On possible use of fold recognition models for molecular replacement

Antonio del Sol Mesa and David T. Jones

Department of Computer Science,
Bioinformatics Unit,
University College London,
Gower Street, London WC1E 6BT, U.K.

This technical report describes part of the work carried out in the following BBSRC-funded research project:

BBSRC Grant Reference: B14023

Title: *Extending threading methods to comparative modelling and molecular replacement for structural genomics*

Institution of grant: University College London

Principal Investigator: Jones D

Abstract

In this study, we look at molecular replacement to see how well methods can find the correct orientation and position of the phasing model in cases where the phasing model has low sequence identity with respect to the target protein (less than 24%). The analysis has been performed on a benchmark set of 34 single chain, single domain proteins (target proteins) exhibiting different folds and with available structure factors. Using a fold recognition method, we have been able to propose usable phasing templates, which are remote homologues for each protein in the list, giving a total of 71 target-template pairs in total.

We have explored the dependence of the quality of the molecular replacement solutions on the completeness of the phasing models by selecting six different atom subsets from the model coordinate files. The performance of molecular replacement in terms of different characteristics of the target and template proteins, such as resolution, is also evaluated.

This study represents an important first step towards the general application of threading methods to the problem of solving a protein structure by X-ray crystallography. The results clearly indicate that molecular replacement is quite tractable for even remote homologues, and as such, the successful development of this approach would be of great benefit in structural genomics initiatives.

Introduction

It is well-known that the 3-D structure of a protein can provide a great deal of information about its biological function and mechanism. The main reason for this is that protein structure is much more highly conserved than protein sequence, and so the tertiary structure of a protein much more readily provides clues as to distant common ancestry. As a result of this, there is now a great deal of interest in determining large numbers of protein structures to assist in the elucidation of the functions of genes. This interest has now manifested itself as a number of ongoing structural genomics initiatives, which aim to solve experimentally the structures of every protein encoded by a bacterial genome, for example, or all of the structures within particular functional classes.

Few techniques are available for determining the structures of proteins to atomic resolution, however. At present, two techniques can provide the three-dimensional structure of a protein to high resolution: X-ray or neutron diffraction analysis and NMR analysis of small proteins in solution. Of these, X-ray crystallography is by far the most widely used approach due to the limitations of NMR techniques in solving the structures of large proteins.

Since the pioneering work by Perutz and Kendrew on the structure of the hemoglobin and myoglobin in the 1950s, there have been great improvements to the basic techniques of X-ray crystallography. For example, the use of synchrotron radiation sources and the introduction of sophisticated computer hardware and software have reduced the time to determine new structures while increasing the accuracy of the results. Although the expression, purification and crystallisation of the target protein still remain as the rate-limiting step in structure determination, great strides have been taken towards their automation, particularly as a result of developments accruing from the structural genomics initiatives. Once the protein is crystallized and the native diffraction data is collected, in order to calculate an electron density map the phase information is needed. Most protein structures are determined by traditional experimental intensity methods such as the heavy metal isomorphous replacement and anomalous scattering methods, but a alternative approach is that of molecular replacement, which can solve a structure with relatively little effort in comparison with other methods. However, molecular replacement has the distinct disadvantage of being seen to require a closely related protein of known 3-D structure to act as a phasing model. No systematic studies have been done to explore the limits of molecular replacement, however.

The molecular replacement method, pioneered by Rossmann and Blow (1), has the goal of orienting and translating a search model in such a way, that it coincides with the position of the unknown protein in the crystal. This is done by calculating the correlation function (or overlapping function), between the observed peaks (Pobs.) and calculated peaks (Pcal.) from the model Patterson functions.

Of course, the higher the structural similarity between the search model and the target protein, the easier it is to identify the correct placement of the phasing model in the unit cell of the crystal. This is usually achieved by choosing a model protein with high sequence similarity to the target protein. However, it is now well established that many proteins sharing little or no obvious sequence similarity can show remarkable similarities in their native folds, such as the various TIM barrel enzymes or the diverse globin superfamily for example. Indeed, some pairs of globins can have backbone RMSDs (Root Mean Square Deviations) of as little as 1.9 Å, despite having less than 20% sequence identity. In further support of this, it has been pointed out (2) that there are some examples of protein structures taken from the PDB, which have apparently been solved using molecular replacement with phasing models exhibiting low sequence identity (less than 20%) with respect to their corresponding target proteins. In the same publication, it was also shown that some structures have been solved using phasing models with a C-alpha RMSD with respect to their target proteins within the interval 2.5-3.0 A, and with a percentage of equivalent residues between the target and the template proteins as little as 50%. These observations suggest that molecular replacement techniques should not be considered limited only to close homologues.

These ideas have motivated us to explore the possibility of using fold recognition or threading methods to provide suitable molecular replacement search models in cases where no close sequence homologue of known 3-D structure is available. Originally, the idea of threading (3) was to recognize folds in the absence of sequence similarity and, therefore, the template sequence information was usually not taken into account. However, with the growth of sequence and structure data banks, a number of new threading methods have been proposed, which incorporate sensitive sequence comparison algorithms; for example, GenTHREADER (4) is a method that uses a traditional sequence alignment algorithm and generates alignments which are evaluated by knowledge-based potentials of mean force. In general, the combination of sequence profile alignment methods with fold recognition has increased the quality of the proposed models even in the cases of very distant homology.

Here, we study the possibility of using threading models as phasing templates for molecular replacement in cases of very low sequence identity (< 20%) between the target and model proteins. We analyse the effectiveness of molecular replacement in finding the right orientation and position of the phasing model respect to the target protein; taking into account different characteristics such as the resolution of the phasing model, the completeness of the initial model, its percentage of secondary structure elements, the quality of the sequence alignment provided by GenTHREADER, the percent of the target aligned residues, and the RMSD (Root Mean Square Deviation) between the target and the model.

We selected a list of 34 single chain, single domain proteins, with their corresponding coordinate and structure factor files deposited in the RCSB Protein Data Bank (Berman et. Al.,2000(5)). These proteins were considered as target proteins, and their structure factor files were our starting point for testing the performance of molecular replacement. Using GenTHREADER, we built the best threading models (according to the GenTHREADER score) for each protein sequence in the list, providing that the sequence identity was less than 24% and the C-alpha RMSD was less than 3.0 Å between the protein and each of its models. We also required each model to have at the least 40% of its residues belonging to secondary structure elements (alpha helices or beta strands). These threading models were considered as phasing templates for MOLREP (7), which is a highly automated implementation of molecular replacement and thus ideal for this study. Since each of the 34 target proteins had more than one suitable phasing template, we eventually built a list of 71 target-template pairs.

For each target-template pair, six different possibilities for the template were considered depending on its completeness: keeping all its atoms, just considering its main chain atoms, just considering its C-alpha atoms, and again the first three combinations, but excluding the most variable parts (loops). This allowed us to see the effect on the success of molecular replacement of the completeness of the phasing model. We also carried out an analysis of the MOLREP performance judged against the characteristics of the model and target protein by comparing those cases where MOLREP succeeded from those where it failed.

Results and Discussion

We have analysed the effectiveness of molecular replacement (as implemented by MOLREP) in finding the right orientation and position of the phasing model in the unit cell for 71 target-model pairs (Table I). The targets (single chain and single domain proteins) were taken from PDB and the models for low sequence similarity homologues of the targets were generated by means of a fold recognition method (GenTHREADER). In all cases the percent of identity was less than 24% and the optimal C-alpha RMSD between the target and the model less than 3.0 Å. The percentage of model residues in secondary structure elements (alpha helices or beta strands) was at least 40%.

For each target-template pair in the list, we tested how the quality of the MOLREP solution depended on the selection of atomic information used in the phasing model. Thus, we built six different models for each phasing template based on its aligned residues (according to the GenTHREADER alignment): keeping all the atoms, just considering the main chain atoms, taking only the C-alpha atoms, and the first three combinations but only for residues in secondary structure elements (removing loops). Each of the six combinations was used to provide a phasing template input for MOLREP in each case.

We considered that MOLREP had found the right orientation and position of the phasing template if the C-alpha RMSD between the target and the rotated and translated template was at most 3 Å away from the optimal C-alpha RMSD between the target and the template as calculated by a rigid body superposition method.

In around 54% of the target-template pairs (38 pairs out of 71) MOLREP found a solution that satisfied out condition for correctness in at least one of the six combinations for the phasing template. These 38 pairs are indicated in Table I with bold face letters.

The results in Fig.1 show that the number of MOLREP successes depends on the method used to derive the phasing model from the initial sequence alignment. Although in general MOLREP finds the right orientation and position in those cases where we keep all the aligned residue atoms or the residue main chain atoms in the phasing model, we see that there are other cases when the right solution has only been found with one of the other combinations. In light of this it is quite difficult to predict in advance which of the strategies would be best for any given case, and therefore the results suggest that all six of the strategies should be used in parallel.

We then proceeded to analyse the dependence of the effectiveness of MOLREP on different characteristics of the target and the template. In order to do that, we calculated for each target-template pair the C-alpha RMSD for the six different atom selection combinations. We considered the minimum RMSD obtained out of the six possibilities to be the "baseline" i.e. the best possible RMSD for any solution generated by MOLREP for a given target-template pair.

Fig. 2a shows the dependence of the best MOLREP RMSD on the optimal RMSD between the target and the template. If we analyse this dependence in the region where the best MOLREP RMSD is ≤ 3.0 Å (Fig.2b), we can see a clear correlation between the best MOLREP RMSD and the optimal RMSD (straight line). This result can be expected because it is well understood that the closer the template fold is to the target fold the more efficient MR methods are in finding the right orientation and position of the phasing model.

In the region where the best MOLREP solutions are greater than 3.0 Å, we do not see a clear correlation, but this is due to the fact that in all these cases the right solution was not found by MOLREP, so MOLREP is not sensitive enough to distinguish among these cases.

The percentage of target residues aligned by GenTHREADER with respect to the phasing template was another factor we saw had some influence on the quality of the MOLREP solution. Clearly phasing models which only match a short region of the target protein are unlikely to provide a clear indication of the correct peak. In Fig.3a this dependence is shown, and in Fig. 3b we clearly notice that in the region where the best MOLREP solution was not greater than 3.0 Å, all the solutions corresponded to cases with no less than 80% of target-template overlap. There was just one solution with an overlap less than 80%, but even this case is quite close to that value. A range of overlaps is observed for the best MOLREP solutions with RMSDs greater than 3.0 Å, which is again a result of the lack of sensitivity of MOLREP for these cases.

One interesting question in this analysis is to what extent the quality of the GenTHREADER alignment is important for the success of MOLREP. To study this point, we selected a list of 50 target-template pairs with existing FSSP files for the target proteins, in such a way that we had the possibility of comparing the GenTHREADER alignment with the structural FSSP alignment in all these cases. Since the analysis of MOLREP is based on the structural information of the phasing template, in other words, MOLREP takes as an input the PDB file of the phasing model containing just the aligned residues (in this case by GenTHREADER), we decided to calculate the percentage of agreement between the FSSP alignment and the GenTHREADER alignment, and to see the dependence of the MOLREP solution quality on this parameter.

Fig. 4a shows the dependence of the best MOLREP solution on the percentage of agreement between the FSSP alignment and the alignment generated by GenTHREADER. Fig. 4b shows that all the best MOLREP solutions with RMSDs smaller than 3.0 Å have a percent of agreement of at least 50%. One interesting observation is the apparent correct solution (2.30 Å RMSD) when none of the residues in the template have apparently been correctly aligned. In this case the target is a transcription regulation protein with a helix hairpin fold (PDB code: 1nkd), and the phasing template is the chain A of a signalling protein (PDB code: 1qu7) with a double helical-bundle structure (Fig.5a,b). The FSSP structural alignment included part of each of the helices with the linking region of the phasing template, while the GenTHREADER alignment included part of one of the helices. This is one of the cases where although the alignments are not in agreement with the optimum structural alignment, the information in the model is sufficient for MOLREP to find the right orientation and position. There are other cases where, for example, the target protein has a symmetrical structure, and MOLREP can find the right orientation and position of the phasing template as long as at least one of the symmetric units of the target was correctly aligned to the phasing template.

Another interesting point is the effect of the percentage of residues in the template belonging to secondary structure elements (alpha helix or beta strand) on the MOLREP solution. This effect is shown in Fig. 6a, and in Fig. 6b we see that at least 50% secondary structure is required to generate the best MOLREP solutions with a RMSD smaller than 3.0 Å. Proteins with substantial fractions of coil residues do not generally make good phasing models.

We also looked at the crystallographic resolution of the phasing templates, but we did not find a clear correlation between the former and the quality of the MOLREP solutions, since in all cases we were dealing with high to medium resolution structures acting as phasing templates (maximum 2.3 Å). However, we still believe this is an important factor to be taken into account for the success of molecular replacement, though our existing benchmark set may not be large enough to show it.

Overall, these results are in broad agreement with our previous rather ad hoc observations (2) where we noted that among 329 PDB entries solved by MR (with existing FSSP files for the target proteins), the majority of the cases exhibited a C-alpha RMSD up to 2.0 Å between the target and the phasing template, and a percentage of aligned target residues above 90%. However, as we have found in this more systematic study, there were some cases where this RMSD was as high as 3.0 Å, and the percentage of aligned target residues was as low as 50%.

Methods

A list of 34 single chain, single domain proteins belonging to different families, and with their corresponding Structure Factor files available in the RCSB Protein Data Bank^5 were selected from the CATH database (6). These chains were taken to be the targets (i.e. the desired experimental structures) for our simulated molecular replacement studies. For each of the 34 targets, we identified template structures which had a maximum of 24% sequence identity and a maximum of 3 Å RMSD from the target and a minimum of 40% of the residues belonging to the secondary structure elements (alpha helices or beta strands). Due to the limited availability of structure factors and the above requirements, the largest benchmark set that we could compile was 71 template-target pairs.

The molecular replacement package we opted to use was MOLREP (7), written by Alexei Vagin. Although several other MR packages are available, MOLREP offers the highest degree of automation, which is an important factor if MR techniques are to be built into a pipeline for high-throughput structure determination. In order to generate "low-homology" comparative models for a number of target-template pairs, GenTHREADER (4) was used to generate a number of models for each protein in the target list with a fold library built from the current release of FSSP (Holm and Sander, 1996).

The program SAP (8) was used to generate the structural alignments between each target and its corresponding threading models. Taking these alignments and using the program ProFit (9), we performed a least squares fits between each target protein structure and each of its threading model structures and thereby calculated the optimal C-alpha RMSD between them.

Since one of the goals of this work was the analysis of the dependence of the MR success on the atomic information of the phasing template, for each target-template pair we proceeded as follows. Using the GenTHREADER alignment a phasing model was constructed from the PDB files (5) using just the aligned residues. Prosthetic groups (heteroatoms), crystallographic waters, and all hydrogen atoms (where present) were excluded. For each model we considered six different methods for constructing the phasing templates: including all the atoms of all residues in the PDB file of the recognition model, including only the residue main chain atoms, including just the residue C-alpha atoms, and repeating the first three models excluding the loops.

Once MOLREP had been run for each of the six models of each target-template pair in the list, we analysed how well MOLREP had determined the correct rotation and translation of the models with respect to the target proteins. In each case, we compared the optimal C-alpha RMSD between the target and its template given by rigid body superposition to the calculated C-alpha RMSD between the target protein (and all of its symmetry related copies) and the template, rotated and translated by MOLREP.

In order to see the effect of the correctness of the alignments used to generate the phasing models on the quality of the MOLREP solutions, we selected a subset of 50 target-template pairs with existing FSSP (10) files for the targets. Using the FSSP alignments as a "gold standard" we compared the GenTHREADER alignments to the FSSP structural alignments and determined a simple percentage of correctly aligned residues.

Conclusions

We have shown here that even for cases where phasing models are based on very remotely related templates (< 24% sequence identity), in 54% of the examples, MOLREP was able to find the right orientation and position for at least one of the six template generating strategies tried in each case. This is quite contrary to the rule of thumb for molecular replacement, which stipulates that only phasing models based on very closely related proteins are likely to succeed. Although we can see some trends which explain why some of the models are not successful, clearly more work is required to rationalise the reasons for failure, and perhaps extend the approach to handle more difficult cases still.

So far, of course, our studies have been limited to single chain, single domain proteins, but we hope to extend our work to consider target proteins with multiple chains and multiple domains. However, a number of difficult technical issues will need to be addressed before this will be possible. Nevertheless, we feel that the results shown clearly demonstrate that fold recognition methods can be routinely applied to the problem of phasing X-ray diffraction data for proteins. Although it is not within the scope of this paper, we are currently working on implementing a completely automatic method for building phasing templates from the best threading models for a given target protein. We would hope that this kind of software would be of tremendous benefit for structural genomics.

Acknowledgments

We are grateful to Dr. Alexei Vagin for providing us with his Molecular replacement program MOLREP, and for useful discussions and comments on this research.

References

1. Rossmann, M. G. & Blow, D. M. (1962). The Detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallogr.* **15**, 24-31.

2. Jones, D. T. (2001). Evaluating the potential of using fold-recognition models for molecular replacement. *Acta Crystallogr. D*, **57**, 1428-1434.

3. Jones, D. T., Taylor, W. R., & Thornton, J. M. (1992). A new approach to protein fold recognition. *Nature*, **358**, 86-89.

4. Jones, D. T. (1999). GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. *J. Mol. Biol.* **287**, 797-815.

5. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Res.* **28**, 235-242.

6. Orengo, C. A., Michie, A. D., Jones, S., Jones, D. T., Swindells, M. B., & Thornton, J. M. (1997). CATH--a hierarchic classification of protein domain structures. *Structure*, **5**,1093-1108.

7. Vagin A. & Teplyakov A. (1997). MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **30**, 1022-1025.

8. Taylor, W. R. (2000). Protein structure comparison using SAP. *Methods Mol. Biol.* **143**,19-32.

9. Martin, A.C.R., http://www.bioinf.org.uk/software/profit/.

10. Holm, L. & Sander, C. (1996). Mapping the protein universe. *Science*, **273**, 595-603.

Figure Legends

Table I. List of target-model pairs with their corresponding percent of identity and RMSD between the target and the model. The examples where MOLREP succeeded in finding the right orientation and position are denoted with bold face letters.

Figure1. Dependence of the best MOLREP solutions on the type of the model.

Figure2a,b. MOLREP RMSD versus optimal RMSD for all solutions up to 13.0 Å (a), and for just the solutions up to 3.0 Å (b) of MOLREP RMSD's.

Figure3a,b. MOLREP RMSD versus percent of target aligned residues for all solutions up to 13.0 Å (a), and for just the solutions up to 3 Å (b) of MOLREP RMSD's.

Figure4a,b. MOLREP RMSD versus percentage of agreement between the FSSP alignment and the fold recognition alignment for all solutions up to 13 Å (a), and for just the solutions up to 3 Å (b) of MOLREP RMSD's.

Figure5a,b. Example where the transcription regulation protein with a helix hairpin fold (1nkd) was used as a target protein (a), and the chain A of a signaling protein (1qu7) with a double helical-bundle structure was used as phasing template.

Figure6a,b. MOLREP RMSD versus percent of template secondary structure elements for all solutions up to 13.0 Å (a), and for just the solutions up to 3.0 Å (b) of MOLREP RMSD's.

TARGET	MODEL	%IDENTITY	RMSD
1a6m	1h97A	18.10	1.90
1a6m	1ithA	15.10	1.50
1a6m	2gdm	17.00	2.30
1a6m	1flp	17.30	1.30
1a6m	1hlb	17.10	1.90
1a6m	**1cg5B**	**19.90**	**1.50**
2lbd	**1a28A**	**15.50**	**2.40**
1nls	1gv9A	18.10	1.60
1thx	**1b9yC**	**15.10**	**1.50**
1thx	**1erv**	**24.30**	**1.30**
1rcf	**1bvyF**	**15.30**	**2.00**
1rcf	**1amoA**	**21.10**	**1.90**
1ido	**1auq**	**17.30**	**2.00**
1ido	**1jeqB**	**6.60**	**2.70**
1ido	1jeqA	9.60	2.70
1btl	**1skf**	**15.80**	**2.40**
1btl	1hd8A	13.30	2.20
1nkd	**1qu7A**	**5.10**	**0.20**
1flp	1h97A	11.30	2.00
1flp	**1cg5B**	**10.40**	**2.00**
1flp	**1a6m**	**17.30**	**1.60**
1onc	**1dytA**	**22.00**	**1.40**
1at6	**1clc**	**7.40**	**2.80**
1hij	**1eteA**	**8.10**	**2.20**
1cpm	**1dypA**	**19.70**	**2.60**
1cpm	1a3k	11.50	2.50
1cpm	**1lcl**	**8.80**	**2.40**
1cpm	**2mprA**	**3.20**	**0.30**
1byh	1a3k	12.20	2.40
1byh	**1d2sA**	**7.90**	**1.90**
1byh	1lcl	8.80	2.40
2bu4	1qcxA	9.80	2.40
1bpi	**1bj5**	**5.20**	**2.90**
1mlu	**1h97A**	**18.10**	**1.80**
1mlu	**1ithA**	**14.30**	**1.60**
1mlu	2gdm	15.50	2.30
1mlu	1flp	17.30	1.30
1mlu	1hlb	15.00	1.80
1myt	2hbg	20.00	1.80
1myt	1ithA	19.90	1.50
1myt	**1h97A**	**19.00**	**2.00**
1myt	**2gdm**	**17.00**	**2.50**
1myt	**1ewaA**	**12.60**	**1.70**
4fiv	1eagA	15.20	2.90
2fmb	1fknA	12.70	2.80
1a33	**1jceA**	**7.50**	**2.20**
1cri	**1ctj**	**11.90**	**1.70**
1emk	**1h4uA**	**10.60**	**2.40**
1emk	**2por**	**9.00**	**1.60**
1hik	**1eteA**	**8.20**	**2.10**
1mho	1alvA	18.30	1.60
1mho	**1dguA**	**11.90**	**0.10**
1ptk	**1ga1A**	**19.10**	**2.30**
1ptk	**1cnv**	**4.90**	**2.90**
1thy	1b5eA	16.00	2.30
1uic	**153l**	**9.20**	**2.30**
2mm1	1hlb	17.10	2.10
2mm1	1flp	16.50	1.40
2mm1	1ithA	16.50	1.60
2mm1	1h97A	19.70	2.10
2mm1	2gdm	16.30	2.40
2mm1	1cg5B	20.70	1.50
3tlh	1fknA	13.30	2.80
3tlh	1eagA	16.70	2.50
1bsy	**1mup**	**15.50**	**1.80**
1bsy	1aqb	17.10	2.30
1lu1	**1gv9A**	**20.30**	**1.70**
1bk1	2nlrA	14.90	1.90
1bk1	**1by5A**	**2.00**	**2.30**
1bk1	**1i5pA**	**6.30**	**2.60**
9ilb	1wba	10.90	2.50

TABLE I

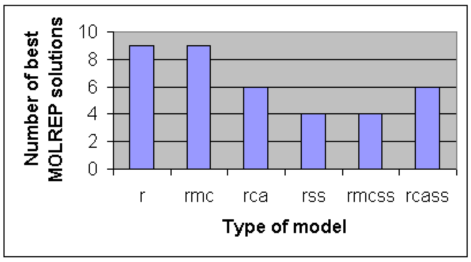

Number of best MOLREP solutions
10
8
6
4
2
0
r
rmc
rca
rss
rmcss
rcass
Type of model

Fig.1

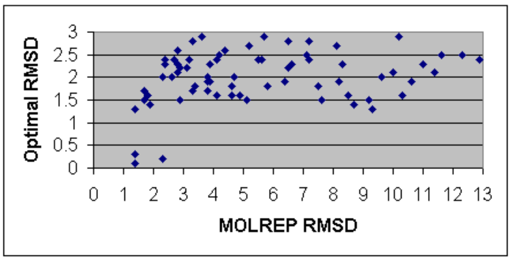

Optimal RMSD
3
2.5
2
1.5
1
0.5
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13
MOLREP RMSD

Fig.2a

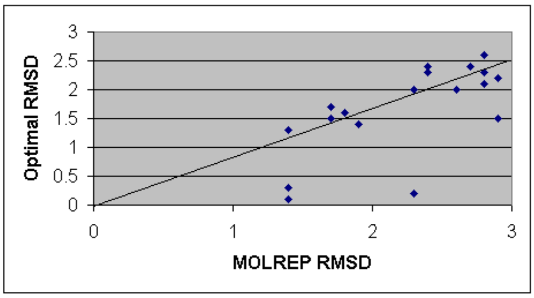

Optimal RMSD
3
2.5
2
1.5
1
0.5
0
0
1
2
3
MOLREP RMSD

Fig.2b

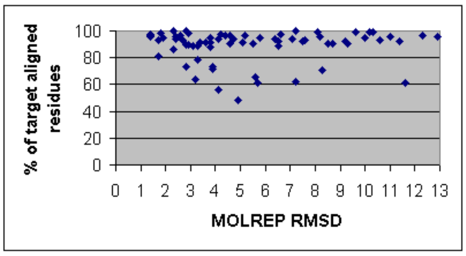

% of target aligned residues
100
80
60
40
20
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13
MOLREP RMSD

Fig.3a

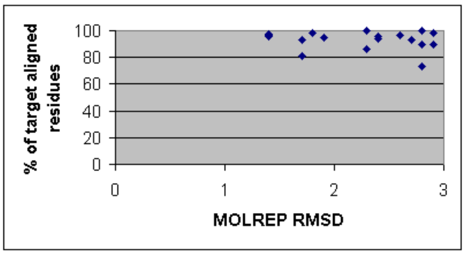
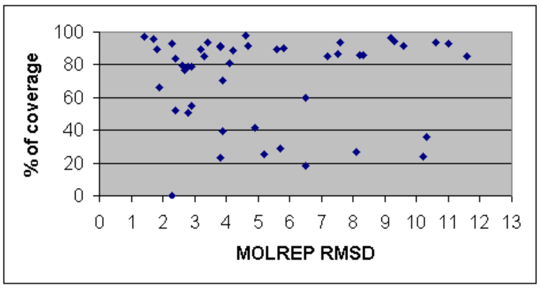

% of target aligned residues
100
80
60
40
20
0
0 1 2 3
MOLREP RMSD

Fig.3b



100
80
60
40
20
0
% of coverage
0
1
2
3
4
5
6
7
8
9
10
11
12
13
MOLREP RMSD

Fig.4a

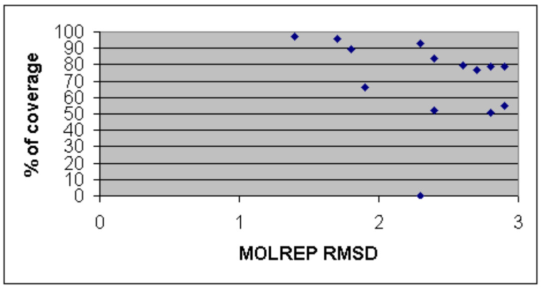


Fig.4b

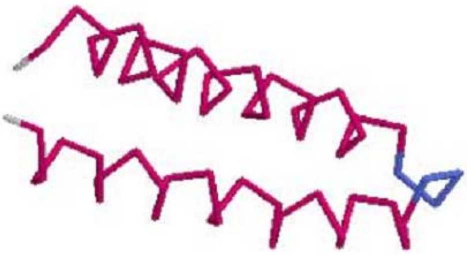


Fig. 5a

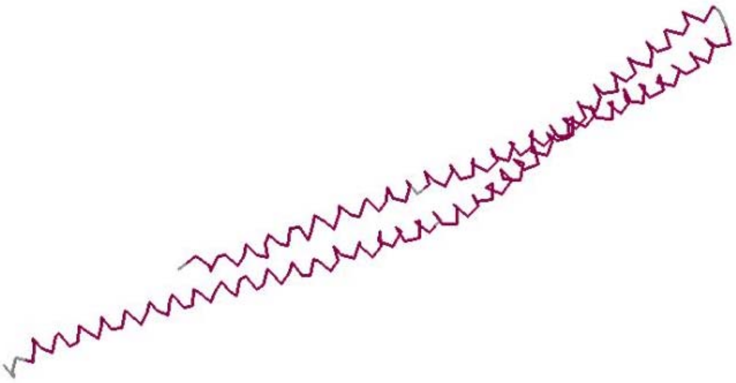


Fig.5b

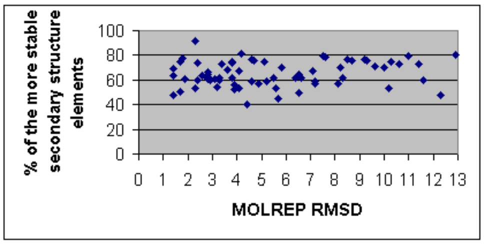

% of the more stable secondary structure elements
100
80
60
40
20
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13
MOLREP RMSD

Fig.6a

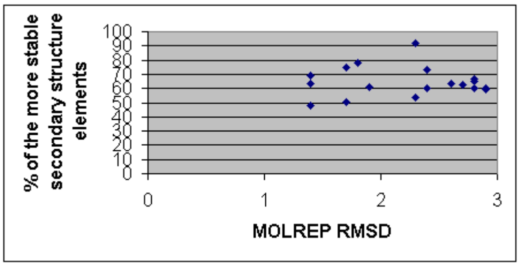

% of the more stable secondary structure elements
100
90
80
70
60
50
40
30
20
10
0
0 1 2 3
MOLREP RMSD

Fig.6b