ARP/wARP and molecular replacement: the next generation

Serge X. Cohen,[a]* Marouane Ben Jelloul,[a] Fei Long,[b] Alexei Vagin,[b] Puck Knipscheer,[a] Joyce Lebbink,[a] Titia K. Sixma,[a] Victor S. Lamzin,[c] Garib N. Murshudov[b] and Anastassis Perrakis[a]

[a]Department of Molecular Carcinogenesis, Netherlands Cancer Institute, Amsterdam, The Netherlands, [b]York Structural Biology Laboratory, Chemistry Department, The University of York, York, England, and [c]EMBL Hamburg Outstation, Hamburg, Germany

Correspondence e-mail: s.cohen@nki.nl

Automatic iterative model (re-)building, as implemented in *ARP/wARP* and its new control system *flex-wARP*, is particularly well suited to follow structure solution by molecular replacement. More than 100 molecular-replacement solutions automatically solved by the *BALBES* software were submitted to three standard protocols in *flex-wARP* and the results were compared with final models from the PDB. Standard metrics were gathered in a systematic way and enabled the drawing of statistical conclusions on the advantages of each protocol. Based on this analysis, an empirical estimator was proposed that predicts how good the final model produced by *flex-wARP* is likely to be based on the experimental data and the quality of the molecular-replacement solution. To introduce the differences between the three *flex-wARP* protocols (keeping the complete search model, converting it to atomic coordinates but ignoring atom identities or using the electron-density map calculated from the molecular-replacement solution), two examples are also discussed in detail, focusing on the evolution of the models during iterative rebuilding. This highlights the diversity of paths that the *flex-wARP* control system can employ to reach a nearly complete and accurate model while actually starting from the same initial information.

1. Introduction

With the advent of structural genomics initiatives (Stevens *et al.*, 2001) and more focused high-throughput structure-determination projects (Banci *et al.*, 2006), the need for advanced methods for structure determination has been emphasized (Lamzin & Perrakis, 2000). A specific research investment in automatic model building has led to significant developments; the advances of software such as *RESOLVE* (Terwilliger, 2004), *TEXTAL* (Ioerger & Sacchettini, 2003) and *Buccaneer* (Cowtan, 2006) are some prime representatives of this effort. These programs are typically based on interpreting electron density directly in terms of small secondary-structure elements and extending the model from such seeds. Typically, the input is a set of structure-factor amplitudes and some phase estimates or phase probability distributions.

Conversely, the *ARP/wARP* software suite (Morris *et al.*, 2004) provides a 'pipeline' which unifies model building with model refinement. The electron-density map is parameterized with a set of representative 'free' atoms. After initial model building, once a fraction of these atoms acquires a chemical identity the 'hybrid' model (consisting of free atoms and atoms with known stereochemistry) is iteratively refined and edited, taking advantage of the improved electron-density maps

research papers

produced by refinement (Perrakis *et al.*, 1999). A main advantage of the iterative editing performed by *ARP/wARP* is the ability to recycle newly derived chemical information into a better estimation of the electron density. A direct effect of this approach is that models from molecular replacement are directly incorporated into the *ARP/wARP* 'pipeline'. As *ARP/wARP* is centred on the use of an atomic 'hybrid' model, handling of models derived from molecular replacement is particularly well suited to the *ARP/wARP* formalisms.

As the number of structures deposited in the PDB increases, molecular replacement is more and more likely to be the method of choice for structure solution. Many efforts are currently under way to enable molecular replacement to work with search models that are either partial (especially in the context of macromolecular complexes) or have low identity to the structure in question.

The molecular-replacement method for crystal structure determination has developed significantly in the past two decades. *AMoRe* (Navaza, 2001) led the way to automation and addressed the need to perform all the core tasks (rotation function, translation function and rigid-body refinement) within calls to tightly unified program modules. As experience accumulated, both *AMoRe* and other software increased in sophistication and required fewer decisions by the end user. For example, *MOLREP* (Vagin & Teplyakov, 2000) included a packing-function term in its search target and an automatic search for multiple copies of the same protein in the asymmetric unit. More recently, more accurate search-target functions have been developed; the program *Phaser* (McCoy *et al.*, 2007) was recently developed to implement a molecular-replacement search using an approximation to the crystallographic maximum-likelihood target function; furthermore, *Phaser* automated complex search strategies. Meanwhile, more effort has been made in the automation and integration of the preparation of search models. For example, *CHAINSAW* (N. D. Stein, work to be published) uses an alignment between the sequence of the protein in the crystal and in the search model to edit the latter (pruning long side chains and deleting unconserved residues). In a similar spirit, the web application *elNémo* (Suhre & Sanejouand, 2004) implements normal-mode analysis to sample possible alternate conformations of the search model before starting the molecular-replacement search. More recently, there has been an effort to further integrate existing software: for example, *MrBUMP* (Keegan & Winn, 2008) and *BALBES* (Long *et al.*, 2008) integrate in different ways the identification and editing of search models with the actual molecular replacement and assessment of the quality of results.

After describing the *ARP/wARP* workflow in the context of a molecular-replacement solution, we present two specific case studies. These are used to build an understanding of how the model evolves during the *ARP/wARP* iterative process, as implemented in the *flex-wARP* control system. The core of the article deals with the systematic study of 129 cases of automatic molecular-replacement solutions for which a corresponding reference structure is available: after evaluating the relative success rate of all three *flex-wARP* standard protocols,

we use the gathered statistics to build an empirical estimator that predicts the quality of the outcome of *flex-wARP* (how good the final map and how complete the final models produced by *flex-wARP* are likely to be) based on the *a priori* known experimental data and the quality of the molecular-replacement solution.

2. *ARP/wARP* workflow for molecular replacement

The *ARP/wARP* general workflow has been described in Perrakis *et al.* (1999) and specific issues relating to molecular replacement have been presented in Perrakis *et al.* (2001). Here, we briefly discuss the general principles and our renewed experience of molecular-replacement protocols, highlighting features specific to the new control system *flex-wARP*.

2.1. General workflow

ARP/wARP typically starts by building a model made of 'free' atoms in order to represent the electron density. All further steps involve the interpretation of the coordinates of the atoms present in the current model and the current electron-density map. The steps taken by *ARP/wARP* are as follows.

Model refinement corresponds to the optimization of the parameters of the current model: refinement of atom coordinates and ADPs, modelling the solvent and scaling the structure factors using maximum-likelihood techniques as implemented in *REFMAC* (Murshudov *et al.*, 1997).

Model update consists of editing the actual model content, adding free atoms in the current likelihood-gradient map and removing atoms that lie in low density in the current likelihood-weighted map (Lamzin & Wilson, 1993).

Model (re-)building following syntactic pattern recognition using the current set of coordinates and the electron density as a starting point to assemble progressively structures of higher complexity: peptides, dipeptides, polypeptide fragments and finally with the addition of sequence information and side chains, a protein (Morris *et al.*, 2002).

In the following examples and discussion we use a new control system for the *ARP/wARP* process, *flex-wARP* (previously also introduced as *pyWARP*). *Flex-wARP* parses the output of the process modules in order to decide at run time what is the action to be taken next: the outcome of every decision is which actions have to be taken and which decision these actions lead to when completed. This concept, described briefly in Cohen *et al.* (2004), can be represented as an oriented graph, in which arrows represent actions and nodes stand for decisions (Fig. 1). As described in the following sections, this graph can be entered at different decisions/nodes when the structure is solved using molecular replacement. A fundamental difference from the 'classic' *ARP/wARP* scheme is that the number of refinement steps (internally in *REFMAC* but also between autobuilding cycles) as well as the total number of cycles are not pre-decided but are dynamically defined according to model evolution. In addition, when the

model is complete enough all free atoms are deleted and only waters are added until convergence. Moreover, a novel loop-building algorithm (Joosten *et al.*, submitted) joins main-chain fragments that are separated in sequence by 14 or fewer residues.

2.2. Starting from the positioned search model

Molecular replacement positions the search model in the correct place in the asymmetric unit. The model is then directly subjected to iterative model refinement, update and rebuilding. The main advantage of this procedure is that it keeps all chemical information contained in the search model (in terms of stereochemical restraints) and provides more information to the initial refinement stage before the first automated model building takes place.

2.3. Starting from a set of coordinates

In most cases, it is beneficial to use the restraints provided by the molecular-replacement search model for the first

Figure 1
A schematic view of the graph that is used in *flex-wARP* to automatically build a model from a molecular-replacement solution. The three different ways a molecular-replacement model can be used as input to the control system are represented in the top left corner. Round-edged boxes correspond to decisions taken at run-time. Arrows connecting the decision nodes correspond to one or a set of actions to be taken on the current model; similarly coloured arrows correspond to the same action set.

refinement steps. However, if most of the restraints are genuinely invalid (*e.g.* when a very poor starting model is used) it may be beneficial to remove the chemical identity of atoms by simply keeping the atomic coordinates but turning all atom types to free atoms, 'DUM', effectively eliminating all stereochemical information. This free-atom model is refined and updated without restraints until the first automated model-building step that will assign new chemical identity.

2.4. Starting from an electron-density map

The positioned search model can also only be used to compute an initial electron-density map, which is then interpreted in terms of a free-atom model (as for any other experimental map). This model is improved by a few cycles iterating refinement and model update. A first model is then built and the process iterates all three actions. It is clear that this procedure is not using any of the chemical information that is contained in the molecular-replacement model; hence, it will rapidly remove most of the bias introduced by the molecular-replacement model.

In the next part we will use two specific examples to illustrate how *ARP/wARP* iteratively edits and optimizes the model output from molecular replacement, leading in these cases to highly complete and accurate structures. We will then report the medium-scale systematic study that was required to draw statistically sound conclusions about the usefulness of the application of *ARP/wARP* procedures to molecular-replacement solutions.

3. Two in-depth examples

3.1. Completing a partial search model

The first example illustrates how *ARP/wARP* can complete a partial molecular-replacement model. It concerns the solution of the structure of a noncovalent complex between Ubc9 (an E2 ubiquitin ligase) and SUMO (a ubiquitin-like regulator of Ubc9 in the context of this complex; Knipscheer *et al.*, 2007). A data set at 1.8 Å was used for this study. Molecular replacement was performed using only Ubc9 as the search model, representing only two-thirds of the ordered asymmetric unit content. The model was directly input to *flex-wARP* using the default protocol; that is, keeping both atom coordinates and chemical information during the first update cycles (as described in §2.2). The evolution of the quality and completeness of the automatically built model is summarized in the top two panels of Fig. 2.

Within 40 steps of refinement and model update, the number of atoms reaches the expected number for the expected asymmetric unit content. This in turn enables the model-building algorithms to build both Ubc9 (which was already provided by the molecular-replacement solution) and SUMO. After step 70, when the *flex-wARP* decision system considers the model to be of good quality, final model clean-up and re-arrangement takes place: the number of atoms decreases sharply (since all free atoms are removed) and over
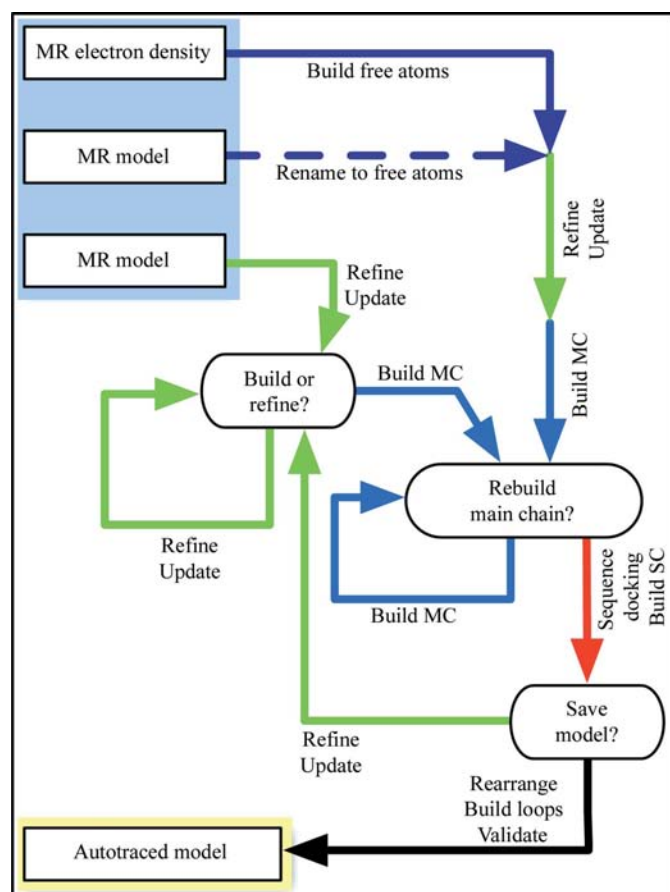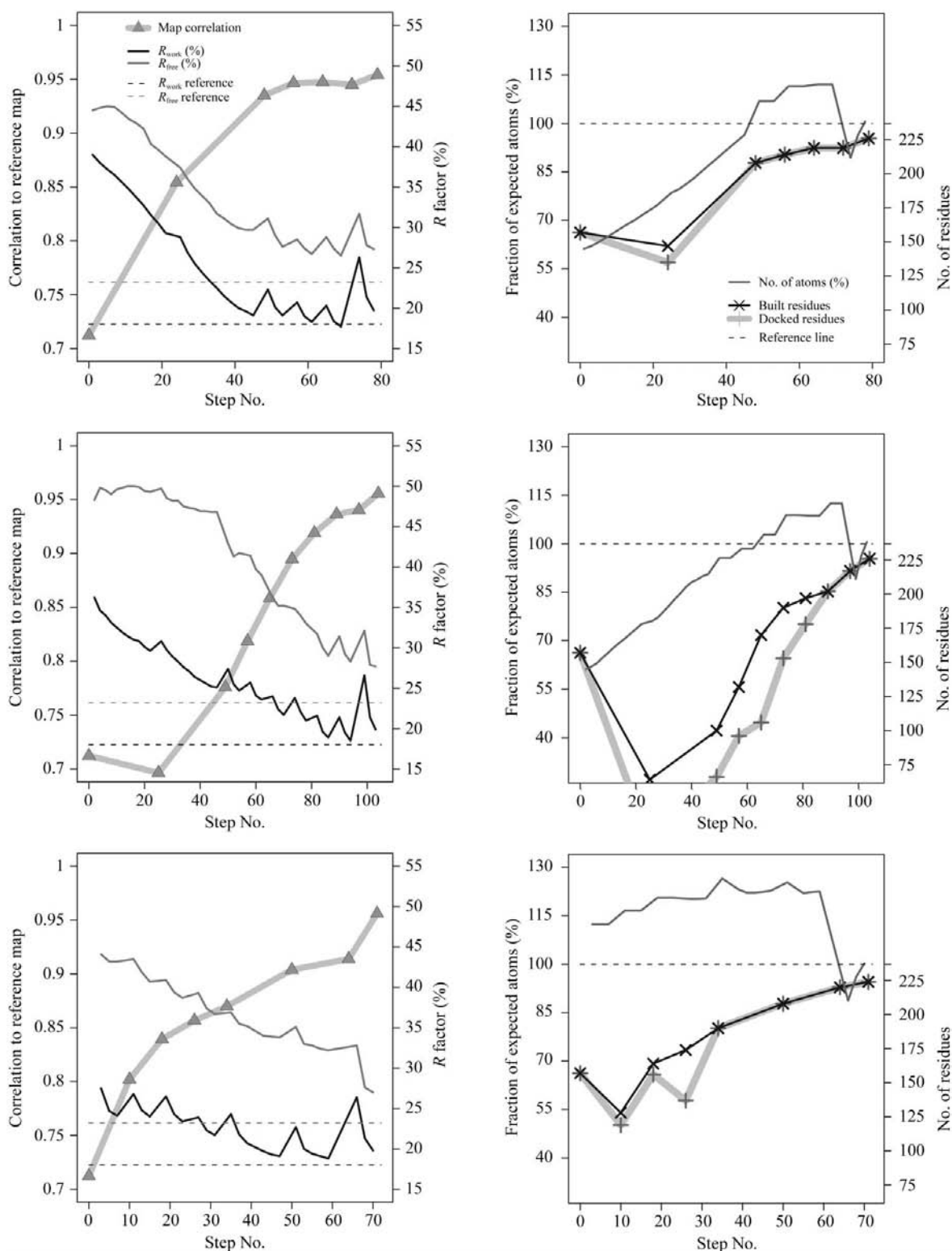
Figure 2
The evolution of the quality and completeness indicators as *flex-wARP* iteratively completes, edits and refines a model for the Ubc9–SUMO complex. The left panels represent the evolution of R_{work} and R_{free} (plain thin lines in black and grey, respectively), with corresponding values from the reference shown as dashed horizontal lines. The thick grey line corresponds to the correlation of the current likelihood-weighted map to the reference electron-density map. The right panels show the completeness of the model over iteration: a thin grey line presents the number of atoms in the current model compared with the number in the reference structure. The thin black and thick grey lines correspond to the number of built residues and the number of residues assigned to the sequence (hence having their side chain built), respectively. In all panels, marks (triangles and cross) represents the steps where main-chain tracing, sequence docking and side-chain building are performed. The top two panels correspond to the default protocol starting from the molecular-replacement model (§2.2). The middle panels correspond to the protocol that starts from the coordinates alone (removing all the restraints, as described in §2.3). Finally, the bottom two panels represent the results obtained by the protocol starting from the electron density alone (described in §2.4).

the next ten steps water molecules are added to model the ordered part of the solvent. At the same time the loop-building algorithm joins a few chains and builds extra residues with their corresponding side chains.

We also tested the two other protocols: starting from the coordinates only or the electron-density map (as described in §§2.3 and 2.4, respectively; middle and bottom panels of Fig. 2). Both tests led to fairly complete models of quality similar to that built using the standard protocol. However, this required 110 and 73 steps, respectively, compared with 81 steps for the default protocol. The faster convergence of the protocol in which all expected free atoms are placed in electron density first is expected since the molecular-replacement solution was a partial model. Finally, since in these two cases restraints take a few cycles to accumulate through model building, the initial steps of refinement show strong overfitting (maintaining up to 22% and 19% difference between R_{work} and R_{free}, respectively), whereas the largest difference between R_{work} and R_{free}, when starting from the search model and maintaining stereochemistry, is 10% (with a mean difference of 8%).

3.2. Recovering from a poor phase set

The second example shows how *ARP/wARP* can help to recover a very poor molecular-replacement solution. It is based on the high-resolution structure determination of the SMR domain of the MutS2 protein from *Helicobacter pylori* (Lebbink, Radicella & Sixma, work to be published). Diffraction data were collected from a crystal diffracting to a resolution of 1.0 Å and containing two copies of the protein in the asymmetric unit. The structure was solved by molecular replacement using an NMR model with 19% sequence identity as search model. Despite the high resolution and a fairly clear molecular-replacement solution provided by *Phaser*, the positioned model did not refine and the resulting electron density was hard to interpret (as reflected by a map correlation of only 32% with the reference model).

The evolution of the iteratively rebuilt model using the standard protocol (as described in §2.2) is shown in the top panels of Fig. 3. As the model is edited and completed (many atoms were missing after the editing of the search model), both R_{work} and R_{free} rapidly decrease. This corresponds to a sharp increase of the map correlation to the reference. After step 25, only minor editing takes place and the final model (with 160 out of 168 residues) is obtained at step 60.

We again tested the other two protocols we describe on this data set. Starting from coordinates alone (removing the restraints, as described in §2.3; middle panels in Fig. 3) works well but takes more time (70 steps instead of 60), which is somewhat surprising; restraints should not matter much when 1.0 Å data are available. Also surprisingly, starting from the map generated by the molecular replacement (protocol described in §2.4; bottom panels in Fig. 3) proved to require a very large number of steps to be able to produce useful results: the first side chains are built at step 297 and *flex-wARP* considers the model to be complete only after step 391. This could be explained by the poor quality of the initial map resulting in wrong positioning of the seeding atoms; somehow the atomic information of the search model contributes very significantly despite the very high resolution of the data. Because *ARP/wARP* peptide recognition is optimized for a resolution in the range of 1.6–2.5 Å, we also applied the protocol described in §2.4 after cutting the data to a resolution of 1.4 and 1.6 Å. This did not lead to any improvement in convergence speed or model quality.

4. Systematic study of example cases

The previous two examples were used to show how the model evolves during automatic iterative model rebuilding, especially in the context of the new *ARP/wARP* control system *flex-wARP*. To reach statistically sound conclusions concerning the benefits of using each of the three proposed protocols (§2) requires a large number of test cases. Here, we present the results of a medium-scale systematic study based on 129 deposited structures.

4.1. Presentation of test results

A large number of statistical indicators are available to evaluate the quality of the produced model when the final result is not known: R_{work}, the figure of merit, the likelihood gradient (during refinement), the number of built atoms *etc.* Since all these indicators are strongly correlated to each other, we chose to show only the fraction of built residues (the number of residues in the final model divided by the expected number of residues given the sequence information and the number of copies in the asymmetric unit) and the fraction of docked residues (the number of residues which were assigned to the sequence and for which the side was built divided by the expected number of residues).

Additional to the above indicators, when a reference model for each data set is available for test reasons, additional metrics can be employed correlating the result of the automated process to the reference set. From these indicators we chose to show the correlation between the reference electron-density map and that obtained from the final model.

For the sake of clarity, for graphical representation we pool the data sets according to the 'initial *R* factor' and to the 'resolution' of the diffraction data, instead of crudely plotting results for each of the 129 data sets. The 'initial *R* factor' is the R_{work} produced by the positioned search model prior to any positional/ADP refinement.

Many other indicators are available and it is also valid to group the data sets in different ways for presentation; the primary data for all tests are available as a set of tables in ASCII text files from http://xtal.nki.nl/~serge/BALBES-1 for the benefit of the curious reader.

4.2. Conclusions from test cases

The default protocol (§2.2; Fig. 4) shows that when the initial R_{work} is better than 30% automatic model building is likely to produce useful results. Conversely, a molecular-replacement solution producing an R_{work} of between 30 and

40% is almost equally likely to be rescued by the default *flex-wARP* procedure or fail to produce results of any use; however, there is a tendency to improve the map quality (as shown by the values of map correlation) but produce fairly incomplete models. When success is assessed as a function of resolution, the fundamental tendencies of *ARP/wARP* show up: when data better than 2.0 Å are available, *ARP/wARP* fails only occasionally (presumably when the starting model is
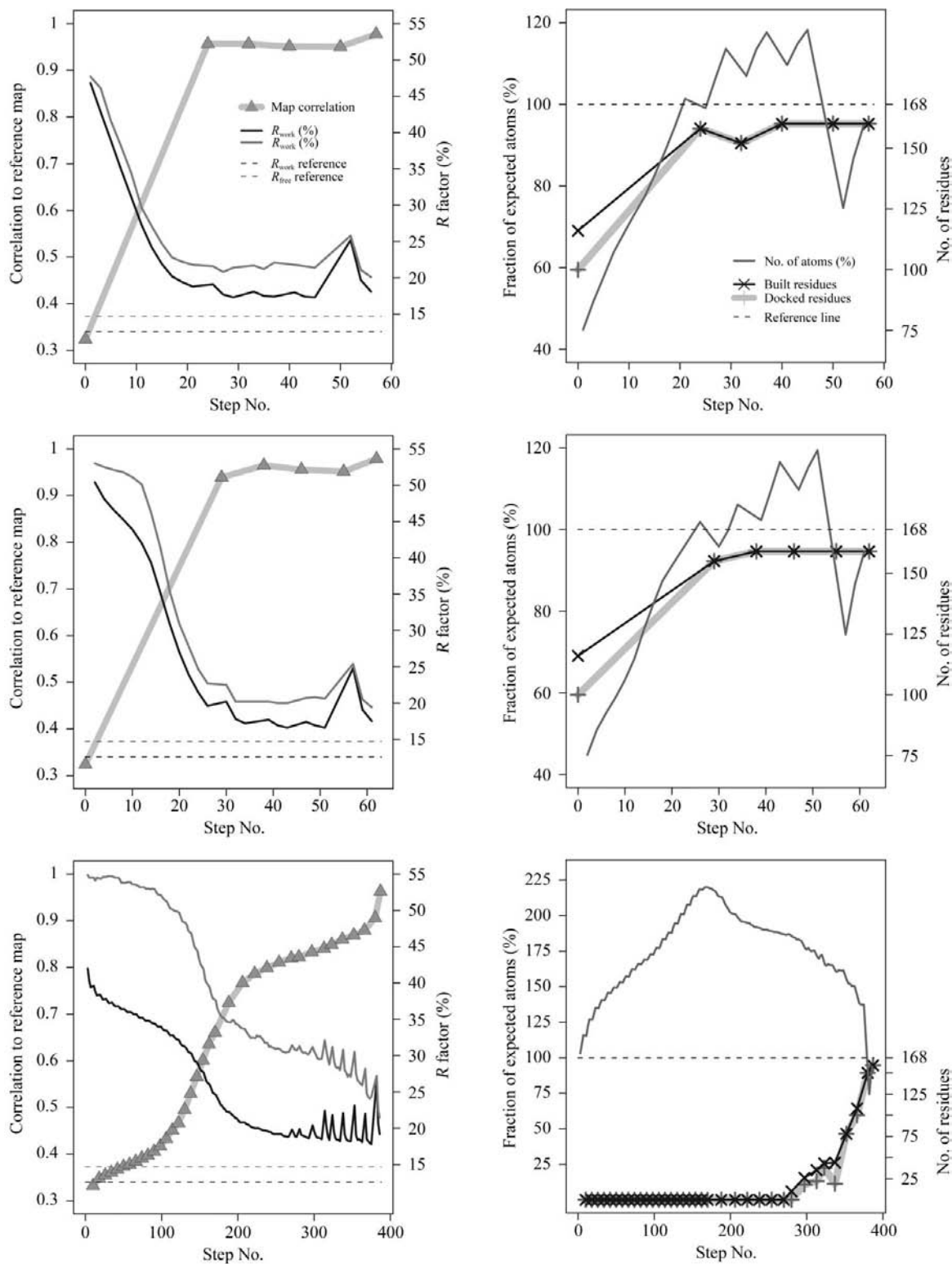

Figure 3
The evolution of the quality and completeness indicators as *flex-wARP* iteratively edits and refines a model for the SMR. The legend is the same as for Fig. 2. Note that the bottom right figure uses a different *y* scale than the top two figures: the number of generated atoms increases to more than 220% of the expected number of atoms.

really very bad). Between 2.0 and 2.5 Å models are in general less complete and more cases tend not to work, but in general the runs are successful. With data weaker than 2.5 Å there are occasional successes that produce models close to 80% completeness, while below 3.0 Å we did not observe a single successful case. These observations are well correlated with the general *ARP/wARP* success rates, but also show that *ARP/wARP* can often produce good model-building results even from data that do not extend beyond 2.5 Å.

The two alternative protocols (§§2.3 and 2.4; Fig. 5) usually produce poorer results than the default. Nevertheless, there are a few exceptions where these alternate possibilities were more successful. Starting from atomic coordinates (§2.3) occasionally works better, but there is no clear tendency. However, it is notable that one case that shows 40% more docked residues built than the default protocol is at rather low resolution (2.5 Å) and has a relatively high starting R_{work} (35%); we performed a detailed visual inspection of the initial
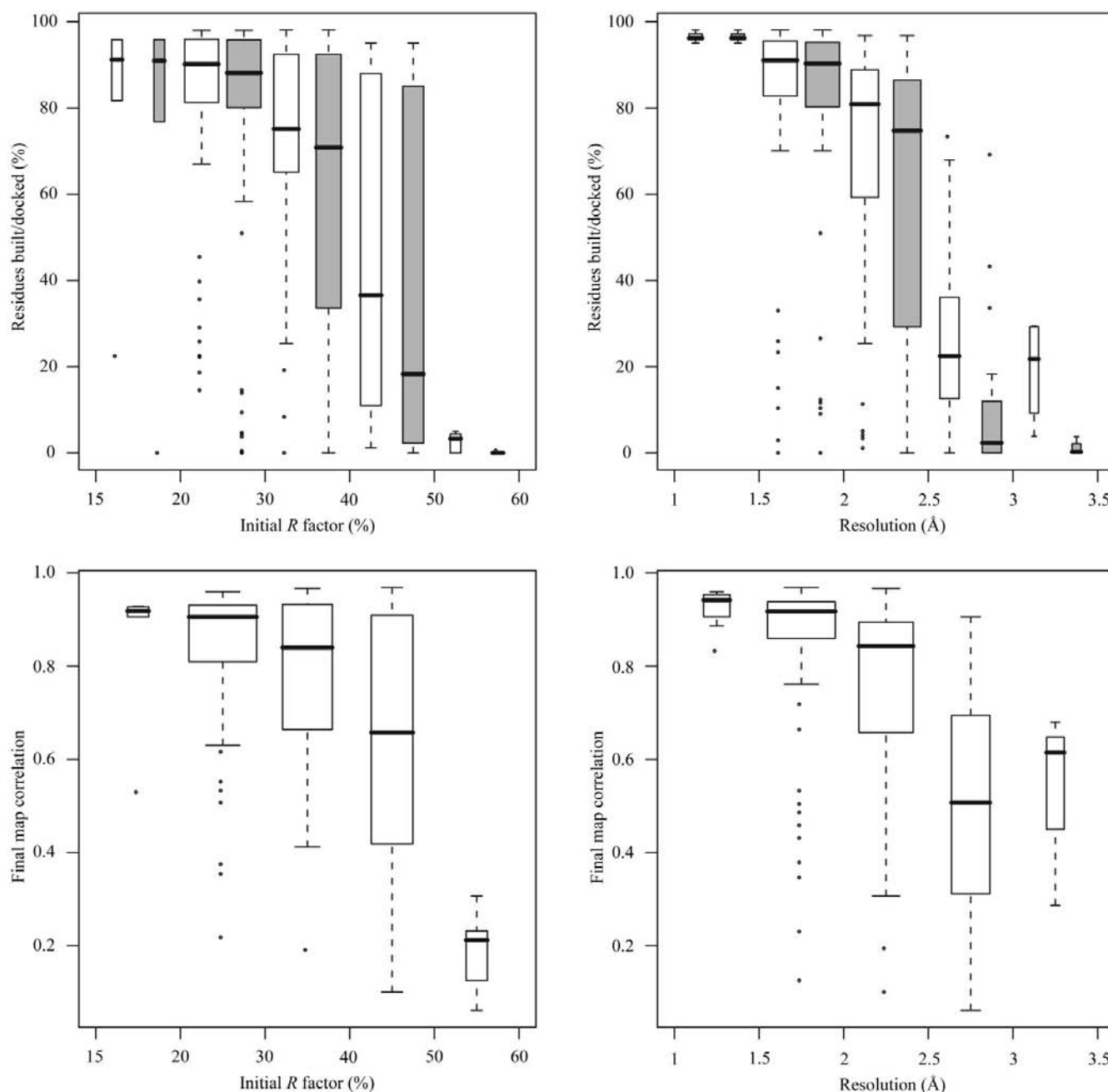
Figure 4

Box plot of the results of *flex-wARP* (running in default mode, keeping the initial model). The data sets were divided into five groups based either on the initial R factor (left column) or its high-resolution limit (right column). The boundaries of each group are labelled on the x axis. In each category the relative width of the box corresponds to the number of data sets in the category; the box itself spans vertically from the first to the third quartiles, whilst the bold line is situated at the median; whiskers represent the full spread of the distribution, whilst open circles represent outliers. The top two graphs represent the fraction of residues built (white boxes) and the fraction of residues assigned to sequence (hence having side chain built; grey boxes). The bottom two graphs give the values of the correlation of the map obtained by *flex-wARP* with the reference map.

(molecular-replacement) model and the reference model but were unable to derive a straightforward explanation for the behaviour of this particular data set. The third protocol (§2.4) does not show any advantage over the other two in general. However, it yields significantly better results when the starting model is bad (as indicated by R_{work}) and the resolution is poorer than 1.5 Å; the difference can be as much as 40% more residues, with 20% more residues being quite often the case. These results could be further enhanced if advantage of NCS averaging or density modification was taken before such runs, but unfortunately the benefits of such a 'pre-treatment' of the electron density before starting the model rebuilding could not be systematically tested here.

4.3. Learning from the experience with the test cases

Having a reference structure, an objective assessment of the quality of the model built by *flex-wARP* can be derived. However, in normal day-to-day use the program is obviously run without knowing the result and, having the benefit of
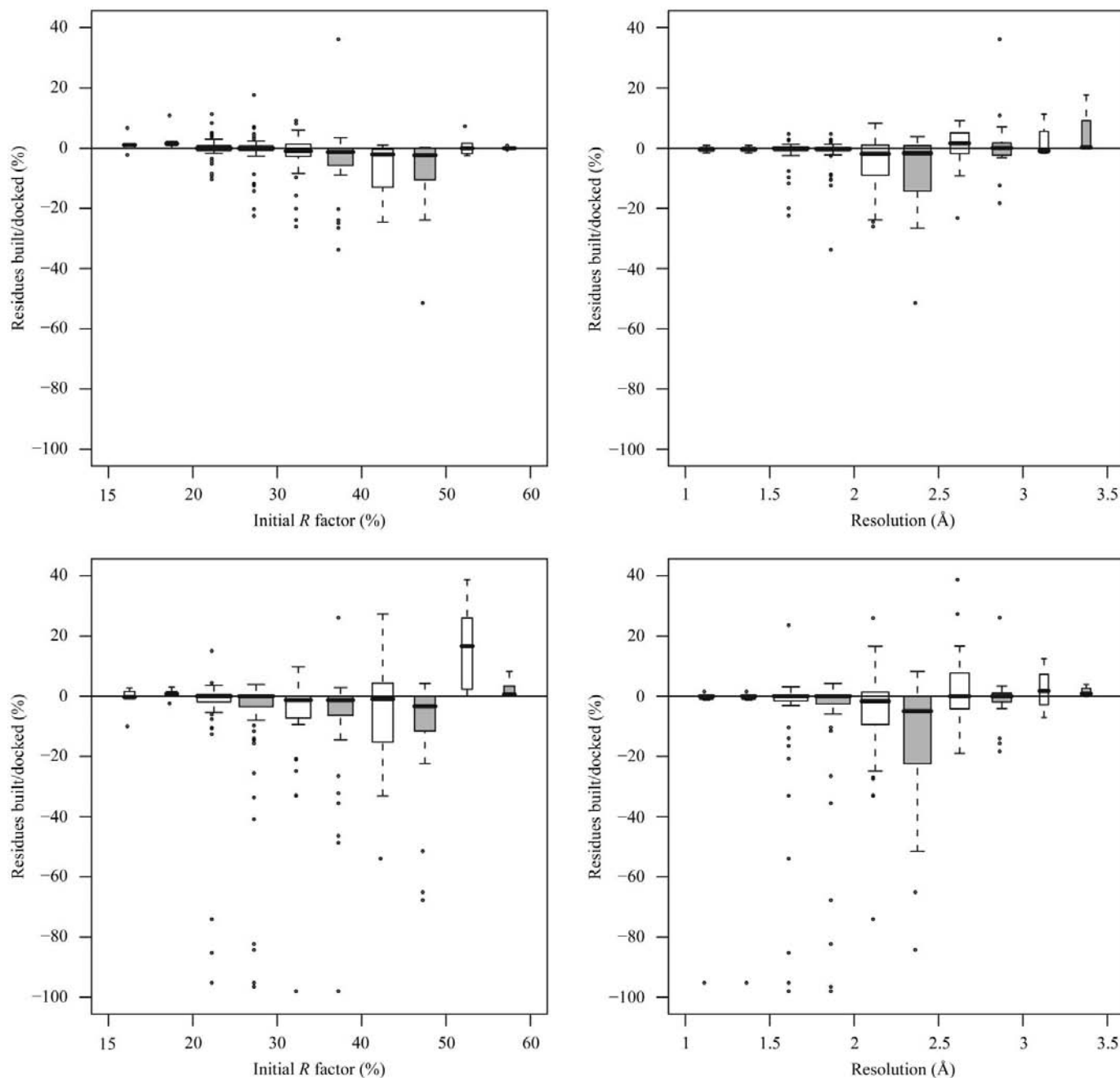

Figure 5
Box plot of the difference between the results obtained with the default protocol and those obtained using only starting-model coordinates (top two figures) and those using the starting model only to compute an electron-density map (bottom figures). The grouping is the same as that used in Fig. 4 (using the *R* factor of the molecular-replacement solution on the left and considering the high-resolution limit on the right). Here, we represent the difference in the fraction of residues built (white boxes) and assigned to sequence (grey boxes).

hindsight, knowing the reference structure. In other words, assessing the quality of the electron-density map produced by *flex-wARP* is not as trivial as computing a correlation with the 'final' reference map. It would be useful to know whether it is possible to use parameters available immediately after successful molecular-replacement solution to predict the quality of the map that *flex-wARP* will produce and ultimately the success of the *ARP/wARP* procedure. Using the relatively large number of test cases presented in this study, we tried to build an estimator of the produced map quality; the final map quality correlates well with the percentage of automatically built residues.

Of the large number of parameters which are available after the molecular-replacement solution, most were found to be of little predictive value or were redundant with the final parameters we chose: the initial R factor (R_{MOLREP}), the high-resolution limit of the data ($Resol_{high}$) and the solvent content (SC). We developed a good-quality estimator using these parameters and an intercept (constant term) with the following formula,

$$MC_{estimate} \simeq 0.86 \times SC + 19.61 \times R_{MOLREP} - \frac{0.52}{Resol_{high}}$$
$$+ \frac{0.44}{R_{MOLREP}} - 0.11 \times Resol_{high}^2 - 23.64 \times R_{MOLREP}^2$$
$$- 4.1. \tag{1}$$

As explained in §5, this estimator is derived solely as an empirical value computed from the statistics obtained by this study and is not based on a particular physical model.
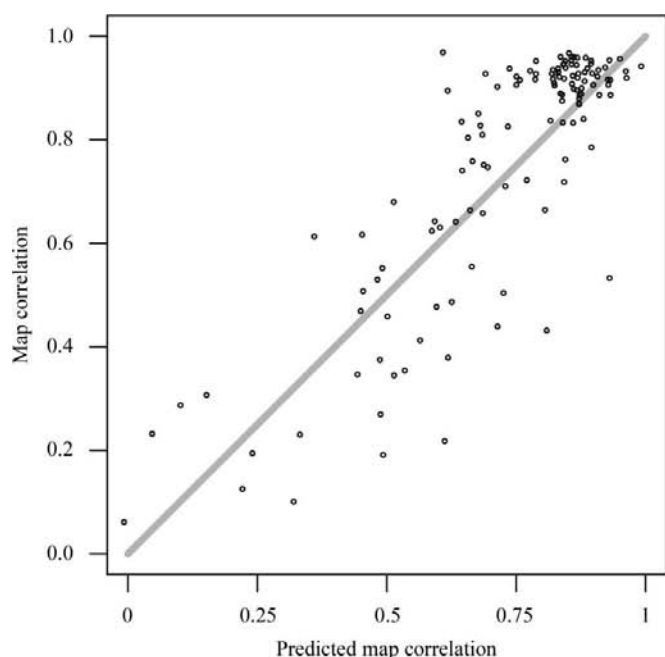
The quality of the estimator is shown in Fig. 6, which shows a scatter plot of the estimated value at the start of the run against the final 'true' value computed against the correct reference. The proposed formula is clearly a crude estimator of the final results of *flex-wARP*, but nonetheless it can prove useful to quickly estimate the quality of an *ARP/wARP* run starting from a molecular-replacement solution. Upon detailed observation of the current mathematical model one sees that its main defect is the inability to express saturation; the target value (map correlation to the reference model) is bounded and the bounds are reached even for non-extreme cases. Hence, the response function is likely to be better modelled by some sort of sigmoid function instead of a linear one; unfortunately, this type of model requires more training data than are currently available.

5. Material and methods

5.1. Ubc9–SUMO complex

The data used in this study were collected to a resolution of 1.80 Å. The structure was solved by molecular replacement using the program *Phaser*; the search model consisted of the structure of Ubc9 alone (PDB code 2grn, 158 residues), which corresponded to only two-thirds of the expected ordered content of the asymmetric unit. Hence, one-third of the structure, the SUMO protein consisting of 79 residues, was missing from the model and still had to be built. The model resulting from molecular replacement produced an R_{work} of 46% and was directly input to *flex-wARP*. The models obtained after each side-chain-building step were compared
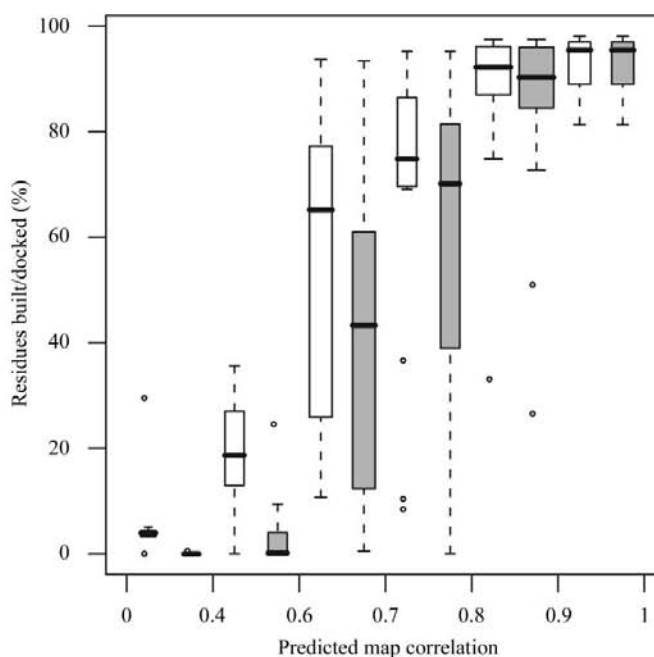

Figure 6
The quality of the final map-quality predictor. On the left hand side, a scatter plot is shown of the predicted value *versus* the true value of the map correlation at the end of the *flex-wARP* run with the reference map. On the right-hand side, a box plot shows the fraction of the residues built (white boxes) and the fraction assigned to sequence (grey boxes). In this box plot, the data sets are grouped by predicted final map quality. Note that the groups have irregular spacing in order to have approximately the same number of data sets per group.

with the reference structure in order to be able to assess the evolution of the map and model quality.

As a reference structure we used a structure of the same complex solved by molecular replacement and refined against 1.4 Å data in a slightly different crystal form (PDB code 1u9a). The difference in crystal unit cells arises from a slight re-orientation of the two proteins in the better diffracting crystal, such that we had to perform molecular replacement (in two steps, one protein at a time) and extensive re-refinement [*MOLREP* (Vagin & Teplyakov, 2000) and *REFMAC*5 (Murshudov *et al.*, 1997) were used] in the crystal form studied here. This protocol was completed by interactive model rebuilding in *Coot* (Emsley & Cowtan, 2004). The overall quality of the reference model is presented in Table 1.

5.2. SMR domain of MutS2

Diffraction data were measured to a resolution of 1.00 Å, practically limited by the beam wavelength and the minimum crystal-to-detector distance (as displayed in Table 1, despite the high resolution of the data the R_{sym} is still fairly low in the outer resolution shell). The structure was solved by molecular replacement using the program *Phaser*; the search model (edited from PDB entry 2d9i) is an NMR structure consisting of 20 conformers and has a low identity (19%) to the crystallized protein. All 20 conformers of the search model were edited and loops with a large r.m.s.d. between conformers were deleted. The *CHAINSAW* program was used to edit nonconserved residue side chains, removing all atoms after C^{γ}. *Phaser* produced a reasonable molecular-replacement solution for both copies (with a log-likelihood gain of 51.4).

The quality of the positioned search model was assessed using rigid-body refinement (one rigid body per protein copy), leading to an R_{work} of 53%. Extensive positional and ADP refinement were not able to reduce the R_{work} below 46%. Both these refinements were performed using *REFMAC*5.

The reference model for comparisons was obtained by submitting the molecular-replacement model to the classic *ARP/wARP* package (v.6.1.1, using the currently distributed control system) for model rebuilding. This was complemented by iterative refinement in *REFMAC*5 (Murshudov *et al.*, 1997) and interactive rebuilding in *Coot* (Emsley & Cowtan, 2004). Anisotropic atomic displacement and multiple conformations were added at the final steps of refinement. The reference model (Table 1) needs improvement before it can be considered 'final'; however, it is good enough to be used as a reference in assessment of the results of automatic model rebuilding.

The quality of the search model was assessed *a posteriori* by computing the r.m.s. deviation between the search model and the reference structure. A least-squares superposition using C^{α} atoms resulted in an r.m.s.d. of 1.14 Å for 58 out of 82 total residues that were conserved in the alignment.

5.3. *BALBES* test sets

Automatic molecular replacement by *BALBES* was systematically attempted on all structures released by the PDB

Table 1
Quality of the data and related reference structures used for the two detailed examples.

The model for SMR still needs to be edited and refined, but is of sufficient quality to assess the results of automatic model rebuilding. Values in parentheses are for the highest resolution shell.

	Ubc9–SUMO	SMR
Data collection		
Space group	$P2_1$	$C2$
Unit-cell parameters (Å, °)	$a = 50.4$, $b = 34.6$, $c = 68.7$, $\alpha = 90$, $\beta = 90.6$, $\gamma = 90$	$a = 32.8$, $b = 47.9$, $c = 93.7$, $\alpha = 90$, $\beta = 91.9$, $\gamma = 90$
Resolution (Å)	50–1.80 (1.90–1.80)	94–1.00 (1.06–1.00)
R_{sym} (%)	3.7 (10.1)	5.6 (10.2)
$I/\sigma(I)$	11.4 (6.9)	13.2 (5.4)
Completeness (%)	95.2 (72.8)	96.1 (95.3)
Redundancy	4.0 (3.2)	2.1 (1.9)
Refinement		
Resolution (Å)	50–1.80	94–1.00
No. of reflections	20183 (1081)	70653 (3726)
R_{work}/R_{free} (%)	18.0/23.2	12.6/14.7
No. of residues (per ASU)	237	168

between 22 September and 9 October 2006; the search models were structures released before 21 September 2006. When *BALBES* produced a molecular-replacement solution according to internal criteria and this model could be related to the deposited structure in the PDB (taking into account potential origin shifts), this example case was added to the set of test structures. A few structures for which we did not manage to reproduce the R factor of the deposited model were excluded. Finally, because *flex-wARP* currently only handles proteins, we removed data sets for which more than 10% of the ordered diffracting matter was not made of proteins. These two filtering criteria removed a total of 33 data sets out of 162.

The test set thus contains structures that could be solved using a non-identical search model that existed in the PDB prior to the deposition of the test case. For each such case, we store the diffraction data deposited by the authors, the final model deposited by the authors to be used as a reference, the sequence of the protein(s) extracted from the final model and the search model positioned at the right place of the asymmetric unit by *BALBES*.

The final set comprises 129 structures and is fairly representative of the content of the PDB, with the resolution of the diffraction data spanning from 1.05 to 3.1 Å (median at 2.1 Å), a solvent content ranging between 32 and 72% (median at 50%) and Wilson B factors in the range 4.1–76.8 Å^2 (median at 19.3 Å^2).

5.4. Systematic test protocol

For each data set the deposited model was moved to the same origin as the output of the molecular replacement and a likelihood-weighted electron-density map was computed in *REFMAC*5 without any positional or ADP refinement (only scaling was applied). The model structure and the electron-density map were then used to assess the quality of the models produced by *flex-wARP*.

All three protocols (as described in §§2.2, 2.3 and 2.4) were tested systematically. Iteration in *flex-wARP* was stopped when more than 95% of the residues were built (including side chains) or once 50 steps of model building had been completed, regardless of the total execution time or the number of refinement/update steps. The default protocol, in which the positioned search model is iteratively edited following refinement without removing restraints, was used as a baseline when evaluating the results of the other two protocols (Fig. 5).

5.5. Selecting an estimator of *flex-wARP* outcome

For each data set from the *BALBES* test set, the following information was gathered before starting the model rebuilding, reflecting the statistics available to the crystallographer just after the end of the molecular-replacement search.

(i) The initial R factor (the R factor produced by the molecular-replacement placed search model).

(ii) The high-resolution limit of the data set.

(iii) The Wilson B factor of the data set.

(iv) The solvent content of the crystal.

(v) The number of residues in the asymmetric unit.

(vi) The completeness of the molecular-replacement model (the number of residues of the search model compared with the number of expected residues in the asymmetric unit).

(vii) The identity of the search model to the content of the crystal (the fraction of the residues of the search model that are identical to residues of the asymmetric unit content).

These last two parameters need further attention. Despite their straightforward definition, calculation can lead to erroneous values of larger than 1.0 in the particular case of multiple copies of a protein in the asymmetric unit and when the search model contains a higher number of copies compared with what is expected. For these data sets we have no access to the expectations of the person who solved the structure, so we had to use the number of copies present in the reference structure as the expected number of copies. Though this does not precisely reflect reality, out of the 129 data sets used in the study only three had search-model completeness above 1.0 and none of the 129 data sets had a relative identity above 1.0.

To train the estimator, we used as a target function the correlation of the final *flex-wARP* map to the reference map. For each tested model, an analysis of variance (with a χ^2 test) was used to assess which parameters were relevant and which were of little predictive value (comparing the tested model with all the models obtained by systematically dropping one term). Different models were then compared using the Akaike Information Criterium that is designed to reward goodness of fit but also includes a penalty that is an increasing function of the number of degrees of freedom (the goal being to avoid overfitting). All models tested were linear models where the input parameters were either one of the metrics listed above, some power of these parameters or a product of different parameters. The modelling was performed using the R envir-

onment (R Development Core Team) and the *MASS* package (Venables & Ripley, 2002).

6. Conclusion

ARP/wARP and the new *flex-wARP* control system are well suited for rebuilding and completing models obtained by molecular replacement. Whilst limited conclusions can be drawn from the two specific examples we present, our medium-scale study based on data of a broader range of quality and resolution provides knowledge that is both more reliable and can be applied to new data sets.

The primary result of this sampling experiment is that *ARP/wARP* is fairly resilient to poor molecular-replacement solution at high and medium resolution (extending to around 2.5 Å), whilst it can also be useful at lower resolution provided that the molecular-replacement solution is close enough to the true structure. Going further in the analysis, we showed that the protocol that uses the model produced by molecular replacement, including the attached chemical restraints, is the most successful one; however, it is still advisable to blindly test the three proposed protocols to be sure to get the most out of *ARP/wARP*. As also illustrated in Fig. 7, in almost one out of five cases (19%) it is worth trying the nondefault protocols. To facilitate systematic tests of all available protocols, we are planning to provide a web service to the community.

Finally, we were able to use a simple linear model approximation to express the relative importance of the resolution of the data and the quality of the molecular-replacement solution in obtaining a complete model and a good-quality map using *flex-wARP*. Overcoming the limitations of the proposed estimator might be achieved by incorporating new training data set as the *BALBES* development team provides the result of more test rounds and moving away from a linear model to a sigmoid response model or some other supervised learning technique. Despite the scatter visible in Fig. 6, this estimator is
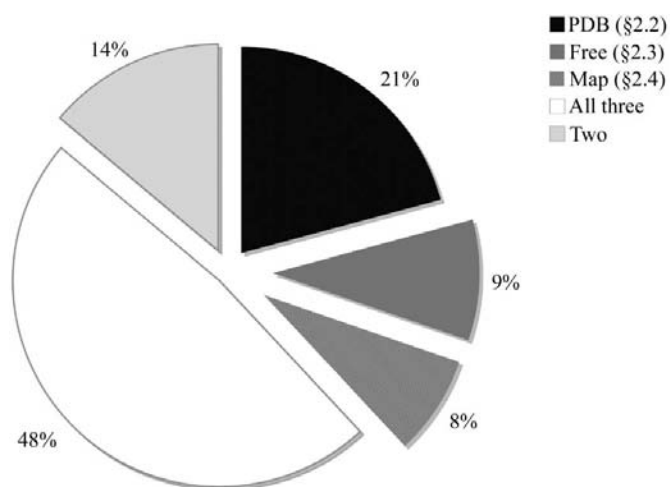

Figure 7
The percentage of case studies for which any of the three protocols described in §2 are significantly better (more than 5% discrepancy) than any other protocol and of cases in which the choice of protocol makes no difference or at least two protocols perform similarly.

useful when 'extreme' values are obtained: obtaining an estimate value under 0.6 is a good reason to spend more time obtaining better phasing information (through a better search model or combining molecular replacement with experimental phasing), while obtaining a value greater than 0.8 should almost guarantee success in automated model building.

References

Banci, L. *et al.* (2006). *Acta Cryst.* D**62**, 1208–1217.

Cohen, S. X., Morris, R. J., Fernandez, F. J., Ben Jelloul, M., Kakaris, M., Parthasarathy, V., Lamzin, V. S., Kleywegt, G. J. & Perrakis, A. (2004). *Acta Cryst.* D**60**, 2222–2229.

Cowtan, K. (2006). *Acta Cryst.* D**62**, 1002–1011.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.

Ioerger, T. R. & Sacchettini, J. C. (2003). *Methods Enzymol.* **374**, 244–270.

Keegan, R. M. & Winn, M. D. (2008). *Acta Cryst.* D**64**, 119–124.

Knipscheer, P., van Dijk, W. J., Olsen, J. V., Mann, M. & Sixma, T. K. (2007). *EMBO J.* **26**, 2797–2807.

Lamzin, V. S. & Perrakis, A. (2000). *Nature Struct. Biol.* **7**, *Suppl.*, 978–981.

Lamzin, V. S. & Wilson, K. S. (1993). *Acta Cryst.* D**49**, 129–147.

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.

Morris, R. J., Perrakis, A. & Lamzin, V. S. (2002). *Acta Cryst.* D**58**, 968–975.

Morris, R. J., Zwart, P. H., Cohen, S., Fernandez, F. J., Kakaris, M., Kirillova, O., Vonrhein, C., Perrakis, A. & Lamzin, V. S. (2004). *J. Synchrotron Rad.* **11**, 56–59.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.

Navaza, J. (2001). *Acta Cryst.* D**57**, 1367–1372.

Perrakis, A., Harkiolaki, M., Wilson, K. S. & Lamzin, V. S. (2001). *Acta Cryst.* D**57**, 1445–1450.

Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.

Stevens, R. C., Yokoyama, S. & Wilson, I. A. (2001). *Science*, **294**, 89–92.

Suhre, K. & Sanejouand, Y. H. (2004). *Nucleic Acids Res.* **32**, W610–W614.

Terwilliger, T. (2004). *J. Synchrotron Rad.* **11**, 49–52.

Vagin, A. & Teplyakov, A. (2000). *Acta Cryst.* D**56**, 1622–1624.

Venables, W. & Ripley, B. (2002). *Modern Applied Statistics with S*, 4th ed. New York: Springer.