We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK
CITATION
INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Contributions of Structure Comparison Methods to the Protein Structure Prediction Field

David Piedra[1], Marco d'Abramo[2] and Xavier de la Cruz[1,3]
[1]IBMB-CSIC
[2]CNIO
[3]ICREA
Spain

1. Introduction

Since their development, structure comparison methods have contributed to advance our understanding of protein structure and evolution (Greene et al, 2007; Hasegawa & Holm, 2009), to help the development of structural genomics projects (Pearl et al, 2005), to improve protein function annotations (D. A. Lee et al), etc, thus becoming an essential tool in structural bioinformatics. In recent years, their application range has grown to include the protein structure prediction field, were they are used to evaluate overall prediction quality (Jauch et al, 2007; Venclovas et al, 2001; Vincent et al, 2005; G. Wang et al, 2005), to identify a protein's fold from low-resolution models (Bonneau et al, 2002; de la Cruz et al, 2002), etc. In this chapter, after briefly reviewing some of these applications, we show how structure comparison methods can also be used for local quality assessment of low-resolution models and how this information can help refine/improve them.

Quality assessment is becoming an important research topic in structural bioinformatics because model quality determines the applicability of structure predictions (Cozzetto et al, 2007). Also, because prediction technology is now easily available and potential end-users of prediction methods, from template-based (comparative modeling and threading) to *de novo* methods, are no longer specialized structural bioinformaticians. Quality assessment methods have been routinely used for many years in structural biology in the evaluation of experimental models. These methods focus on several features of the protein structure (see (Laskowski et al, 1998) and (Kleywegt, 2000) and references therein). Because a number of quality issues are common to both experimental and predicted models, the use of these methods has been naturally extended to the evaluation of structure predictions. For example, in the case of homology modeling, a widely used structure prediction technique, evaluation of models with PROCHECK (Laskowski et al, 1993), WHAT-CHECK (Hooft et al, 1997), PROSA (Sippl, 1993), and others (see (Marti-Renom et al, 2000) and references therein) is part of the standard prediction protocol; WHATIF (Vriend, 1990) and PROSA (Sippl, 1993) have also been used in the CASP experiment to assess comparative models (Venclovas, 2001; Williams et al, 2001); etc.

Some quality assessment problems are unique to the structure prediction field, given the specific characteristics of computational models, and have led to the development of

methods aimed at: the recognition of near-native predictions from a set of decoys (Jones & Thornton, 1996; Lazaridis & Karplus, 2000; Sippl, 1995); identification of a target's protein family (Bonneau et al, 2002; de la Cruz et al, 2002); overall quality assessment of predictions (Archie et al, 2009; Benkert et al, 2009; Cheng et al, 2009; Larsson et al, 2009; Lundstrom et al, 2001; McGuffin, 2009; Mereghetti et al, 2008; Wallner & Elofsson, 2003; 2005; Z. Wang et al, 2009; Zhou & Skolnick, 2008); and, more recently, residue-level quality assessment (Benkert et al, 2009; Cheng et al, 2009; Larsson et al, 2009; McGuffin, 2009; Wallner & Elofsson, 2006; 2007; Z. Wang et al, 2009). However, in spite of these promising efforts, quality assessment of protein structure predictions remains an open issue(Cozzetto et al, 2009).

Here we focus on the problem of local quality assessment, which consists on the identification of correctly modeled regions in predicted structures (Wallner & Elofsson, 2006; 2007), or, as stated by Wallner and Elofsson(Wallner & Elofsson, 2007): "The real value of local quality prediction is when the method is able to distinguish between high and low quality regions.". In many cases, global and local quality estimates are produced simultaneously (Benkert et al, 2009; Cheng et al, 2009; Larsson et al, 2009; McGuffin, 2009). However, in this chapter we separate these two issues by assuming that, irrespective of its quality, a structure prediction with the native fold of the corresponding protein is available. From a structural point of view this is a natural requirement, as a correct local feature (particularly if it is one which, like a β-strand (Chou et al, 1983), is stabilized by long-range interactions) in an otherwise wrong structure can hardly be understood. From a practical point of view, successful identification of correct parts within incorrect models may lead to costly errors. For example, identifying a correctly modeled binding site within a structurally incorrect context should not be used for drug design: it would surely have incorrect dynamics; the long-range terms of the interaction potential, like electrostatics, would be meaningless; false neighboring residues could create unwanted steric clashes with the substrate, thus hampering its docking; or, on the contrary, absence of the true neighbors could lead to unrealistic docking solutions; etc. In the remaining of the chapter we describe how structure comparison methods can be applied to obtain local quality estimates for low-resolution models and how these estimates can be used to improve the model quality.

2. A simple protocol for local quality assessment with structure comparison methods

As mentioned before, an important goal in local quality assessment(Wallner & Elofsson, 2006; 2007) is to partition the residues from a structure prediction in two quality classes: high and low. This can be done combining several predictions; however, in the last two rounds of the CASP experiment -a large, blind prediction experiment performed every two years(Kryshtafovych et al, 2009)- evaluators of the Quality Assessment category stressed that methods aimed to assess single predictions are needed(Cozzetto et al, 2007; Cozzetto et al, 2009). These methods are particularly important for users that generate their protein models with *de novo* prediction tools, which are still computationally costly(Jauch et al, 2007), particularly for large proteins.

Here we describe a single-molecule approach, based on the use of structure comparison methods, that allows to partition model residues in two sets, of high and low quality respectively. In this approach (Fig. 1), the user's model of the target is first structurally aligned with a target's homolog. This alignment, which constitutes the core of the procedure, is then used to separate the target's residues in two groups: aligned and

unaligned. The main assumption of this approach is that aligned residues are of higher quality than the average. The validity of this assumption is tested in the next section. In section 3 we discuss the conditions that determine/limit the applicability and usefulness of the method.
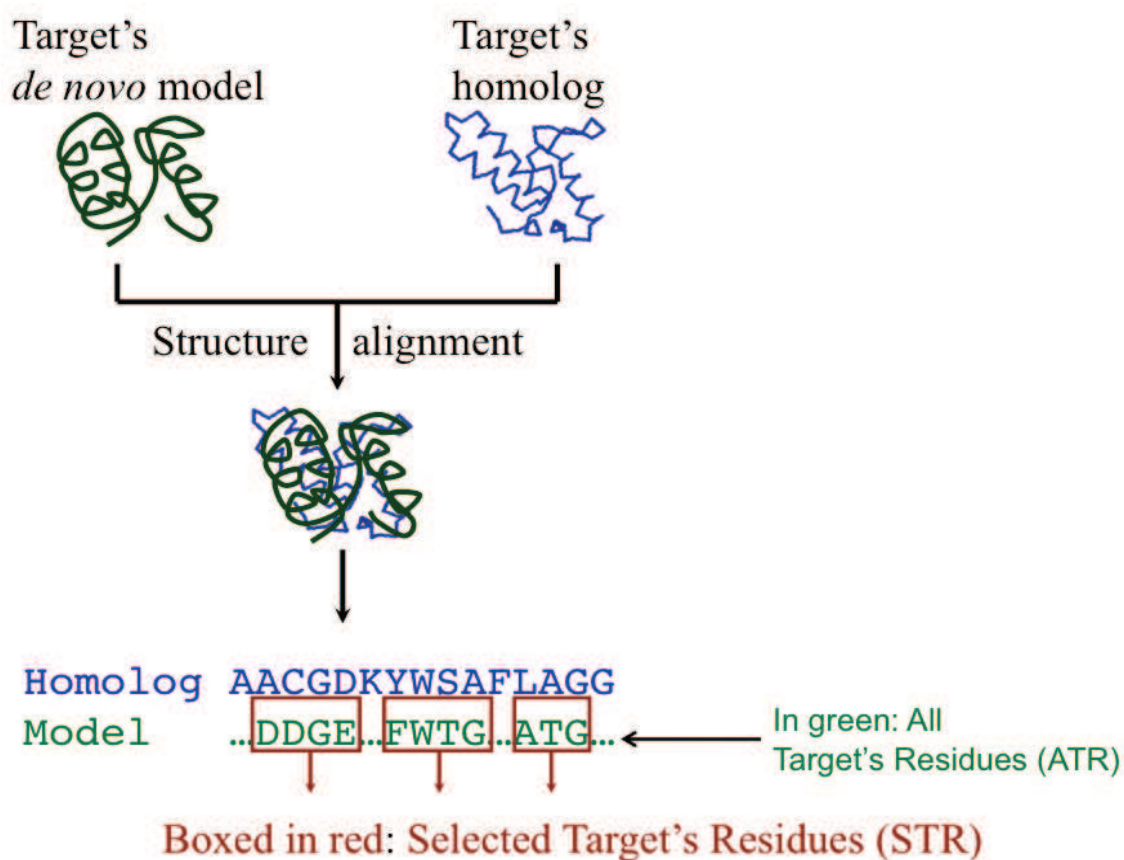

Fig. 1. Use of structure comparison methods for local quality assessment of structure predictions

2.1 Performance of structure comparison methods in local quality assessment

To show that structurally aligned residues are usually of higher quality we used a set of *de novo* predictions with medium/low to very low resolution, obtained with Rosetta(Simons et al, 1997). Although several *de novo* prediction programs have shown promising results in the CASP experiment(Jauch et al, 2007; Vincent et al, 2005), we used Rosetta predictions because: (i) Rosetta is a well known *de novo* prediction program that has ranked first in successive CASP rounds (Jauch et al, 2007; Vincent et al, 2005); (ii) many Rosetta predictions are available at the server of Baker's group, thus allowing a consistent test of our approach, with predictions from the same source; and (iii) the program is available for interested users (http://www.rosettacommons.org/software/).

We downloaded the protein structure predictions from the server of Baker's laboratory (http://depts.washington.edu/bakerpg/drupal/). This set was constituted by 999 *de novo* models generated with Rosetta(Simons et al, 1997) for 85 proteins, i.e. a total of 84915 models. Application of the protocol studied here (Fig. 1) requires that the structure of a homolog of the target is available, and that the predictions used have the fold of the target.

The former was enforced by keeping only those proteins with a CATH relative at the T-level(Pearl et al, 2005) (homologs with the same fold of the target protein, regardless of their sequence similarity). The second condition was required to focus only on the local quality assessment problem, and was implemented by excluding those models not having the fold of their target protein. Technically, this meant that we only kept those models structurally similar to any member of the target's structural family: that is, those models giving a score higher than 5.25 for the model-homolog structure comparison done with MAMMOTH (Ortiz et al, 2002), for at least one homolog of the target protein. This step was computationally costly, as it involved 7,493,499 structure comparisons, and could only be carried using MAMMOTH(Ortiz et al, 2002); the 5.25 score threshold was taken from MAMMOTH's article (Ortiz et al, 2002). The final dataset was constituted by 68 target proteins and a total of 17180 models.

The properties of the selected target's residues (STR; to avoid meaningless results we only considered STR sets larger than 20 residues) were characterized with four parameters: two structure-based, and two sequence-based. The former were used to check if STR really were of better quality, comparing their parameters' values with those obtained for the set of all the target residues (ATR), i.e. the whole model structure. It has to be noted that: (i) STR and ATR sets are constituted by residues from the target protein, more precisely STR is a subset of ATR; and (ii) three possible STR sets were produced, because we checked our procedure using three structure comparison methods (MAMMOTH (Ortiz et al, 2002), SSAP(Orengo & Taylor, 1990) and LGA(Zemla, 2003)). The sequence-based properties were utilized to describe how STR spread along the sequence of the target, which helps to assess the usefulness of the protocol. Below we provide a brief description of each parameter, together with the results obtained from their use.

2.1.1 Structural quality: rmsd

Rmsd(Kabsch, 1976) is a quality measure widely employed to assess structure models: it corresponds to the average distance between model atoms and their equivalent in the native structure. Small rmsd values correspond to higher quality predictions than larger values.

In Fig. 2 we see the STR and ATR rmsd distributions. Regardless of the structure comparison method used (MAMMOTH (Ortiz et al, 2002), SSAP(Orengo & Taylor, 1990) and LGA(Zemla, 2003) in blue, yellow and red, respectively), STR distributions are shifted towards lower rmsd values relative to ATR distributions (in grey). This confirms the starting assumption: it shows that model residues structurally aligned to the protein's homolog usually have a higher structural quality. A consensus alignment (in black), which combined the results of the three structure comparison methods, gave better results at the price of including fewer residues; for this reason we excluded the consensus approach from subsequent analyses.

An interesting feature of STR rmsd distributions was that their maxima were between 3.5 Å and 6.5 Å, and that a majority of individual values were between 3 Å and 8 Å, and below 10 Å. To further explore this issue, we plotted the values of rmsd for STR against ATR (Fig. 3, grey boxes). In accordance with the histogram results, STR rmsd tended to be smaller than ATR rmsd. We distinguished two regions in the graph: in the first region (ATR rmsd between 0 Å and 6-8 Å) there was a roughly linear relationship between ATR and STR rmsds; however, for ATR rmsd values beyond 8 Å, STR rmsd reached a plateau. This plateau is at the origin of the thresholds observed in the histograms (Fig. 2), and confirms

that structure alignments can be used to identify subsets of model residues with better rmsd than the rest.
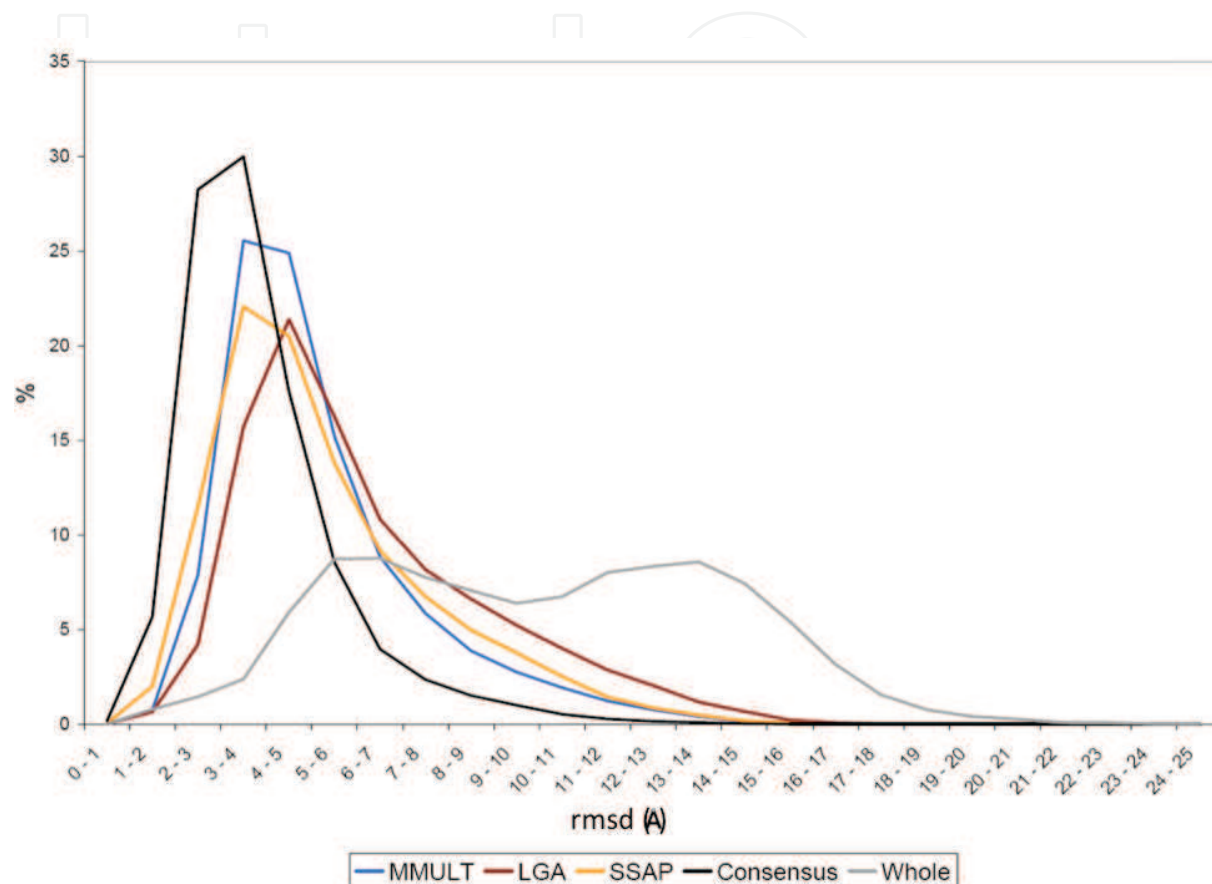

Fig. 2. Quality of structurally aligned regions vs. whole model, rmsd frequency histogram.

As a performance reference we used the PROSA program (Sippl, 1993) (white boxes) which provides a residue-by-residue, energy-based quality assessment, and is a single model method, therefore comparable to the approach presented here. PROSA was executed with default parameters, and we took as high quality residues those having energies below zero. In Fig. 3 we see that for good models, i.e. those with low ATR values, PROSA results (in white) were as good as those obtained with structure comparison methods (in grey). However, as models became poorer, PROSA results became worse, particularly after structure comparison methods reached their plateau. This indicates that when dealing with poor predictions use of structure alignments can improve/complement other quality assessment methods.
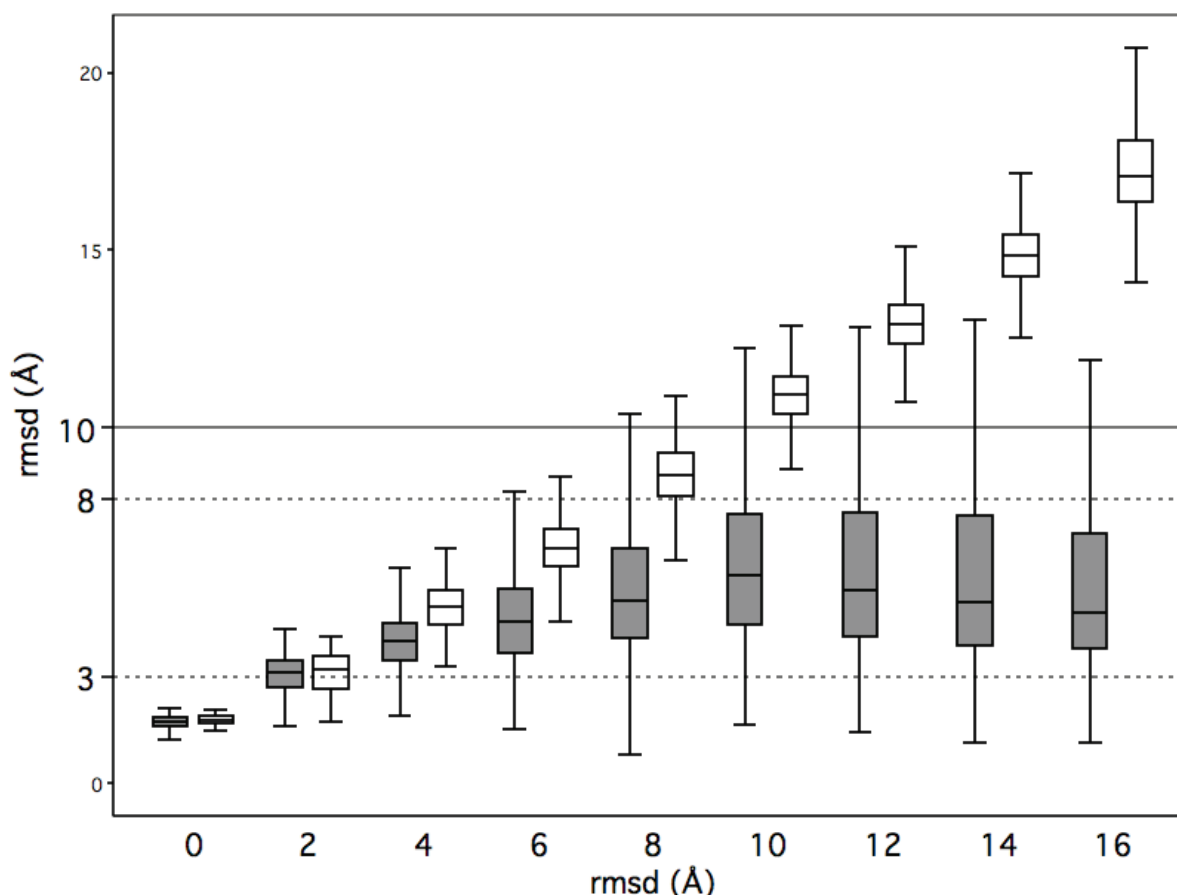
Fig. 3. Quality of structurally aligned (obtained with MAMMOTH (Ortiz et al, 2002)) regions vs. whole model, rmsd of selected residues vs. all-residues. Grey: structure comparison-based protocol (Fig. 1); white: PROSA(Sippl, 1993) results.

2.1.2 Structural quality: GDT_TS

GDT_TS is a quality measure routinely utilised by evaluator teams in the CASP community experiment (Jauch et al, 2007; Vincent et al, 2005): it is equal to the average of the percentages of model residues at less than 1 Å, 2 Å, 4 Å and 8 Å from their location in the correct structure. It was computed following the procedure described by Zemla(Zemla, 2003), using Cα atoms to compute residue-residue distances. GDT_TS varies between 0 and 100, with values approaching 100 as models become better.

We found that STR GDT_TS was in general better than ATR GDT_TS (Fig. 4); this was particularly true when the latter was below 40-50. Overall, this shows that STR is enriched in good quality sub-structures relative to ATR, particularly for poor models.

Consistency with rmsd analysis was observed when comparing the performance of structure comparison-based quality assessment (in grey) with that of PROSA (in white): for good models (GDT_TS values above 60-70) both approaches had a similar behavior; however, as model quality decreased, use of structure alignments showed increasingly better performance than PROSA at pinpointing correct substructures.
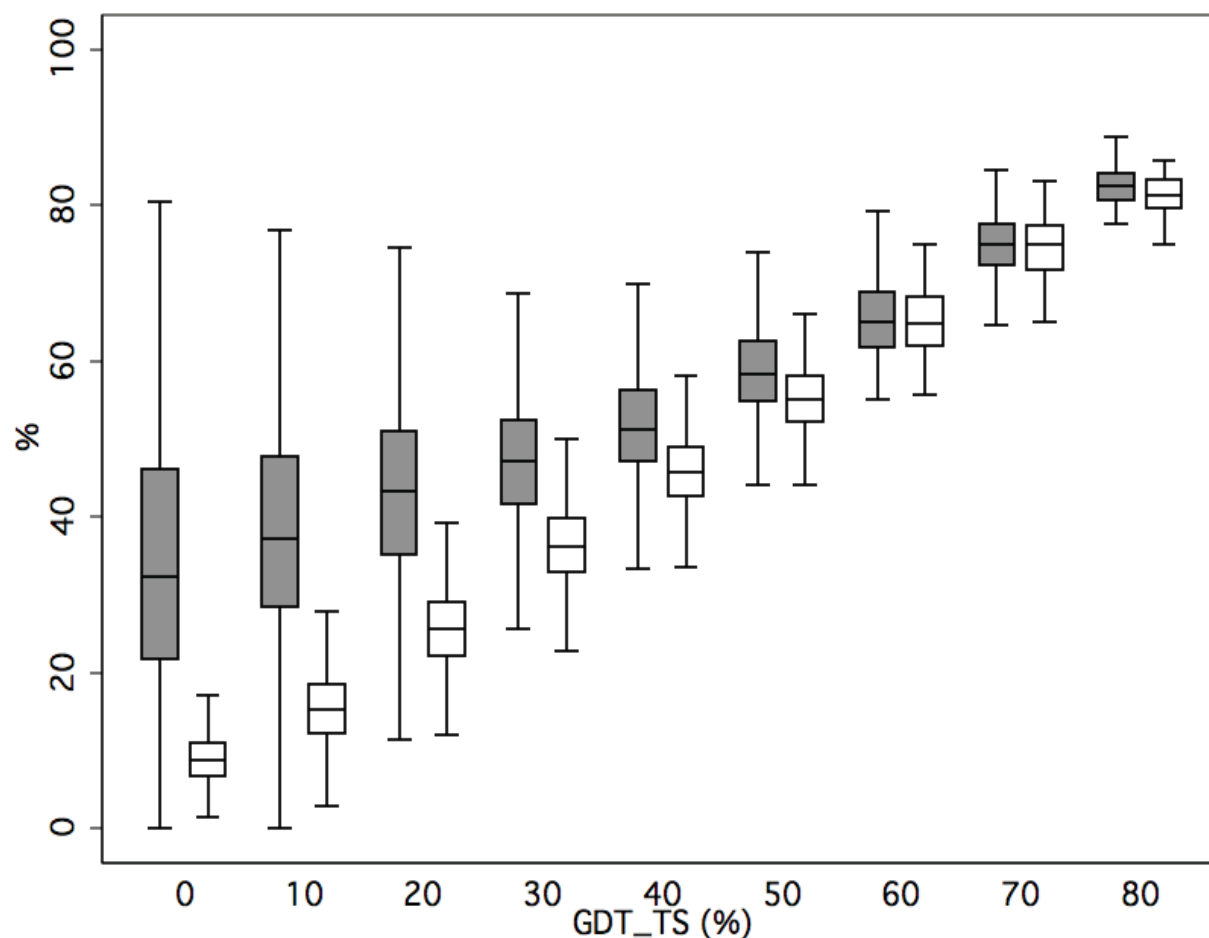
Fig. 4. Quality of structurally aligned (obtained with MAMMOTH (Ortiz et al, 2002)) regions vs. whole model, GDT_TS of the selected residues vs. all-residues GDT_TS. Grey: structure comparison-based protocol (Fig. 1); white: PROSA(Sippl, 1993) results.

2.1.3 Distribution of high quality residues along the protein sequence

Usually, STR do not form a continuous block, they tend to scatter along the sequence. The nature of this distribution is of interest for some applications of quality assessment methods (like model refinement) for which STR sets may be of little value if the involved residues are either too close in sequence, or contain too many orphan residues.

To characterize the distribution of STR along the sequence we used two measures: maximum distance (MD) between STR runs and normalized size distribution of STR runs (SAS). Both are based on the fact that, for a given model, STR sets are constituted by residue runs of varying size. MD corresponds to the largest sequence distance between STR runs (i.e. the number of residues between the rightmost and leftmost STR runs), divided by whole sequence length. MD values near 1 indicate that STR runs are spread over the whole protein, while smaller values point to a tighter residue clustering. SAS corresponds to the normalized (again by whole sequence length) size distribution for all runs constituting STR sets. SAS gives a view of how the sequence coverage is done: either by large sequence chunks, by small residue clusters, or by a mixture of both. When the alignment is constituted by small, evenly distributed residue clusters the SAS distribution will approach zero.

Our results showed that MD values are more frequent above 0.5, and more than 50% of them were higher than 0.8 (Fig. 5). The three structure comparison methods showed similar distributions, although LGA was slightly nearer to 1. This indicates that STR spread over a substantial part of the predicted protein.
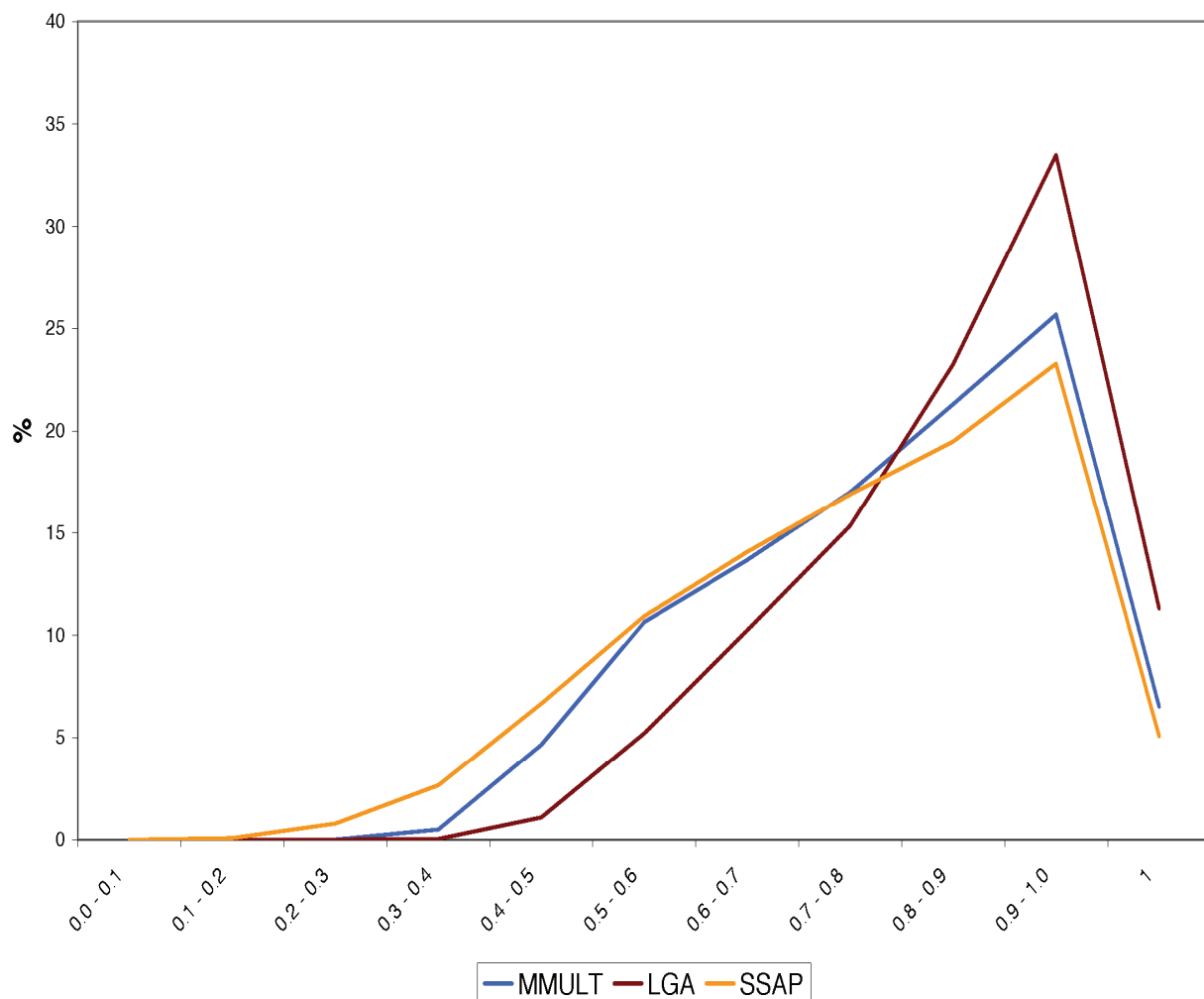

Fig. 5. Frequency distribution of the selected residues along the target sequence: normalized maximum distance between STR runs (unitless parameter).

Results for SAS (Fig. 6) showed that while ~50 % of STR formed clusters of size lower than 10 % of the whole sequence (i.e. SAS values below 0.1), the remaining residues were grouped in medium to large stretches. This means that for a 100 residue protein, clusters of more than 10 residues (which is roughly the size of an average α-helix) are frequent. In addition, for 95 % of the cases, the largest run of adjacent residues was above 30 % of the target length.

The picture arising from MD and SAS distributions is that STR usually extend over the protein length. Although STR sets are constituted by somewhat heterogeneous runs they do not contain too many orphan residues, as they include one large run (the size of a supersecondary structure motif, or larger) and several, smaller runs (the size of secondary structure elements).
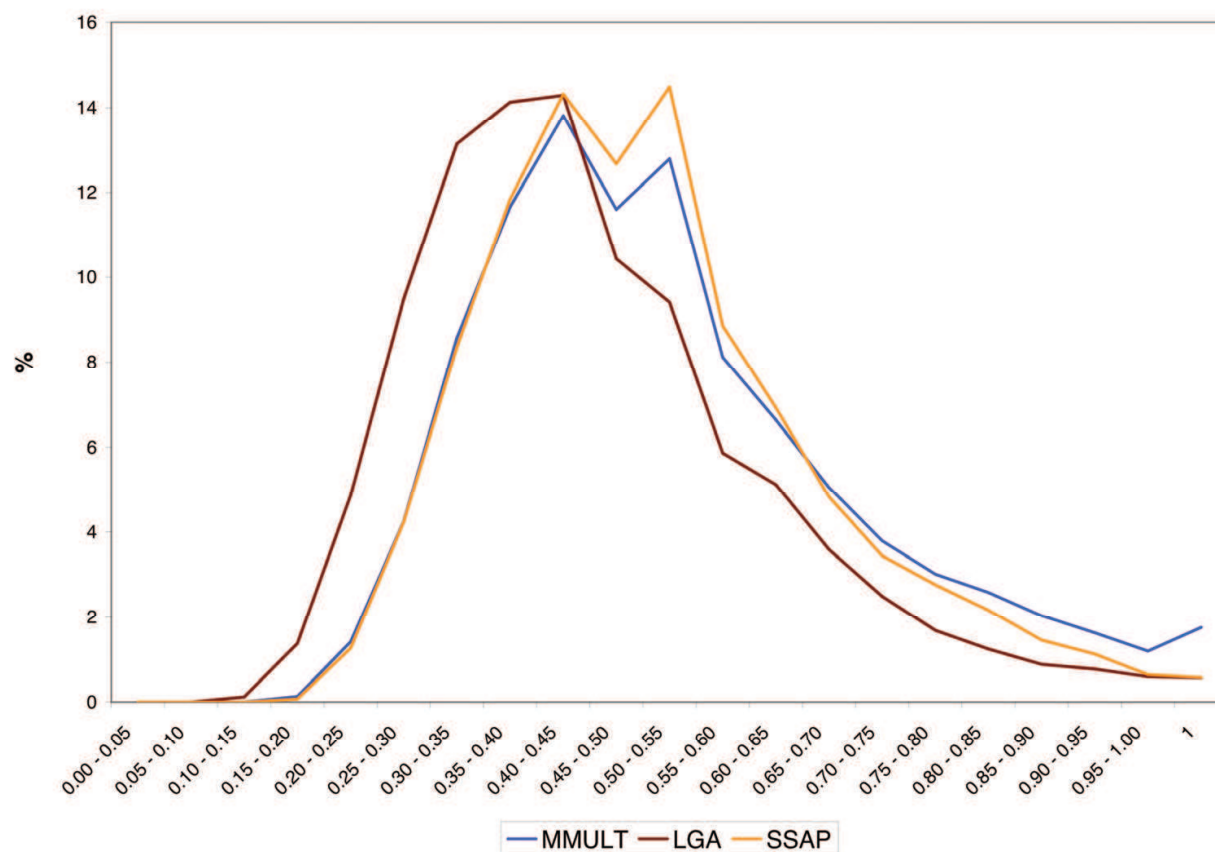
Fig. 6. Frequency distribution of the selected residues along the target sequence: normalized maximum distance between STR runs (unitless parameter).

3. Applicability range of structure comparison methods in local quality assessment

The approach described here is easy to use and has few computational requirements; however, it cannot be arbitrarily applied to any model or in any prediction scenario. In this section we describe which are its limits regarding prediction methods, target proteins and protein model nature.

3.1 Prediction methods

As far as the target protein has a homolog of known structure, model-homolog structure alignments can be computed and the quality assessment protocol (Fig. 1) can be applied, regardless of the prediction method originating the model. However, the approach presented here reaches its maximum utility when models are obtained with *de novo* structure prediction methods (methods originally devised to work using only the target's sequence and a physico-chemical/statistical potential, irrespective of the availability of homologs). This may seem somewhat contradictory, as one can think that the existence of target's homologs favors the use of comparative modeling methods instead of *de novo* methods. However, this is not the case: while *de novo* methods were initially developed with the *de novo* scenario in mind (only sequence information is available for the target protein), this situation is changing rapidly. Actually, when prediction problems become difficult, or a

given method gives an unclear answer, using more than one technique is considered a good approach within the prediction community, as the evaluators of the *de novo* section in the CASP6 experiment explain (Vincent et al, 2005): "Many predicting groups now use both de novo and homology modeling/fold recognition techniques to predict structures in all categories". In addition, it has been shown that de novo methods can compete with template-based methods in the prediction of difficult targets (Jauch et al, 2007; Raman et al, 2009; Vincent et al, 2005). In this situation, which implies the existence of target's homologs, our method can be used to score the local quality of *de novo* predictions.

In addition, a completely new field of application for *de novo* methods has been unveiled by the growing interest in knowing the structure of alternative splicing isoforms (C. Lee & Wang, 2005). Due to the very localized nature of sequence changes (Talavera et al, 2007), structure prediction of alternative splicing variants seems a trivial exercise in comparative modeling. However, template-based methods fail to reproduce the structure changes introduced by alternative splicing (Davletov & Jimenez, 2004). *De novo* approaches with their ability to combine first principles with deep conformational searches are ideal candidates to tackle this problem; in this case, availability of the structure of only one isoform would allow the application of our method.

3.2 Target proteins
Proteins to which our approach can be applied must have a homolog of known structure. The number of these proteins is increasing due to: (i) the progress of structural genomics projects (Todd et al, 2005) (this will increase the number of both easy/medium and hard targets); (ii) the growing number of alternative splicing variants of unknown structure (C. Lee & Wang, 2005).

3.3 Protein models
The approach proposed (Fig. 1) is a local, not a global, quality assessment method and should only be applied to models that have the native fold of the target (see above). Present *de novo* methods still cannot consistently produce models with a native-like fold (Moult et al, 2009). Therefore, researchers must ascertain that the model's fold is correct (irrespective of its resolution). This can be done using global quality assessment methods like PROSA (Sippl, 1993), the Elofsson's suite of programs (Wallner & Elofsson, 2007), etc.

4. Applications

Once available, local quality information can be used with different purposes. For example, it may help to identify those parts of a theoretical model that are more reliable for mutant design, or to interpret the results of mutagenesis experiments; it may be used for *in sillico* docking experiments involving *de novo* models, to decide which parts of the models must be employed preferentially; etc. One the most promising applications of quality assessment methods is the refinement of low-resolution models (Wallner & Elofsson, 2007). In this section we illustrate how the results of the procedure here described can be used for this purpose.

Among the possible options available for model refinement, we propose to use the alignment resulting from the structural superimposition between a *de novo* model and the target's homolog (Fig. 1) as input to a comparative modeling program. We applied this

strategy to 15 proteins (five from each of the three main CATH structural classes: alpha, beta and alpha/beta) from our initial dataset. These 15 proteins contributed a total of 2693 *de novo* models that resulted in 8033 model-homolog alignments (obtained with MAMMOTH (Ortiz et al, 2002)). These alignments were subsequently used as input to the standard homology modeling program MODELLER (Marti-Renom et al, 2000), which was run with default parameters. For the aligned regions we found (Fig. 7) that most of the refined models had lower model-native rmsd than the starting *de novo* models, i.e. they were closer to the native structure. A similar, although milder, trend was also observed when considering the whole set of protein residues (i.e. aligned as well as unaligned residues) (Fig. 8). These results show that this simple, computationally cheap model refinement protocol, based on the use of structure comparison local quality analysis, clearly helps to refine/improve low-resolution *de novo* models to an accuracy determined by the closest homolog of the target.
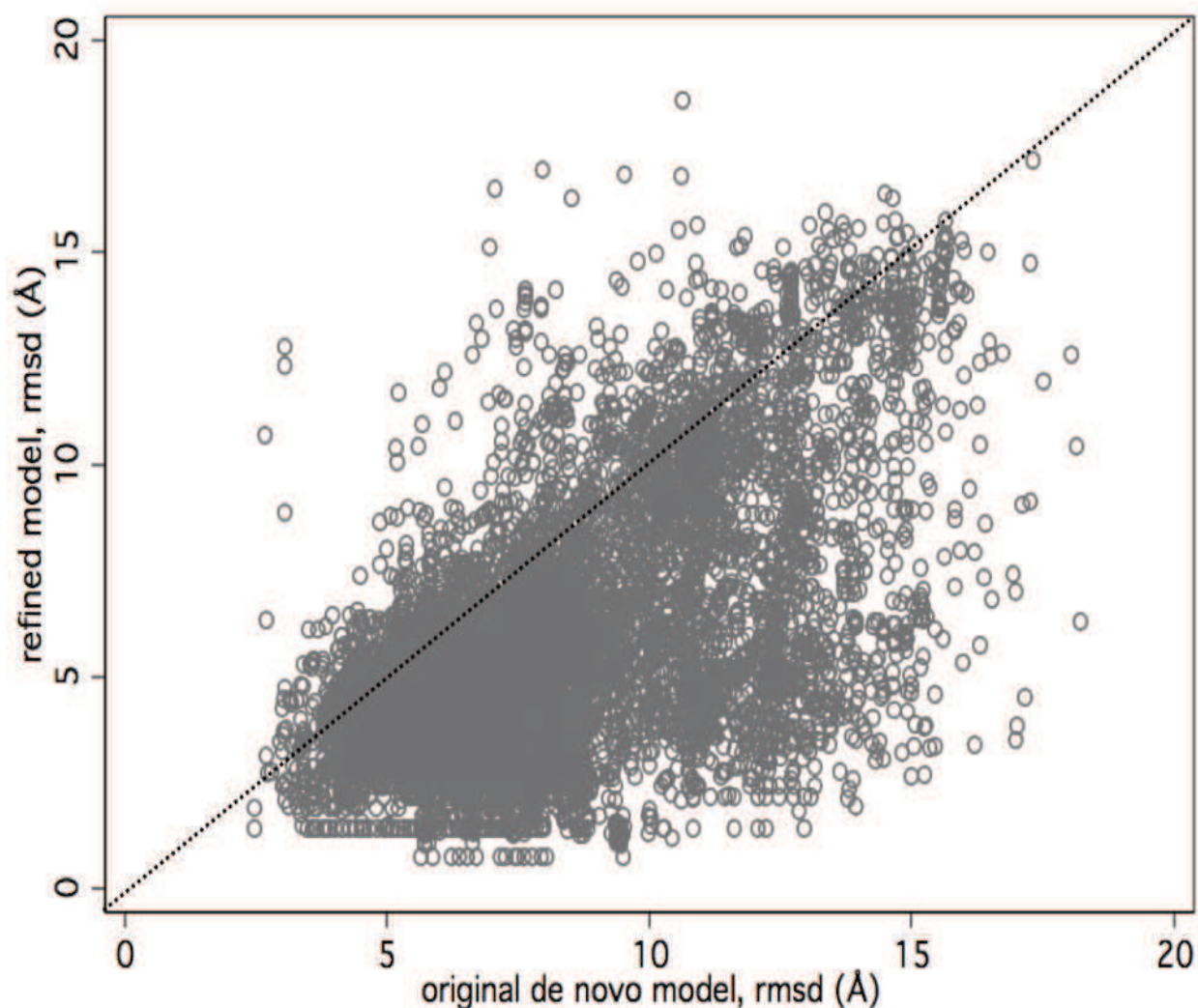
Fig. 7. Model refinement using structure comparison-based local quality assessment: rmsd of refined models vs. rmsd of original *de novo* models, subset of aligned residues. Points below the dotted line correspond to refinement-improved models.
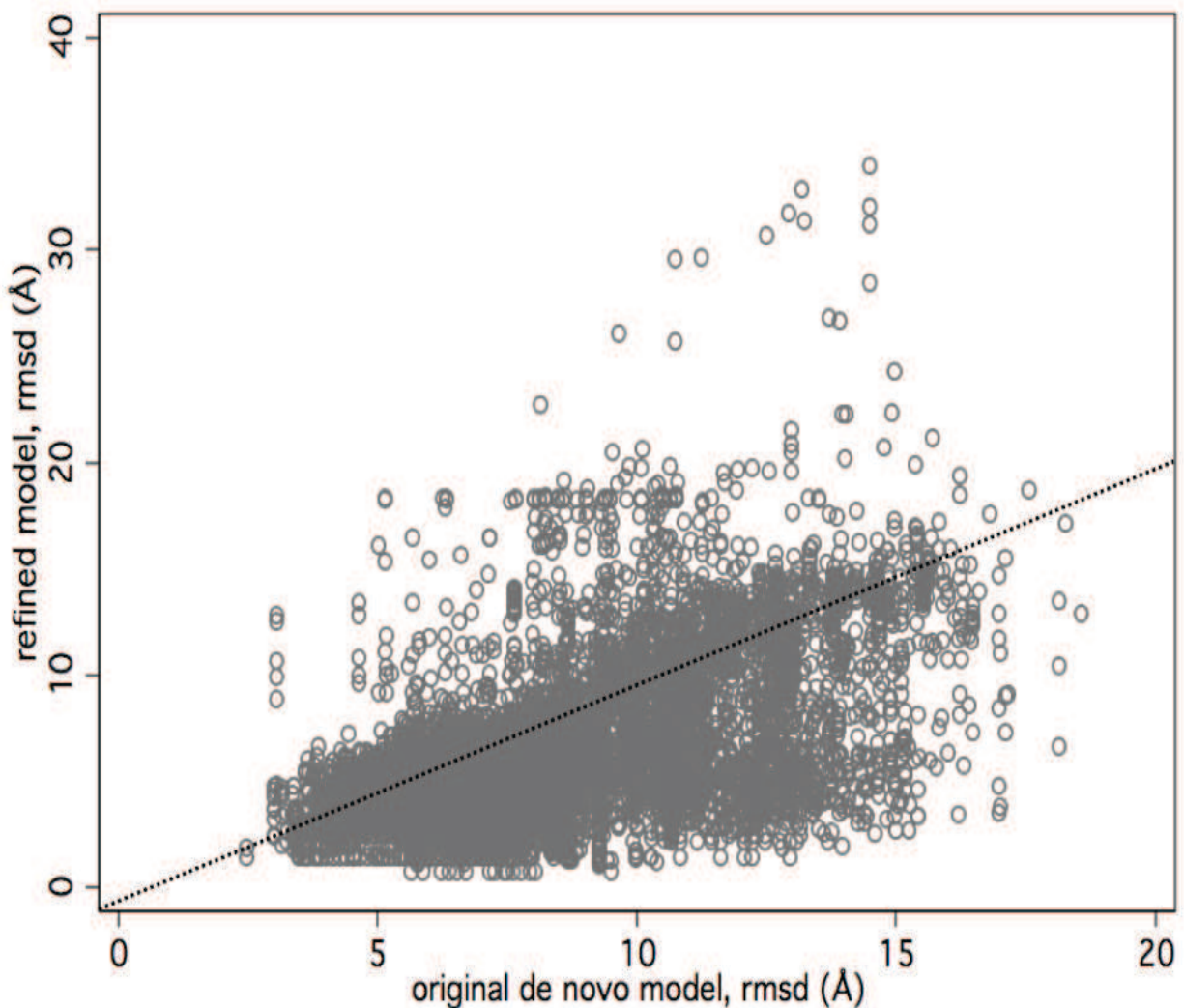
Fig. 8. Model refinement using structure comparison-based local quality assessment: rmsd of refined models vs. rmsd of original *de novo* models, all protein residues. Points below the dotted line correspond to refinement-improved models.

5. Conclusions

In this chapter we have described and tested a protocol for local quality assessment of low-resolution predictions based on the use of structure comparison methods. The testing was carried with *de novo* predictions, and the results showed that structure comparison methods allow the partitioning of the model's residues in two sets of high and low quality, respectively. This result holds even when only remote homologs of the target protein are available. The simplicity of the approach leaves room for future improvements and fruitful combination with other quality assessment methods. Two conditions determine the application range of this approach: the target protein must have at least one homolog of known structure, and models reproducing the fold of the target are required. However, results indicating that we may be near a full coverage of the proteins' fold space, together with advances in overall quality scoring indicate that these two problems are likely to become minor issues in the near future. Finally, our procedure suggests a simple refinement

strategy based on the use of comparative modeling programs that may be used to improve low-resolution *de novo* models.

6. Acknowledgment

7. References

Archie, J.G.; Paluszewski M. & Karplus K. (2009) Applying Undertaker to quality assessment. Proteins Vol.77 Suppl 9191-195, ISSN 0887-3585.

Benkert, P.; Tosatto S.C. & Schwede T. (2009) Global and local model quality estimation at CASP8 using the scoring functions QMEAN and QMEANclust. Proteins Vol.77 Suppl 9173-180, ISSN 0887-3585.

Bonneau, R.; Strauss C.E.; Rohl C.A.; Chivian D.; Bradley P.; Malmstrom L.; Robertson T. & Baker D. (2002) De novo prediction of three-dimensional structures for major protein families. J Mol Biol Vol.322, No.1, pp. 65-78, ISSN 0022-2836.

Cheng, J.; Wang Z.; Tegge A.N. & Eickholt J. (2009) Prediction of global and local quality of CASP8 models by MULTICOM series. Proteins Vol.77 Suppl 9181-184, ISSN 0887-3585.

Chou, K.C.; Nemethy G. & Scheraga H.A. (1983) Role of interchain interactions in the stabilization of the right-handed twist of beta-sheets. J Mol Biol Vol.168, No.2, pp. 389-407, ISSN 0022-2836.

Cozzetto, D.; Kryshtafovych A.; Ceriani M. & Tramontano A. (2007) Assessment of predictions in the model quality assessment category. Proteins Vol.69 Suppl 8175-183, ISSN 0887-3585.

Cozzetto, D.; Kryshtafovych A. & Tramontano A. (2009) Evaluation of CASP8 model quality predictions. Proteins Vol.77 Suppl 9157-166, ISSN 0887-3585.

Davletov, B. & Jimenez J.L. (2004) Sculpting a domain by splicing. Nat Struct Mol Biol Vol.11, No.1, pp. 4-5,

de la Cruz, X.; Sillitoe I. & Orengo C. (2002) Use of structure comparison methods for the refinement of protein structure predictions. I. Identifying the structural family of a protein from low-resolution models. Proteins Vol.46, No.1, pp. 72-84, ISSN 0887-3585.

Greene, L.H.; Lewis T.E.; Addou S.; Cuff A.; Dallman T.; Dibley M.; Redfern O.; Pearl F.; Nambudiry R.; Reid A.; Sillitoe I.; Yeats C.; Thornton J.M. & Orengo C.A. (2007) The CATH domain structure database: new protocols and classification levels give a more comprehensive resource for exploring evolution. Nucleic Acids Res Vol.35, No.Database issue, pp. D291-297, ISSN 0305-1048.

Hasegawa, H. & Holm L. (2009) Advances and pitfalls of protein structural alignment. Curr Opin Struct Biol Vol.19, No.3, pp. 341-348, ISSN 0959-440X.

Hooft, R.W.; Sander C. & Vriend G. (1997) Objectively judging the quality of a protein structure from a Ramachandran plot. Comput Appl Biosci Vol.13, No.4, pp. 425-430, ISSN 0266-7061.

Jauch, R.; Yeo H.C.; Kolatkar P.R. & Clarke N.D. (2007) Assessment of CASP7 structure predictions for template free targets. Proteins Vol.69 Suppl 857-67, ISSN 0887-3585.

Jones, D.T. & Thornton J.M. (1996) Potential energy functions for threading. Curr Opin Struct Biol Vol.6, No.2, pp. 210-216, ISSN 0959-440X.

Kabsch, W.A. (1976) A solution for the best rotation to relate two sets of vectors. Acta Crystallogr A Vol.A32922-923,

Kleywegt, G.J. (2000) Validation of protein crystal structures. Acta Crystallogr D Biol Crystallogr Vol.56, No.Pt 3, pp. 249-265, ISSN 1399-0047.

Kryshtafovych, A.; Fidelis K. & Moult J. (2009) CASP8 results in context of previous experiments. Proteins Vol.77 Suppl 9217-228, ISSN 0887-3585.

Larsson, P.; Skwark M.J.; Wallner B. & Elofsson A. (2009) Assessment of global and local model quality in CASP8 using Pcons and ProQ. Proteins Vol.77 Suppl 9167-172, ISSN 0887-3585.

Laskowski, R.A.; MacArthur M.W.; Moss D.S. & Thornton J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr Vol.26283-291, ISSN 1600-5767.

Laskowski, R.A.; MacArthur M.W. & Thornton J.M. (1998) Validation of protein models derived from experiment. Curr Opin Struct Biol Vol.8, No.5, pp. 631-639, ISSN 0959-440X.

Lazaridis, T. & Karplus M. (2000) Effective energy functions for protein structure prediction. Curr Opin Struct Biol Vol.10, No.2, pp. 139-145, ISSN 0959-440X.

Lee, C. & Wang Q. (2005) Bioinformatics analysis of alternative splicing. Brief Bioinform Vol.6, No.1, pp. 23-33,

Lee, D.A.; Rentzsch R. & Orengo C. (2010) GeMMA: functional subfamily classification within superfamilies of predicted protein structural domains. Nucleic Acids Res Vol.38, No.3, pp. 720-737, ISSN 0305-1048.

Lundstrom, J.; Rychlewski L.; Bujnicki J. & Elofsson A. (2001) Pcons: a neural-network-based consensus predictor that improves fold recognition. Protein Sci Vol.10, No.11, pp. 2354-2362,

Marti-Renom, M.A.; Stuart A.C.; Fiser A.; Sanchez R.; Melo F. & Sali A. (2000) Comparative protein structure modeling of genes and genomes. Annu Rev Biophys Biomol Struct Vol.29291-325, ISSN 1056-8700.

McGuffin, L.J. (2009) Prediction of global and local model quality in CASP8 using the ModFOLD server. Proteins Vol.77 Suppl 9185-190, ISSN 0887-3585.

Mereghetti, P.; Ganadu M.L.; Papaleo E.; Fantucci P. & De Gioia L. (2008) Validation of protein models by a neural network approach. BMC Bioinformatics Vol.966, ISSN 1471-2105.

Moult, J.; Fidelis K.; Kryshtafovych A.; Rost B. & Tramontano A. (2009) Critical assessment of methods of protein structure prediction - Round VIII. Proteins Vol.77 Suppl 91-4, ISSN 0887-3585.

Orengo, C.A. & Taylor W.R. (1990) A rapid method of protein structure alignment. J Theor Biol Vol.147, No.4, pp. 517-551,

Ortiz, A.R.; Strauss C.E. & Olmea O. (2002) MAMMOTH (matching molecular models obtained from theory): an automated method for model comparison. Protein Sci Vol.11, No.11, pp. 2606-2621,

Pearl, F.; Todd A.; Sillitoe I.; Dibley M.; Redfern O.; Lewis T.; Bennett C.; Marsden R.; Grant A.; Lee D.; Akpor A.; Maibaum M.; Harrison A.; Dallman T.; Reeves G.; Diboun I.; Addou S.; Lise S.; Johnston C.; Sillero A.; Thornton J. & Orengo C. (2005) The CATH Domain Structure Database and related resources Gene3D and DHS provide comprehensive domain family information for genome analysis. Nucleic Acids Res Vol.33, No.Database issue, pp. D247-251,

Raman, S.; Vernon R.; Thompson J.; Tyka M.; Sadreyev R.; Pei J.; Kim D.; Kellogg E.; DiMaio F.; Lange O.; Kinch L.; Sheffler W.; Kim B.H.; Das R.; Grishin N.V. & Baker D. (2009) Structure prediction for CASP8 with all-atom refinement using Rosetta. Proteins Vol.77 Suppl 989-99, ISSN 0887-3585.

Simons, K.T.; Kooperberg C.; Huang E. & Baker D. (1997) Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. J Mol Biol Vol.268, No.1, pp. 209-225,

Sippl, M.J. (1993) Recognition of errors in three-dimensional structures of proteins. Proteins Vol.17, No.4, pp. 355-362, ISSN 0887-3585.

Sippl, M.J. (1995) Knowledge-based potentials for proteins. Curr Opin Struct Biol Vol.5, No.2, pp. 229-235, ISSN 0959-440X.

Talavera, D.; Vogel C.; Orozco M.; Teichmann S.A. & de la Cruz X. (2007) The (in)dependence of alternative splicing and gene duplication. PLoS Comput Biol Vol.3, No.3, pp. e33,

Todd, A.E.; Marsden R.L.; Thornton J.M. & Orengo C.A. (2005) Progress of structural genomics initiatives: an analysis of solved target structures. J Mol Biol Vol.348, No.5, pp. 1235-1260, ISSN 0022-2836.

Venclovas, C. (2001) Comparative modeling of CASP4 target proteins: combining results of sequence search with three-dimensional structure assessment. Proteins Vol.Suppl 547-54, ISSN 0887-3585.

Venclovas, C.; Zemla A.; Fidelis K. & Moult J. (2001) Comparison of performance in successive CASP experiments. Proteins Vol.Suppl 5163-170, ISSN 0887-3585.

Vincent, J.J.; Tai C.H.; Sathyanarayana B.K. & Lee B. (2005) Assessment of CASP6 predictions for new and nearly new fold targets. Proteins Vol.61 Suppl 767-83, ISSN 0887-3585.

Vriend, G. (1990) WHAT IF: a molecular modeling and drug design program. J Mol Graph Vol.8, No.1, pp. 52-56, 29, ISSN 0263-7855.

Wallner, B. & Elofsson A. (2003) Can correct protein models be identified? Protein Sci Vol.12, No.5, pp. 1073-1086, ISSN 0961-8368.

Wallner, B. & Elofsson A. (2005) Pcons5: combining consensus, structural evaluation and fold recognition scores. Bioinformatics Vol.21, No.23, pp. 4248-4254, ISSN 1367-4803.

Wallner, B. & Elofsson A. (2006) Identification of correct regions in protein models using structural, alignment, and consensus information. Protein Sci Vol.15, No.4, pp. 900-913, ISSN 0961-8368.

Wallner, B. & Elofsson A. (2007) Prediction of global and local model quality in CASP7 using Pcons and ProQ. Proteins Vol.69 Suppl 8184-193, ISSN 0887-3585.

Wang, G.; Jin Y. & Dunbrack R.L., Jr. (2005) Assessment of fold recognition predictions in CASP6. Proteins Vol.61 Suppl 746-66, ISSN 0887-3585.

Wang, Z.; Tegge A.N. & Cheng J. (2009) Evaluating the absolute quality of a single protein model using structural features and support vector machines. Proteins Vol.75, No.3, pp. 638-647, ISSN 0887-3585.

Williams, M.G.; Shirai H.; Shi J.; Nagendra H.G.; Mueller J.; Mizuguchi K.; Miguel R.N.; Lovell S.C.; Innis C.A.; Deane C.M.; Chen L.; Campillo N.; Burke D.F.; Blundell T.L. & de Bakker P.I. (2001) Sequence-structure homology recognition by iterative alignment refinement and comparative modeling. Proteins Vol.Suppl 592-97, ISSN 0887-3585.

Zemla, A. (2003) LGA: A method for finding 3D similarities in protein structures. Nucleic Acids Res Vol.31, No.13, pp. 3370-3374,

Zhou, H. & Skolnick J. (2008) Protein model quality assessment prediction by combining fragment comparisons and a consensus C(alpha) contact potential. Proteins Vol.71, No.3, pp. 1211-1218, ISSN 0887-3585.
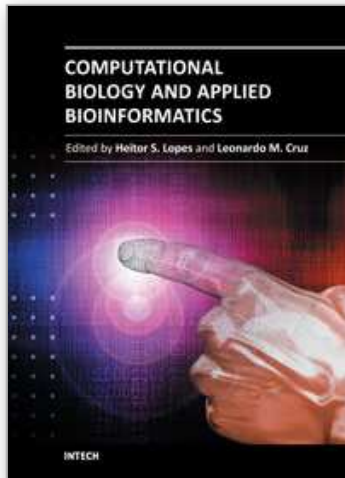
Computational Biology and Applied Bioinformatics

Edited by Prof. Heitor Lopes

ISBN 978-953-307-629-4

Hard cover, 442 pages

Publisher InTech

Published online 02, September, 2011

Published in print edition September, 2011

Nowadays it is difficult to imagine an area of knowledge that can continue developing without the use of computers and informatics. It is not different with biology, that has seen an unpredictable growth in recent decades, with the rise of a new discipline, bioinformatics, bringing together molecular biology, biotechnology and information technology. More recently, the development of high throughput techniques, such as microarray, mass spectrometry and DNA sequencing, has increased the need of computational support to collect, store, retrieve, analyze, and correlate huge data sets of complex information. On the other hand, the growth of the computational power for processing and storage has also increased the necessity for deeper knowledge in the field. The development of bioinformatics has allowed now the emergence of systems biology, the study of the interactions between the components of a biological system, and how these interactions give rise to the function and behavior of a living being. This book presents some theoretical issues, reviews, and a variety of bioinformatics applications. For better understanding, the chapters were grouped in two parts. In Part I, the chapters are more oriented towards literature review and theoretical issues. Part II consists of application-oriented chapters that report case studies in which a specific biological problem is treated with bioinformatics tools.

INTECH
open science | open minds