Structural bioinformatics

An evaluation of automated homology modelling methods at low target–template sequence similarity

James A. R. Dalton and Richard M. Jackson*

Institute of Molecular and Cellular Biology, Faculty of Biological Sciences, University of Leeds, Leeds LS2 9JT, UK

Received on February 13, 2007; revised on April 25, 2007; accepted on May 9, 2007

Advance Access publication May 17, 2007

ABSTRACT

Motivation: There are two main areas of difficulty in homology modelling that are particularly important when sequence identity between target and template falls below 50%: sequence alignment and loop building. These problems become magnified with automatic modelling processes, as there is no human input to correct mistakes. As such we have benchmarked several stand-alone strategies that could be implemented in a workflow for automated high-throughput homology modelling. These include three new sequence-structure alignment programs: 3D-Coffee, Staccato and SAAlign, plus five homology modelling programs and their respective loop building methods: Builder, Nest, Modeller, SegMod/ENCAD and Swiss-Model. The SABmark database provided 123 targets with at least five templates from the same SCOP family and sequence identities ≤50%.

Results: When using Modeller as the common modelling program, 3D-Coffee outperforms Staccato and SAAlign using both multiple templates and the best single template, and across the sequence identity range 20–50%. The mean model RMSD generated from 3D-Coffee using multiple templates is 15 and 28% (or using single templates, 3 and 13%) better than those generated by Staccato and SAAlign, respectively. 3D-Coffee gives equivalent modelling accuracy from multiple and single templates, but Staccato and SAAlign are more successful with single templates, their quality deteriorating as additional lower sequence identity templates are added. Evaluating the different homology modelling programs, on average Modeller performs marginally better in overall modelling than the others tested. However, on average Nest produces the best loops with an 8% improvement by mean RMSD compared to the loops generated by Builder.

Contact: r.m.jackson@leeds.ac.uk.

Supplementary information: Supplementary data are available at Bioinformatics online.

1 INTRODUCTION

Due to the success of genomic sequencing and the time-consuming nature of experimental protein structure determination, a large gap has developed between the number of known protein sequences and the number of known protein structures (Hillisch et al., 2004). With structural genomics unable to keep up with the number of newly discovered genes, many of which may be candidates for rational drug design, computational structure prediction methods are needed to bridge the gap—the most accurate of which is homology (or comparative) modelling (Martin et al., 1997; Moult, 2005). Homology modelling requires a template of known 3D structure that is structurally similar to the unknown. The initial step, template identification, has been greatly strengthened by hidden-Markov-model, position-specific scoring matrix, and profile techniques (Moult, 2005). However, the subsequent step, sequence alignment of template(s) to target, continues to be a difficult area particularly when sequence similarity is low (Kopp and Schwede, 2004). Indeed model accuracy often correlates with the level of sequence identity between target and template (Kopp and Schwede, 2004), which in turn infers the extent of structural similarity (Chothia and Lesk, 1986).

Below 50% identity, modelling confidence deteriorates and below 25% identity, confidence in accuracy is weak (Kopp and Schwede, 2004). The final step, model construction, also has its difficulties, particularly where gaps are present in the sequence alignment. These gaps correspond to loop regions in the model, which have to be built without template guidance, and the larger the gaps, the greater the difficulty in identifying the native conformation (Fiser et al., 2000).

In order to address the sequence alignment problem, new methods are being developed that use 3D structural information to improve overall alignment accuracy. Three of these are 3D-Coffee (O’Sullivan et al., 2004), Staccato (Shatsky et al., 2006) and SAAlign (Madhusudhan et al., 2006).

3D-Coffee implements the Fugue (Shi et al., 2001) threading program for sequence–structure alignment, and the structure alignment program (SAP) (Taylor and Orengo, 1989) for structure–structure alignment. These external pair-wise alignments are combined with the T-Coffee library to generate a multiple sequence alignment. The more structures that are incorporated, the more accurate the alignment becomes, with a linear correlation between the two (O’Sullivan et al., 2004). The inclusion of one structure on a set of distantly related sequences increases accuracy by ~4%, suggesting more than one structure is needed for a significant improvement (O’Sullivan et al., 2004).

Staccato calculates a multiple structural alignment with the MultiProt program (Shatsky et al., 2004) and generates the final multiple sequence alignment by applying an iterative profile–profile alignment procedure. The profile representing the structure(s) incorporates information regarding amino

*To whom correspondence should be addressed.
acid type, respective distances and secondary structure. The profiles of the sequence and structure(s) are merged to produce the end alignment. Staccato generates results comparable with those from the HOMSTRAD database (Shatsky et al., 2006).

SAAlign incorporates structural information into the target-template alignment by operating a variable gap penalty scheme that attempts to keep insertions/deletions away from predicted secondary structural elements, the protein core, and between spatially distant residues. The method was tested on ∼200 sequence pairs of known structure, all with <40% sequence identities. The authors describe an increased alignment accuracy of ∼3.5% compared to a conventional affine gap penalty function (Madhusudhan et al., 2006).

In order to compare these three programs, a thorough analysis of high throughput, fully automated homology modelling is described here, with the accuracies of the different alignment strategies assessed by the quality of models generated. All modelling was carried out under testing conditions: at a target–template sequence identity ≤50%. Furthermore, a comparison of modelling based on single or multiple templates was also undertaken to determine which is the more effective strategy. According to previous studies the use of multiple parents, if properly selected, produces better models than those constructed from a single parent (Moult, 2005). However, another study suggested multiple templates work well at high levels of sequence similarity but not necessarily at lower levels (Venelovas and Margelevicuis, 2005).

In addition, we assessed five modelling programs at low target-template sequence similarity, specifically in relation to the accuracy of loop building. This was done in a ‘real world’ fashion from automatically generated sequence alignments, which do not always give the best local environment for loop construction, as is often the case when modelling difficult targets, e.g. in CASP: critical assessment of protein structure (Fiser et al., 2000; Rohl et al., 2004). The five programs tested: Builder (Koehl and Delarue, 1995), SegMod/ENCAD (Levitt, 1992), Modeller (Sali and Blundell, 1993), Nest (Petrey et al., 2003) and Swiss-Model (Schwede et al., 2004) have been reviewed elsewhere (Wallner and Elofsson, 2005); however, this previous report looked at overall performance, not loop building. This collection of programs represents the different techniques used in homology modelling, e.g. rigid-body assembly (Nest, Builder, Swiss-Model), segment-matching (SegMod/ENCAD), satisfaction of spatial restraints (Modeller).

Loop modelling generally falls in one of two categories: database or ab initio. Builder and SegMod/ENCAD use the former, Modeller and Nest use the latter and Swiss-Model uses a combination of the two. Builder and SegMod/ENCAD fit fragments that best match local geometry and sequence, selected from a local database, and subsequently underwent energy minimization. Modeller builds loops by optimizing a series of ‘probability density functions’ describing backbone geometry based on amino-acid type, and then refined with an energetic minimization procedure (Fiser et al., 2000). Nest uses the Loopy algorithm (Xiang et al., 2002), which generates a set of conformations for each loop, with the best selected by a ‘colony energy’ term that favours conformations low in energy but also close in structure to other conformations. Swiss-Model initially explores conformational space with constraint space programming and uses a force field-based scoring scheme to determine the best loop conformation. However, when this process fails or when loops are longer than 10 residues, a predetermined loop library is utilized (Schwede et al., 2004).

2 METHODS

All software implemented in this study constituted the latest versions available at the time. All programs were operated in their default modes and installed locally, except for Swiss-Model, which was accessed via its web server at the following address: http://swissmodel.expasy.org//SWISS-MODEL.html. Programs were linked where necessary with Perl scripts that were generated and executed on a 2GHz Suse 9.3 Linux workstation.

2.1 Test dataset

The modelling dataset was constructed from the SABMark database that contains sequences of low sequence identity (0–50%) (Van Walle et al., 2005). Models were produced in an automatic batch mode with ≤50% target–template sequence identities. This was to test the sequence alignment and modelling programs, not to necessarily produce the best models. 156 targets were extracted from SABMark. These targets and their respective templates (five for each target) constituted single-domain sequences from within the same SCOP Family, ensuring structural similarity amongst members (Murzin et al., 1995).

2.2 Template evaluation

Prior to modelling, template suitability was evaluated by sequence similarity to the target sequence with Blast E-values (Altschul et al., 1990) and Fasta sequence identities (Pearson, 2000), and by structural similarity to group cotemplates with CE-MC mean Z-scores (Guda et al., 2004). As all five templates in each group belong to the same SCOP family, structural similarity and mean Z-scores were expected to be high. This is an important pre-condition when using multiple templates and needs to be fulfilled for effective modelling (Bates et al., 2001).

2.3 Sequence alignment benchmark

The modelling protocol used three different sequence alignment programs: SAlign from Modeller v8.2 (Madhusudhan et al., 2006), Staccato v1 (Shatsky et al., 2006) and 3D-Coffee v3.79 (O’Sullivan et al., 2004) and one modelling program: Modeller v8.2 (Sali and Blundell, 1993). Multiple- and single-template modelling was performed for each target using each of the three sequence alignments generated by the respective programs mentioned earlier. Therefore, six models were generated for each target sequence.

A necessary step in modelling from multiple templates with Modeller is the multiple superposition of parent structures. Both SAlign and Staccato have an integrated multiple superposition method, but 3D-Coffee does not. Therefore, MultiProt (Shatsky et al., 2004) was implemented for this purpose.

Modelling was performed on 123 targets which passed cut-off thresholds designed to remove inappropriate templates. All templates had Blast E-values <1 and CE-MC average Z-scores ≥4 (Guda et al., 2004). Each target was modelled with 3, 4 or 5 templates (i.e. multiple template modelling) or with just one, the best template in the group selected by lowest E-value (i.e. single-template modelling).
In a second analysis, the effect of template number on sequence alignment accuracy and modelling quality was investigated with a smaller, more stringent dataset of 67 targets, each having five templates. All templates had an E-value <0.99, Z-score ≥4, and at least one template with an E-value ≤1 × 10^-3 and sequence identity ≥25%. Three models were generated for each of the 67 targets based on five templates, using the three sequence alignment programs described previously. The worst template (with highest E-value) was removed, leaving four templates for each target, and modelling was repeated. This process was continued through three templates, two templates and one template.

2.4 Model evaluation

Model quality was measured by automatic comparison to the known native structure. This was done using two methods:

1. Superposition of model onto native structure with the structure alignment program STructal (Gerstein and Levitt, 1998) and calculation of root mean square deviation (RMSD) of Ca atoms with ‘compare-structures’ function from Modeller v8.2.

2. Generation of Z-score, a measure of statistical significance between matched structures, for the model using the structure alignment program CE, with scores ≥4 indicating good structural similarity (Shindyalov and Bourne, 1998).

Model Z-score was also normalized by comparison to the (average) Z-score of the template structure(s), with the latter calculated in relation to the native structure of the sequence being modelled, not in relation to coterminates, as done previously in template evaluation. The resulting ratio of Z-scores reflects the degree of modelling success. A ratio >1 implies efficient modelling and a model that is better than the template(s). A ratio <1 suggests modelling underachievement and a model that is worse than the template(s).

2.5 Modelled loops benchmark

Five homology modelling programs were benchmarked based on their suitability for large-scale analysis and modelling methodology: Modeller v8.2 (Sali and Blundell, 1993), Nest v1.0 (Petrue et al., 2003), Builder v1.0 (Koehl and Delarue, 1995), SegMod/ENCAD v1.0 (Levitt, 1992) and Swiss-Model (Schwede et al., 2003). Nest and Modeller were each limited to one round of optimization (i.e. their default modes), with the output of single loop conformations. A modelling protocol was constructed for each program using single templates. Five models, one from each modelling program, were built for 111 targets, with all template E-values ≤0.001. All alignments were generated with 3D-Coffee for fair comparison. This simulates the real-world situation when loop building takes place with alignments that are not always optimal. As some models miss beginning/end residues, only equivalent residues present in all models were included in RMSD calculations, again for fair comparison.

The accuracy of loops was assessed with RMSD calculations that only involved loop residues. Loops are defined as areas in the target sequence that correspond to gaps in the template (also called insertions), but any gaps at the ends of the target-template sequence alignment are excluded. The two residues on either side of a gap are also classified as part of that particular loop, meaning the shortest loop modelled here is five residues long and results from an insertion of one residue. The longest loop modelled here is 16 residues long and results from an insertion of 12 residues.

3. RESULTS

3.1 Sequence alignment benchmark

A comparison of modelling based on three different sequence alignment programs and two separate template strategies, i.e. single and multiple, was made using a dataset of 123 targets. From the distributions of model RMSDs generated by Modeller from single and multiple templates (Figs 1 and 2), the biggest difference between the sequence alignment methods is observed in multiple-template modelling, where 3D-Coffee generates noticeably more models of low RMSD and fewer models of high RMSD compared to the two other methods. However, when modelling with the best available single template, the sequence alignment methods produce similar results, although SAAlign shows greater variability than the other two.

When average modelling quality is assessed, the difference amongst the three sequence alignment methods using a multiple-template strategy is clear (Table 1, Fig. 3). Based on these results 3D-Coffee achieves a 28% reduction in mean RMSD compared to SAAlign (t-test, P-value = 2.60 × 10^-10).
and a 15% reduction compared to Staccato ($t$-test, $P$-value = $1.74 \times 10^{-4}$). According to model $Z$-score, 3D-Coffee achieves a 3% increase in mean modelling quality compared to SAlign ($t$-test, $P$-value = $6.14 \times 10^{-3}$) and a 1% improvement compared to Staccato ($t$-test, $P$-value = $2.14 \times 10^{-1}$). The difference between the sequence alignment methods is less marked when using single templates (Table 1). 3D-Coffee outperforms SAlign and Staccato by the respective lesser margins of 13% ($t$-test, $P$-value = $1.52 \times 10^{-5}$) and 4% ($t$-test, $P$-value = $9.92 \times 10^{-2}$) with RMSD criteria, and by $Z$-score the three methods are equivalent.

Staccato yields a 15% improvement in model RMSD with multiple templates ($t$-test, $P$-value = $2.00 \times 10^{-5}$), and a 9% improvement with single templates when compared to SAlign ($t$-test, $P$-value = $1.52 \times 10^{-3}$, Table 1). In modelling with multiple templates, Staccato also gives a 2% increase in model $Z$-score compared to SAlign ($t$-test, $P$-value = $1.21 \times 10^{-1}$). Both Staccato and 3D-Coffee show a $Z$-score ratio $>1$ in multiple-template modelling but SAlign does not (Table 1). This indicates both 3D-Coffee and Staccato, when used with Modeller, combine template information efficiently, and on average produce a model that is better than the templates. However, this is less apparent when using single templates, with all methods showing a ratio $\leq 1$, suggesting that on average, modelling is not actually improving on the template and in some cases making it worse.

The use of single templates rather than multiple templates is the most productive strategy for SAlign and Staccato. When multiple templates are employed over the best single template, average model RMSD deteriorates by 16% with SAlign ($t$-test, $P$-value = $4.45 \times 10^{-3}$) and by 10% with Staccato ($t$-test, $P$-value = $4.75 \times 10^{-2}$), or using $Z$-score, 4% ($t$-test, $P$-value = $3.40 \times 10^{-3}$) and 3% ($t$-test, $P$-value = $2.29 \times 10^{-2}$), respectively. However, this is not true for 3D-Coffee, which is equally effective using either a multiple- or single-template strategy.

The exact influence of template number on modelling quality was investigated further with a second benchmark, made using a smaller dataset and more stringent template criteria resulting in better modelling quality throughout (Table 2). Models were constructed with Modeller from a varying number of templates (1–5) using each sequence alignment strategy.

### Table 1
Average modelling results from 123 targets (plus standard errors) with single/multiple templates and three sequence alignment methods

<table>
<thead>
<tr>
<th>Templates</th>
<th>SAlign</th>
<th>Staccato</th>
<th>3D-Coffee</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSD (Å) $Z$-score</td>
<td>$Z$-score ratio</td>
<td>RMSD (Å) $Z$-score</td>
</tr>
<tr>
<td>Single</td>
<td>7.34 (±0.53)</td>
<td>5.81 (±0.11)</td>
<td>0.99 (±0.01)</td>
</tr>
<tr>
<td>Multiple</td>
<td>8.77 (±0.52)</td>
<td>5.57 (±0.11)</td>
<td>0.99 (±0.01)</td>
</tr>
</tbody>
</table>

Fig. 3. Average modelling quality from 123 targets (standard error bars shown) with single/multiple templates and three sequence alignment methods. Modelling program: Modeller.
alignment technique. Figure 4 shows average modelling quality deteriorates as template number increases with SAAlign according to RMSD (and also by Z-score, Table 2), and to a lesser extent with Staccato, however, 3D-Coffee effectively maintains the same quality throughout and appears to be as effective with five templates as with one. The Z-score ratio also increases with template number for 3D-Coffee but not significantly so with Staccato or SAAlign (Table 2), further indicating 3D-Coffee, in combination with Modeller, maximises the structural content of multiple templates. All the alignment methods on average make models better, or as good as, the template(s), with no average Z-score ratios <1 (Table 2).

When model accuracy is compared to the (average) sequence identity between target and template(s), average model RMSD improves as sequence identity increases regardless of the number of templates used and the sequence alignment method implemented (Figs 5 and 6). Furthermore, the variation (standard error) in modelling quality also lessens as sequence identity increases. Although 3D-Coffee is generally the most effective across the entire sequence identity range here (20–50% identity), particularly it is more accurate than the other methods at very low levels (>30% ID), especially when a higher number of templates are used (Fig. 6).

### 3.2 Loop modelling benchmark

The loop building capabilities of five homology modelling programs: Builder, Nest, Modeller, SegMod/ENCAD and Swiss-Model, were tested using a dataset of 111 targets. Of these 111 targets, Builder failed to complete 11 models, and Swiss-Model failed to generate 15, resulting in a refined dataset of 88 targets that could be modelled by all five methods. Loops are defined as areas in the model, referred to as insertions that cannot be modelled from the template. Loop accuracy was measured by comparison with the native structure across the residues in question. In order to place loop accuracy in context, overall modelling was also assessed, i.e. the modelling of template-directed regions together with variant loops. Regarding this measure, the average modelling quality generated by the five programs is not dramatically different, especially in relation to the respective standard errors (Table 3, Fig. 7). However, Modeller marginally produces the best results compared to the other four programs.

For example, it gives an overall model RMSD that is 6% better than Swiss-Model ($t$-test, $P$-value = $4.44 \times 10^{-3}$) and a Z-score that is 3% better than Builder ($t$-test, $P$-value = $3.88 \times 10^{-4}$).

When the loop regions of the models are evaluated in isolation, the difference between the modelling programs is a little clearer (Table 3, Fig. 7). In the 88-target dataset, 120 loops were defined as areas in the model, referred to as insertions that cannot be modelled from the template. Loop accuracy was measured by comparison with the native structure across the residues in question. In order to place loop accuracy in context, overall modelling was also assessed, i.e. the modelling of template-directed regions together with variant loops. Regarding this measure, the average modelling quality generated by the five programs is not dramatically different, especially in relation to the respective standard errors (Table 3, Fig. 7). However, Modeller marginally produces the best results compared to the other four programs.

For example, it gives an overall model RMSD that is 6% better than Swiss-Model ($t$-test, $P$-value = $4.44 \times 10^{-3}$) and a Z-score that is 3% better than Builder ($t$-test, $P$-value = $3.88 \times 10^{-4}$).

When the loop regions of the models are evaluated in isolation, the difference between the modelling programs is a little clearer (Table 3, Fig. 7). In the 88-target dataset, 120 loops were defined as areas in the model, referred to as insertions that cannot be modelled from the template. Loop accuracy was measured by comparison with the native structure across the residues in question. In order to place loop accuracy in context, overall modelling was also assessed, i.e. the modelling of template-directed regions together with variant loops. Regarding this measure, the average modelling quality generated by the five programs is not dramatically different, especially in relation to the respective standard errors (Table 3, Fig. 7). However, Modeller marginally produces the best results compared to the other four programs.
required modelling with the average loop approximately 7 residues in length. Nest appears to yield the best results for loop modelling with Modeller second, SegMod/ENCAD third and Swiss-Model fourth. Nest generates loops that are on average 8% better by RMSD than those generated by Builder (t-test, P-value = 5.14 × 10⁻¹⁰) and 6% better than those generated by Swiss-Model (t-test, P-value = 3.40 × 10⁻⁹). Nest is 4% better than those generated by SegMod/ENCAD (t-test, P-value = 2.08 × 10⁻²) and 2% better than those generated by Modeller (t-test, P-value = 3.86 × 10⁻¹). As loop accuracy is affected by general backbone quality, which can differ from model to model, it is interesting that Nest produces the best results for loop modelling even though it does not appear to be quite as efficient as Modeller or SegMod/ENCAD in overall modelling.

For all five modelling programs, average loop RMSD generally increases with loop length. The shortest loops, five residues in length, are on average modelled with the greatest accuracy, and the longest loops, nine or more residues in length, are on average modelled the worst, with accuracy decreasing with loop length (Fig. 8).

4 DISCUSSION

In the context of the homology modelling methods benchmarked here, the choice of sequence alignment strategy is of greater importance in generating accurate models than the choice of modelling program, with the most significant improvement in modelling quality obtained by using the best available sequence alignment technique. This critical importance of sequence alignment has been established in previous studies (e.g. Martin et al., 1997; Venclovas and Margelevicius, 2005).

In relation to automatic modelling at low levels of sequence identity, obtaining an optimum sequence alignment is very unlikely. As a result the standard of modelling here is fairly low with many models clearly suffering from imperfect alignments or unsuitable template choice (despite the fact that homologous templates are used in all cases). Blast E-values and Fasta sequence identities were used to judge template suitability and both appear to be fair in this regard. The modelling performed from templates with E-values ≤ 1 × 10⁻³ (in the loops benchmark and second sequence alignment benchmark) is noticeably of higher quality than the modelling performed from templates with E-values < 1 (in the first
sequence alignment benchmark), as might be expected. Modelling quality is also seen to correlate with sequence identity; the higher the similarity, the better the modelling, a fact which has been noted previously (Tramontano, 1998). Indeed there appears to be a threshold at \(\sim30\%\): above this value generally produces acceptable models, but less than this value, modelling becomes unreliable. This indicates that despite the extra structural information used by the three sequence alignment techniques tested here, a reasonable level of sequence similarity is still required in order to have confidence in modelling. A previous study also noted a sequence identity threshold in homology modelling, in this case \(\geq25\%\), whereby most alignment methods can generate an acceptable model (Elofsson, 2002).

It might be anticipated that when operating at low sequence similarity, multiple templates might generate a better model than one generated from a single template. For example, multiple templates should provide more comprehensive coverage across the target sequence. However, the results show that with automatic modelling, using multiple templates rarely improves on a model generated from the best single template, and in some cases may even make it worse. It may be the case that multiple templates would be more effective if implemented in a manual protocol. For example, visually inspecting the compatibility between templates, and dictating which template should be used for modelling different segments of the target. However, this was beyond the scope of our study.

Of the three sequence alignment programs tested, 3D-Coffee in combination with Modeller appears to generate the most accurate modelling results. We suggest this may be due to the extra step that is implemented with the Fugue program—a sequence-structure alignment program in its own right. 3D-Coffee also appears to be equally effective when using either single or multiple templates. This is not the case with SAlign or Staccato, which appear to be more effective with a low number of templates. The key factor in multiple-template modelling is in maximizing the structural information contained in the templates and coping with any structural variation between them. 3D-Coffee with Modeller appears to be best able to do this, particularly at very low sequence similarity, where the differences between the sequence alignment programs are most evident. Therefore, the answer to the question of whether to use a single template or multiple templates is not a simple one because it depends on the sequence alignment technique and the level of target-template(s) sequence identity. However, generally speaking, using the best available single template is probably the safest strategy when sequence similarity is low and modelling is fully automatic.

Concerning loop building with the homology modelling programs, there is far less variation here than there is between the sequence alignment methods. Even though the margins involved are small, Modeller just outperforms the other programs in the low sequence identity range, producing good modelling accuracy overall. However, when loops are taken in isolation, Nest appears to produce the most accurate loop conformations, although a t-test does reveal that the difference between Modeller and Nest in loop building is not statistically significant. Therefore, it should be concluded that they are equal in accuracy. It is also interesting to note that the two leading loop building programs (i.e. Nest and Modeller) are both \textit{ab initio}, which may indicate that this type of technique is the superior method in loop generation.

Clearly though, all the loop building methods suffer as loop length increases, and on average loops longer than six residues are modelled inaccurately. This reflects the difficulty of constructing loops when modelling an entire target from scratch at low target-template sequence identity, where any inaccuracy in the sequence alignment is probably going to compound inaccuracy in loop building. Therefore in the ‘real world’, the accuracy of a particular modelling method is probably not the limiting factor when predicting long loop regions, especially as good levels of accuracy for loops up to 12 residues long have been demonstrated within native environments (Fiser \textit{et al.}, 2000; Rohl \textit{et al.}, 2004). This level of accuracy is clearly not possible when modelling loops within homology models, as demonstrated here, and it is likely that the quality of sequence alignment and respective loop take-off points are the most important factors when building loops longer than just a few residues.

**ACKNOWLEDGEMENTS**

We thank all the authors of the programs used in this study for their help with implementation. We also thank the BBSRC for funding in the form of a studentship for J.A.R.D.

**Conflict of Interest:** none declared.

**REFERENCES**


