLDO* may not be the optimal search fragment for helical structures containing markedly curved or otherwise distorted helices. This is the case for the 90-amino-acid, 1.5 Å resolution structure with PDB code 3l32 in space group $I4_122$. Solution with *ARCIMBOLDO_LITE* failed, whereas using a collection of 162 models for differently distorted helices of 18 amino acids extracted from the PDB and gathered in a library rendered a solution. PDB entry 3l32 is one of the 38 cases solved with libraries of similar models with a single-computer implementation of *ARCIMBOLDO_BORGES*.

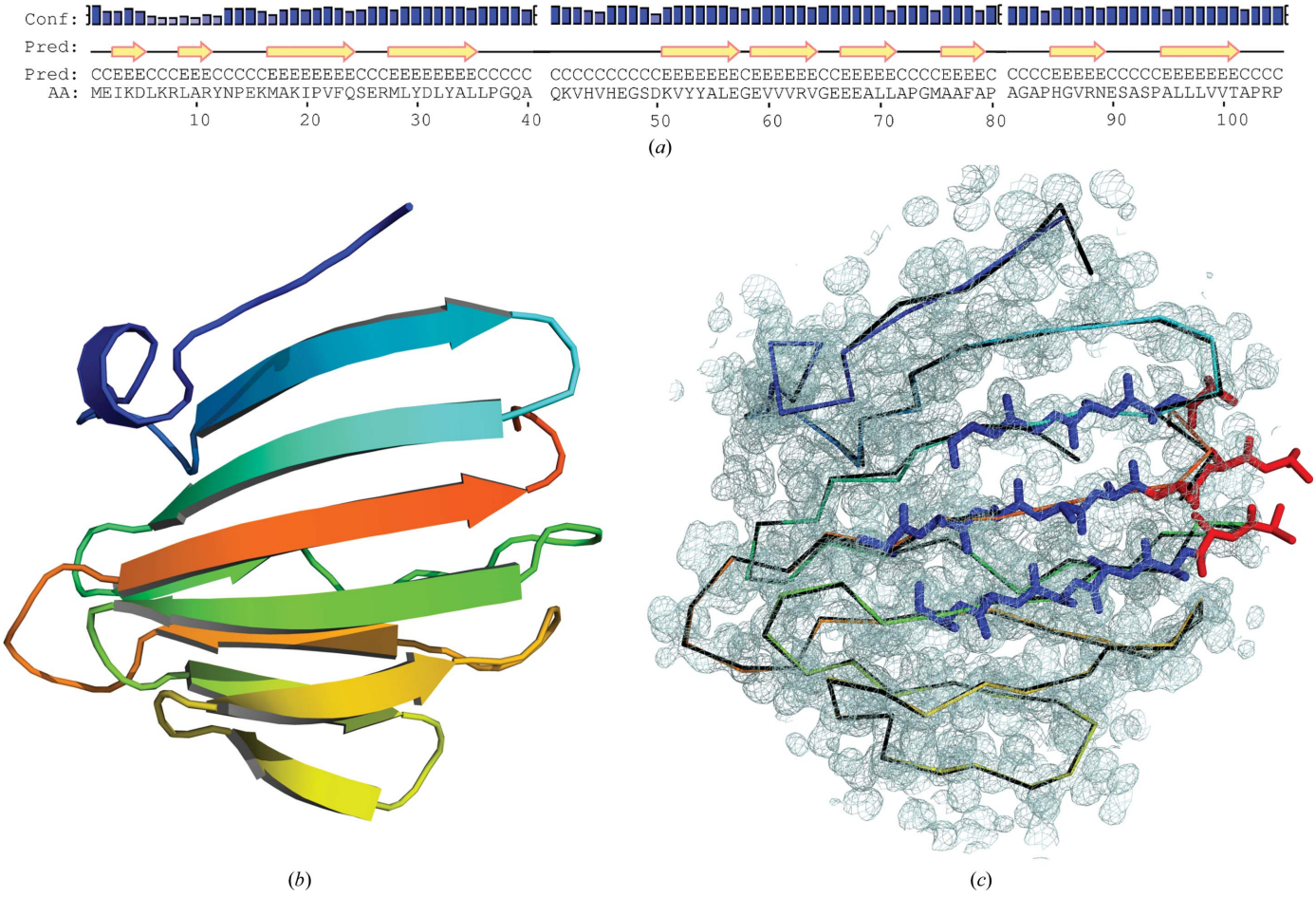


Figure 4
ARCIMBOLDO_BORGES solution of PDB entry 1v70. (*a*) Secondary-structure prediction. (*b*) Structure in rainbow-coloured cartoon representation. (*c*) *SHELXE* electron-density map of a solution; the trace is shown as a coil and the placed antiparallel three-stranded fragments are shown as sticks, with residues eliminated during PDB optimization displayed in red.

3.4. Performance of *BORGES* libraries on test structures

The pool of test structures contains 147 cases in which β-strands are the only secondary-structure fragments or make up most of the fold. In addition, 20 structures feature four or more disulfide bridges anchoring a peculiar fold, as is typical in small toxin inhibitors. Such cases are outside the scope of an *ARCIMBOLDO_LITE* secondary-structure fragment search, but can be solved with libraries of small folds in *ARCIMBOLDO_BORGES* (Sammito *et al.*, 2013). Although the method is better suited to large computing resources, the new implementation can run either connecting to a local or a remote grid or on a single multicore machine. Again, automatic parameterization is related to the available hardware, and only four rotation clusters will be sequentially evaluated by default when the program is run on a single machine. The test structures in this pool are small enough to be computed on the eight-core workstations used for the *ARCIMBOLDO_LITE* tests. The typical run time was 2–5 d. The same precomputed libraries were used in all tests. They are available for download from the *ARCIMBOLDO* web page. Three libraries contain β-sheets with three strands in parallel, parallel–antiparallel and antiparallel arrangements. The latter have been the most successful, being more frequent; 131 of the cases in the group of β-sheet-containing structures display this arrangement of strands. In addition to the library of single distorted helices mentioned, two libraries contain two contiguous helices in parallel and antiparallel dispositions, respectively. They are especially indicated for coiled coils, as single helices tend to be placed in the same position in fragment searches. Finally, a library of two tetrapeptides linked by a disulfide bridge has also been tested.

Two test cases were solved with the library of helices, and 31 with the antiparallel, two with the parallel–antiparallel and one with the parallel three-stranded library, respectively. One structure was solved with the disulfide-linked peptide library. The optimal parameterization has been found to be different for structures mainly composed of β-strands and is accordingly set as default for *ARCIMBOLDO_BORGES* as exemplified in Fig. 4. If the program is launched with one of the nonhelical libraries, the default expansion stage will not search for helices, will use a lower solvent content than in the helical case, will keep the original fragment for *K* cycles and will perform more autotracing iterations. As in *ARCIMBOLDO_LITE*, a line in the `.bor` file setting the parameterization for *SHELXE* will override any of the inbuilt defaults, but leaving this line unset will apply the resolution-tailored and hardware-tailored values described.

3.5. Statistics and performance

Table 2 summarizes the single-workstation performance of *ARCIMBOLDO_LITE* and *ARCIMBOLDO_BORGES* on the 294 test structures. Beyond the 181 structures (62%) solved as described, some could be solved if run on a computer with more cores or if the parameterization was further adapted to run a more exhaustive attempt or even with less restrictions on a grid of computers. For example, the 114-amino-acid structure with PDB code 3f40 in space group $C222_1$ at 1.27 Å resolution renders correct substructures composed of two helices of 12 and eight alanines, respectively, but these are not prioritized among the seven expanded helices. Running *ARCIMBOLDO_LITE* on a machine with more cores or a grid would solve it. Alternatively, the number of expansions to try may be increased by the user. Regarding *ARCIMBOLDO_BORGES*, the 105-amino-acid structure with PDB code 2haz at 1.7 Å resolution is solved automatically on a double Xeon workstation with similar features to those used for benchmarking performance in this study but endowed with 3 GB rather than 2 GB of RAM per core. This case has not been counted as solved in our results.

High-resolution cases yield a higher success rate: 78% for atomic resolution structures, decreasing to 64% for data in the 1.6–1.3 Å resolution range and to 53% below 1.6 Å resolution. For structures mainly composed of β-strands, phasing with the secondary-structure fragments has not found any practical use within *ARCIMBOLDO*, whereas precomputed libraries of β-sheets have been successful in 22% of the identified cases. Not all have been tested, as structures previously solved with helices were eliminated from the pool. In the same manner, the more frequent antiparallel disposition of three strands was tested first and only structures where this fold would not work were further considered for phasing with parallel–antiparallel or parallel libraries. Six structures were phased both with libraries containing three-stranded and four-stranded models. They are counted only once in the total of 181 structures. The four-stranded libraries are also available for download and might be useful for larger structures. In the same way, two structures were phased with the library of disulfide bridges and one with a library of polyalanine helices extracted from real structures and presenting different deviations from the regular model helix. This structure, PDB entry 3l32, could not be solved with a model helix, and the use of the library illustrates that a starting model with a lower r.m.s.d may succeed where the model fails and that the more exhaustive run needed may be worthwhile to solve an unknown structure.

4. Concluding remarks

ARCIMBOLDO_LITE succeeds in solving 143 out of a pool of 294 test structures on an eight-core workstation, with sizes ranging from 40 to 120 residues and resolutions between 2.2 and 0.54 Å. The fragments placed are predominantly straight polyalanine helices and, in three cases, iron cofactors. *ARCIMBOLDO_BORGES* solves a further 38 structures using precomputed libraries of small folds consisting of contiguous helices, three-stranded β-sheets and disulfide-linked tetrapeptides. The run times for *ARCIMBOLDO_LITE* jobs are typically a couple of hours, whereas *ARCIMBOLDO_BORGES* requires one or two days on a single machine.

The results of this study have been used to incorporate default parameterization as a function of data resolution and search fragment. Leaving the *SHELXE* line unset in the input `.bor` file will activate *ARCIMBOLDO* defaults, which may differ from *SHELXE* defaults.

The binaries for Linux and MAC OS have been validated for the latest distributions and can be downloaded from http://chango.ibmb.csic.es/ARCIMBOLDO_LITE. The same binaries that run on a single workstation can be run to distribute jobs on a local or remote Condor, SGE/Opengrid or Torque/MOAB grid. The precomputed libraries can be downloaded together with the *ARCIMBOLDO_BORGES* binary.

Acknowledgements

We thank George M. Sheldrick, Randy J. Read and Airlie McCoy for helpful discussions. This work was supported by grants BFU2012-35367 and BIO2013-49604-EXP (the Spanish Ministry of Economy and Competitiveness) and Generalitat de Catalunya (2014SGR-997). RJB acknowledges a grant from the Brazilian Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.
Bibby, J., Keegan, R. M., Mayans, O., Winn, M. D. & Rigden, D. J. (2012). *Acta Cryst.* D**68**, 1622–1631.
Bibby, J., Keegan, R. M., Mayans, O., Winn, M. D. & Rigden, D. J. (2013). *Acta Cryst.* D**69**, 2194–2201.
Bieniossek, C., Schütz, P., Bumann, M., Limacher, A., Usón, I. & Baumann, U. (2006). *J. Mol. Biol.* **360**, 457–465.
Burla, M. C., Carrozzini, B., Cascarano, G. L., Giacovazzo, C. & Polidori, G. (2011). *J. Appl. Cryst.* **44**, 1143–1151.
Burla, M. C., Carrozzini, B., Cascarano, G. L., Giacovazzo, C. & Polidori, G. (2012). *J. Appl. Cryst.* **45**, 1287–1294.
Burla, M. C., Giacovazzo, C. & Polidori, G. (2010). *J. Appl. Cryst.* **43**, 825–836.
Caliandro, R., Carrozzini, B., Cascarano, G. L., De Caro, L., Giacovazzo, C., Mazzone, A. & Siliqi, D. (2008). *J. Appl. Cryst.* **41**, 548–553.
Caliandro, R., Carrozzini, B., Cascarano, G. L., De Caro, L., Giacovazzo, C. & Siliqi, D. (2005). *Acta Cryst.* D**61**, 1080–1087.
Caliandro, R., Dibenedetto, D., Cascarano, G. L., Mazzone, A. & Nico, G. (2012). *Acta Cryst.* D**68**, 1–12.
Cole, C., Barber, J. D. & Barton, G. J. (2008). *Nucleic Acids Res.* **36**, W197–W201.
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.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Fujinaga, M. & Read, R. J. (1987). *J. Appl. Cryst.* **20**, 517–521.
Glykos, N. M. & Kokkinidis, M. (2003). *Acta Cryst.* D**59**, 709–718.
Hendrickson, W. A. (1991). *Science*, **254**, 51–58.
Hendrickson, W. A. (2013). *Acta Cryst.* A**69**, 51–59.
Hissa, D. C., Bezerra, G. A., Birner-Gruenberger, R., Silva, L. P., Usón, I., Gruber, K. & Melo, V. M. M. (2014). *Chembiochem*, **15**, 393–398.
Karle, J. & Hauptman, H. (1956). *Acta Cryst.* **9**, 635–651.
Keegan, R. M., Bibby, J., Thomas, J., Xu, D., Zhang, Y., Mayans, O., Winn, M. D. & Rigden, D. J. (2015). *Acta Cryst.* D**71**, 338–343.
Lunin, V. Y. & Woolfson, M. M. (1993). *Acta Cryst.* D**49**, 530–533.
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.
McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. (2005). *Acta Cryst.* D**61**, 458–464.
Millán, C., Sammito, M. & Usón, I. (2015). *IUCrJ*, **2**, 95–105.
Miller, R., DeTitta, G. T., Jones, R., Langs, D. A., Weeks, C. M. & Hauptman, H. A. (1993). *Science*, **259**, 1430–1433.
Morris, R. J., Blanc, E. & Bricogne, G. (2004). *Acta Cryst.* D**60**, 227–240.
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.
Oeffner, R. D., Bunkóczi, G., McCoy, A. J. & Read, R. J. (2013). *Acta Cryst.* D**69**, 2209–2215.
Pröpper, K., Meindl, K., Sammito, M., Dittrich, B., Sheldrick, G. M., Pohl, E. & Usón, I. (2014). *Acta Cryst.* D**70**, 1743–1757.
Qian, B., Raman, S., Das, R., Bradley, P., McCoy, A. J., Read, R. J. & Baker, D. (2007). *Nature (London)*, **450**, 259–264.
Refaat, L. S. & Woolfson, M. M. (1993). *Acta Cryst.* D**49**, 367–371.
Rigden, D. J., Keegan, R. M. & Winn, M. D. (2008). *Acta Cryst.* D**64**, 1288–1291.
Robertson, M. P., Chi, Y.-I. & Scott, W. G. (2010). *Methods*, **52**, 168–172.
Robertson, M. P. & Scott, W. G. (2008). *Acta Cryst.* D**64**, 738–744.
Rodríguez, D. D., Grosse, C., Himmel, S., González, C., de Ilarduya, I. M., Becker, S., Sheldrick, G. M. & Usón, I. (2009). *Nature Methods*, **6**, 651–653.
Rodríguez, D., Sammito, M., Meindl, K., de Ilarduya, I. M., Potratz, M., Sheldrick, G. M. & Usón, I. (2012). *Acta Cryst.* D**68**, 336–343.
Rossmann, M. G. (1972). *The Molecular Replacement Method*. New York: Gordon & Breach.
Sammito, M. D., Meindl, K., de Ilarduya, I. M., Millán, C., Artola-Recolons, C., Hermoso, J. A. & Usón, I. (2014). *FEBS J.* **281**, 4029–4045.
Sammito, M. D., Millán, C., Rodríguez, D. D., de Ilarduya, I. M., Meindl, K., De Marino, I., Petrillo, G., Buey, R. M., de Pereda, J. M., Zeth, K., Sheldrick, G. M. & Usón, I. (2013). *Nature Methods*, **10**, 1099–1101.
Sheldrick, G. M. (2002). *Z. Kristallogr.* **217**, 644–650.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Sheldrick, G. M. (2010). *Acta Cryst.* D**66**, 479–485.
Sheldrick, G. M., Gilmore, C. J., Hauptman, H. A., Weeks, C. M., Miller, R. & Usón, I. (2012). *International Tables for Crystallography*, 2nd online ed., edited by E. Arnold, D. M. Himmel & M. G. Rossmann, pp. 413–429. Chester: International Union of Crystallography.
Sheldrick, G. M. & Gould, R. O. (1995). *Acta Cryst.* B**51**, 423–431.
Shiono, M. & Woolfson, M. M. (1992). *Acta Cryst.* A**48**, 451–456.
Shrestha, R., Berenger, F. & Zhang, K. Y. J. (2011). *Acta Cryst.* D**67**, 804–812.
Shrestha, R. & Zhang, K. Y. J. (2015). *Acta Cryst.* D**71**, 304–312.
Storoni, L. C., McCoy, A. J. & Read, R. J. (2004). *Acta Cryst.* D**60**, 432–438.
Thorn, A. & Sheldrick, G. M. (2013). *Acta Cryst.* D**69**, 2251–2256.
Usón, I., Stevenson, C. E. M., Lawson, D. M. & Sheldrick, G. M. (2007). *Acta Cryst.* D**63**, 1069–1074.
Vagin, A. A., Steiner, R. A., Lebedev, A. A., Potterton, L., McNicholas, S., Long, F. & Murshudov, G. N. (2004). *Acta Cryst.* D**60**, 2184–2195.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Xu, D. & Zhang, Y. (2012). *Proteins*, **80**, 1715–1735.
Yao, J., Dodson, E. J., Wilson, K. S. & Woolfson, M. M. (2006). *Acta Cryst.* D**62**, 901–908.
Yao, J., Woolfson, M. M., Wilson, K. S. & Dodson, E. J. (2005). *Acta Cryst.* D**61**, 1465–1475.