BIOINFORMATICS ORIGINAL PAPER

Vol. 28 no. 9 2012, pages 1209–1215
doi:10.1093/bioinformatics/bts103

Structural bioinformatics

Advance Access publication March 6, 2012

Evolutionary inaccuracy of pairwise structural alignments

M. I. Sadowski* and W. R. Taylor

Division of Mathematical Biology, MRC National Institute for Medical Research, The Ridgeway, Mill Hill, London NW71AA, UK

Associate Editor: Alfonso Valencia

ABSTRACT

Motivation: Structural alignment methods are widely used to generate gold standard alignments for improving multiple sequence alignments and transferring functional annotations, as well as for assigning structural distances between proteins. However, the correctness of the alignments generated by these methods is difficult to assess objectively since little is known about the exact evolutionary history of most proteins. Since homology is an equivalence relation, an upper bound on alignment quality can be found by assessing the consistency of alignments. Measuring the consistency of current methods of structure alignment and determining the causes of inconsistencies can, therefore, provide information on the quality of current methods and suggest possibilities for further improvement.

Results: We analyze the self-consistency of seven widely-used structural alignment methods (SAP, TM-align, Fr-TM-align, MAMMOTH, DALI, CE and FATCAT) on a diverse, non-redundant set of 1863 domains from the SCOP database and demonstrate that even for relatively similar proteins the degree of inconsistency of the alignments on a residue level is high (30%). We further show that levels of consistency vary substantially between methods, with two methods (SAP and Fr-TM-align) producing more consistent alignments than the rest. Inconsistency is found to be higher near gaps and for proteins of low structural complexity, as well as for helices. The ability of the methods to identify good structural alignments is also assessed using geometric measures, for which FATCAT (flexible mode) is found to be the best performer despite being highly inconsistent. We conclude that there is substantial scope for improving the consistency of structural alignment methods.

Contact: msadows@nimr.mrc.ac.uk

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on August 23, 2011; revised on January 13, 2012; accepted on February 24, 2012

1 INTRODUCTION

Despite its apparent simplicity, the problem of aligning pairs of protein structures has attracted a significant level of research effort. Methods vary in the details of their objective function, problem representation, null model of comparison statistics and approaches to searching alignment space (Alesker *et al.,* 1996; Birzele *et al*., 2007; Chen and Crippen, 2005; Holm and Sander, 1993; Kifer *et al.,* 2011; Kolodny *et al*., 2005; Lackner *et al*., 2000; Novosad *et al*., 2010; Pandit and Skolnick, 2008; Shibberu and Holder, 2011; Shindyalov and Bourne, 1998; Taylor, 1999; Vesterstrom and Taylor, 2006; Zhang and Skolnick, 2004), Reviewed in Carugo (2007). Common variations include the use of flexible alignment (Mosca *et al*., 2008; Rocha *et al.,* 2009; Salem *et al*., 2010; Shatsky *et al.,* 2004; Ye and Godzik, 2003) or using fragments and topological filters for initial alignments to improve quality and speed (Budowski-Tal *et al.,* 2010; Gibrat *et al*., 1996; Krissinel and Henrick, 2004; Veeramalai *et al*., 2009).

Two previous benchmarks of pairwise structure alignment methods have been published in the last decade (Kolodny *et al*., 2005; Mayr *et al*., 2007). These considered the degree to which the methods tested find a good solution as judged by geometric criteria (Kolodny *et al*., 2005) and the agreement of the aligned residues with a set of manually curated 'gold standard' alignments (Mayr *et al*., 2007). Both studies are important contributions but a recent study which covers current methods is lacking, as is a study which considers both geometric performance and ability to find homologous relationships between positions simultaneously.

We would ideally like to assess the ability of these aligners to find homologous relationships as well as geometric similarities for a large number of proteins but this is problematic since for distantly related proteins we rarely know how individual positions are related. As an alternative to using gold standards and limiting the size of the dataset, we propose to make use of the fundamental property that homology is transitive: if A and B are homologous, B and C are homologous then A and C must also be homologous. Homology, therefore, establishes a set of equivalence classes over the residues in sets of related protein structures (symmetry and self-identity being obvious properties) and the more closely a structural alignment method approaches this situation the better its performance. The exception to this occurs only if a set of residues are related by a star phylogeny—for example where a gene duplication has resulted in duplicate internal structures such as are found in repeat proteins (Taylor and Sadowski, 2010a).

In this study, we use this idea to compare the most widely-used methods for pairwise structural alignment, in addition to considering alignment accuracy relative to other annotation sources: DSSP structural classes (Kabsch and Sander, 1983) and solvent accessibilities. Additionally, following Kolodny *et al*. (2005) we consider the quality of the scores implemented by the methods with respect to external annotations [SCOP folds (Andreeva *et al*., 2008), GO annotations (Morais *et al*., 2011) and topological distances (Hollup *et al*. 2011; Sadowski and Taylor, 2010b)] and several geometric scores. Seven methods were chosen: the choice was based on their free availability for general academic use, their importance for publicly available resources or being widely used as judged by

*To whom correspondence should be addressed.

citation counts. We were interested in examining not only the relative performance of the methods as judged on several criteria but also the relationship between geometric accuracy and the accuracy of homology assignment.

We find that the different assessment methods highlight different strengths and weaknesses of each of the methods, although TM-align and its newer sibling Fr-TM-align generally perform very well overall. We note that for the FATCAT method, flexible alignment increases geometric accuracy at the expense of self-consistency. Finally, we observe a relationship between structural complexity and consistency of alignments, suggesting that care is needed when aligning repetitive structures.

2 METHODS

2.1 Data set

A set of 1863 domains was derived from the ASTRAL SCOP10 database (SCOP version 1.73; Brenner *et al*., 2000). The set was restricted to high quality structures by requiring a SPACI score >0.5 (roughly equivalent to requiring at least 2Å resolution) and excluding NMR structures and those with missing residues.

2.2 Structural alignment methods

All-versus-all pairwise structural alignments were generated using seven methods: SAP (Taylor, 1999), DALIlite (Holm and Sander, 1993), MAMMOTH-MULT (Ortiz *et al*., 2002), FATCAT (Ye and Godzik, 2003), CE (Shindyalov and Bourne, 1998), TM-align (Zhang and Skolnick, 2005) and Fr-TM-align (Pandit and Skolnick, 2008). FATCAT alignments were run both in flexible and non-flexible mode to permit a direct comparison, leading to eight methods in total. We selected this set of methods on the basis that they are easily available, well-established and in many cases are used to compute large sets of alignments for publicly available resources (e.g. FATCAT and CE for the PDB, DALI for the DALI FSSP database) or have been used to draw conclusions about fold-space (SAP, TM-align, DALI and MAMMOTH). All methods were used with default parameters. Andreas Prlic's Java implementations of FATCAT and CE were used.

2.3 Inconsistency measure

Inconsistency was assessed for all positions in any triplet of aligned structures where all pairwise alignments between the three were above a particular threshold. Where a gap was found at that position in any of the three alignment sequences, the position was ignored. For each position we determined whether the condition

$$E(A_i,B_j)\cap E(B_j,C_k)\cap E(A_i,C_k)$$

was true, where the predicate $E(X_i, Y_j)$ is defined as meaning position i in sequence X is aligned to position j in sequence Y. Since we are measuring inconsistency, we score 1 if the condition is false, 0 otherwise. The proportion of inconsistent positions was found for all aligned triples for each method at each threshold and calculated as a percentage. For analysis across all residues, we use the percentage as defined above (absolute inconsistency) whereas for subsets of residues with particular annotations we report values normalized as a percentage of the overall inconsistency for the same method across all residue types (relative inconsistency).

2.4 Calibration of data

The potential for inconsistencies in equivalences increases with structural distance, so it is necessary to compare levels of inconsistency between alignments of similar quality. We therefore sought to rank alignments as fairly as possible without deviating from the native scoring system unduly. To this end, the RMSD and coverage values were used to approximate TM-scores for the alignments generated by each method, we refer to the approximate TM scores as fTM scores. In each case a D_0 value was found using the following equation

$$D_0 = 1.24\sqrt[3]{L-15}-1.8$$

(Zhang and Skolnick, 2004), with L being the shorter of the lengths of the domains being compared. The approximate TM-score (fTM) was calculated as

$$\mathrm{fTM} = \frac{C}{L\left(1+\left(\frac{R}{D_0}\right)^2\right)}$$

with C being the number of aligned residues, L being the mean length of the structures and R the reported RMSD. This calculation is very simple and requires only an alignment length and RMSD as calculated by the method, which all of the methods tested here report.

Self-comparisons of approximate and real TM-scores for TM-align and Fr-TM-align produced correlations of 0.981 and 0.985, respectively, showing that the approximation is very good. Approximate TM-scores for the other methods were correlated with TM-align as follows: SAP 0.739, DALI 0.643, FATCAT 0.774, FATCAT (flexible mode) 0.639, CE 0.837 and Fr-TM-align 0.923.

We then compared the fTM scores with the methods own summary scores to determine which was likely to provide the best ranking. As a measure of correctness in ranking, we determined how well each score correctly identified pairs with the same SCOP fold in the SCOP10 dataset using a ROC statistic. The mean area under the ROC curve (AUC) was determined up to a 5% false positive rate for ten 50% partitions of the data in a bootstrap analysis to allow the significance of any differences to be assessed.

Having chosen the optimal score for each method by these means, we ranked alignments for each method into 15 bins from 99.9% down to 98.5% of alignments. From 2 115 472 pairs of proteins, 0.1% corresponds to ∼2000 alignments, thus we examined the top 30 000 alignments for each method in cumulative increments of 2000. Table S1 (Supplementary Material) details the exact threshold values used for each method at each increment.

2.5 Other geometric measures

Following Kolodny *et al*., (2005) we assessed geometric quality of reported alignments using the following measures:

$$\mathrm{SI} = \frac{R\times\min(L_1,L_2)}{C}$$

$$\mathrm{MI} = 1-\frac{C+1}{(1+\frac{R}{W_0})(1+\min(L_1,L_2)}$$

$$\mathrm{SAS} = \frac{R\times 100}{C}$$

with R = RMSD and C = alignment coverage as above, L_1 and $L_2$the lengths of the two sequences and W_0 a weighting parameter, here set to 1.5 as in Kolodny *et al*. (2005).

2.6 Residue annotations

The catalytic site atlas annotations (Porter *et al*., 2004) were combined with annotations from PDB SITE records to produce datasets of functional residues. Secondary structure assignments were taken from DSSP (Kabsch and Sander, 1983) using the original eight-state assignments without further smoothing. Accessibility values were also taken from DSSP and residues were binned into five equal-width bins of relative accessibility.

When assessing inconsistency for secondary structure or burial, we considered the annotated residue state to be the majority state of the three residues compared, thus BBE would map to B, HHC to H etc. In cases where no majority was found, the set was discarded. Assessment of the consistency of the annotations was assessed separately using chi-square tests (Supplementary Material). Distances from gaps were calculated as the minimum distance from gap for each of the three residues.

2.7 Assessment of symmetry

Symmetry values for protein structures were derived using the Fourier transform-based approach described by Taylor *et al.* (2002), using the power of the Fourier spectrum to measure symmetry of each structure. Inconsistency values per domain were the mean for all methods at the highest threshold (closest structures), which had 803 members; domains with fewer than 5 neighbours for TM-align were culled from the set, leaving 207 domains.

3 RESULTS

3.1 Choosing a score for ranking: ROC assessment

To compare methods we needed to rank their alignments meaningfully. However, since some methods produce multiple scores whereas others only report an RMSD and alignment length it is necessary to have an objective basis for ranking each method which does not penalize any method for producing a poor (or no) summary score.

To address this problem, the scores produced by the methods were compared with respect to their ability to find relationships classified by SCOP database at the fold level. Although SCOP tends to favour evolutionary relationships over strict structural similarities (Sadowski and Taylor, 2010b), the number of such overlaps in this dataset is small and does not significantly affect results for low error rates. Figure 1 shows mean AUC values for ROC curves derived from each possible score for the methods presented. AUCs are taken up to 5% false positive rate, 10 bootstrap replicates over samples of 50% of the data were generated. We tested the scores produced by the methods themselves in addition to the approximate TM-score (fTM) defined in methods to allow comparison with methods that do not produce a length-corrected summary.

The ROC results show that for most of the methods the approximate TM-scores (see Section 2) are significantly better than the method's own score. The exceptions to this are TM-align and MAMMOTH, although MAMMOTH does relatively poorly on this benchmark. CE, DALI, FATCAT (rigid mode) and Fr-TM-align all perform excellently when scored using the approximate TM-score, while MAMMOTH, FATCAT (flexible mode) and SAP all perform less well regardless of score. Interestingly the use of flexible alignment in this instance led to less accurate scoring. DALI's Z-score was found to perform poorly in this test, although the underlying coverage and RMSD values could be used to produce an accurate score. Additionally in many cases simply using a coverage score was sufficient to produce a good AUC value on this test. Using this test we chose the TM-score for TM-align, Z-score for MAMMOTH, *P*-value for FATCAT (flexible mode) and the fTM score for all other methods.

3.2 Assessment of self-consistency for structural aligners

The inconsistency of the top 1.5% of alignments for each method were assessed in increments of 0.1% (exact thresholds used are detailed in Supplementary Material, Table S1). Figure 2 shows a plot of the absolute inconsistencies for each of the 8 methods.

For similar proteins, SAP and DALI are the most consistent, however even for proteins with high similarity, the rate of inconsistency is high with ~20% of positions being inconsistently aligned. The degree of inconsistency increases to 50% for SAP

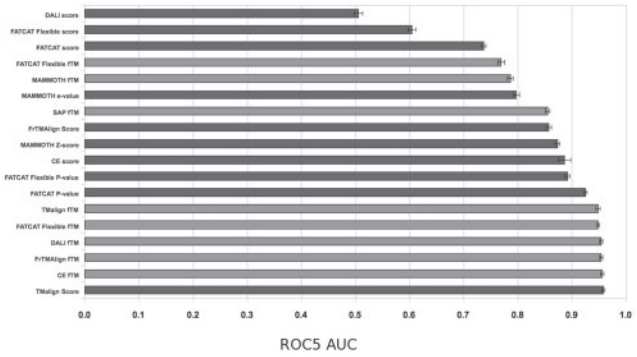

Fig. 1. ROC5 AUC values (mean of $n = 10$ replicates using 50% of the data) for native scores (blue) and approximate TM-scores (pink) for the eight methods. True positives were in the same SCOP fold as the query.

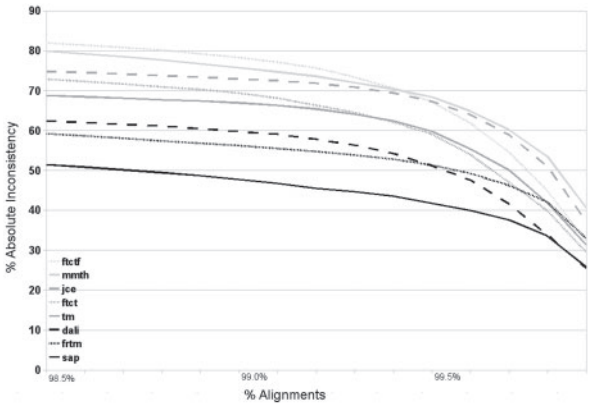

Fig. 2. Inconsistency of pairwise structural alignments. The proportion of positions failing transitive consistency is shown for all alignment pairs in the relevant fraction of the set. The methods appear in the order FATCAT-flexible, MAMMOTH, CE, FATCAT, TM-align, DALI, Fr-TM-align, SAP from top to bottom on the left-hand edge of the graph.

at greater distances, while for the least consistent method—MAMMOTH—the level of inconsistency reaches 85%. At functional residues the level of inconsistency is lower for more structurally similar proteins but reaches the same level as for all residues for proteins which are less similar. At the level of very low structural similarity (bottom 1.5% of alignments) we find that the degree of inconsistency is between 98% and 100% for all of the methods, showing that for unrelated proteins the expected level of correct homology assignment would be 0, as it should be.

From an evolutionary point of view, these inconsistencies must at some level represent errors since the homology relation is transitive. However from a structural point of view even with inconsistent alignments the set of residue equivalences is usually sufficient to produce a good score. More generally at larger evolutionary distances the strict transitivity of relationships might become unsatisfiable: if several residues are descended from a single ancestral residue then the strict homology relationship might be many-one (as has been proposed to have occurred multiple times for repeat proteins such as the beta propellors). In this case, the limitation is not the result of heuristics but simply the requirement

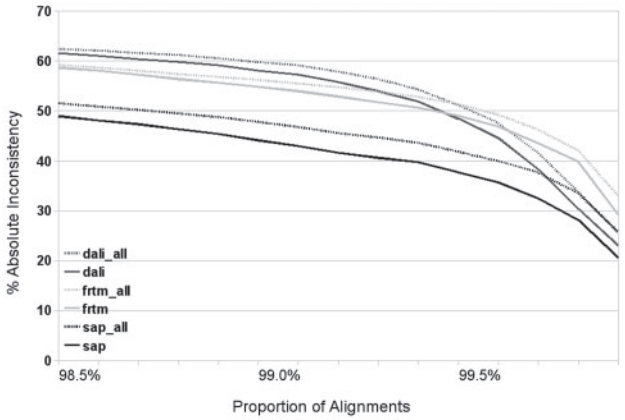

Fig. 3. Improved consistency at residues marked functional. Absolute rates of inconsistency are shown for functional residues (solid lines) and all residues (dashed lines) for the three most consistent methods. These appear in the order DALI, Fr-TM-align, SAP from the top downwards along the left-hand edge.

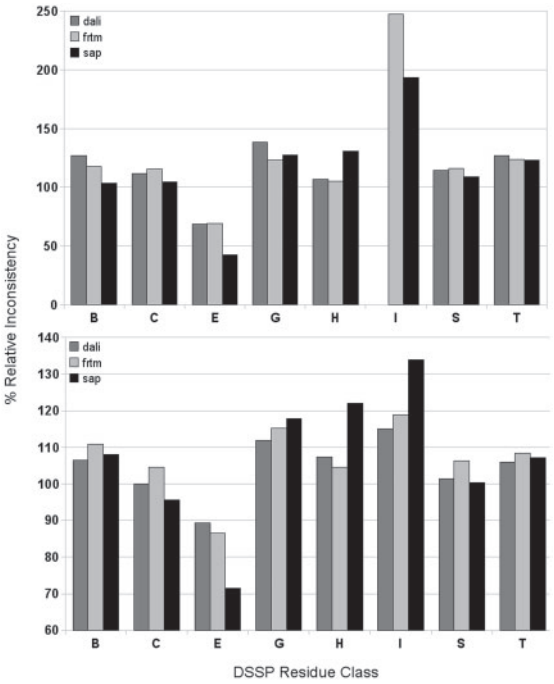

Fig. 4. Relative inconsistencies for DSSP residue classes. Inconsistencies are shown as a percentage of the absolute value for each method. The upper panel shows results for the top 0.01% of alignments, the bottom the top 1.5%.

that alignments are one–one. To shed some light on this question, we determined which domains and regions were most associated with inconsistent alignments.

3.3 Determining structural features associated with inconsistencies

Figures 3–6 show breakdowns of the degree of inconsistency for functional residues, residue classes, solvent accessibility classes and distance from gaps (see Section 2). For clarity only SAP, Fr-TM-align and DALI are shown in Figures 3–5. Each plot shows the results for the shortest and longest structural distances considered (99.5%, 98.5%).

All of the methods show greater consistency for functional residues than non-functional, even at greater distances, meaning that these assignments are a little more trustworthy than general alignments (Fig. 3). The difference is roughly constant at larger distances but generally greatest for most closely similar structures.

The patterns of relative consistency for residues in different secondary structure types is similar for the most consistent methods (Fig. 4) although it varies for other, less consistent, methods (not shown). In these cases beta strand residues are significantly more consistently aligned than the background, with coils typically being close to the background level and helical types (G, H, I) less consistent. For closely similar structures the difference between the most consistent (beta strands) and least consistent (π helices) is very large, as is the difference between these and the background level of consistency. At greater distances the differences are smaller, a feature seen generally for all the breakdowns.

Buried residues are also found to be more consistently aligned than exposed residues, with increasing inconsistency being associated with increasing solvent exposure (Fig. 5), a trend consistently repeated for all methods at all levels of structural distance. The strongest effect is seen for gap distance (Fig. 6). For all methods the level of inconsistency is highest for positions adjacent to gaps in at least one of the three sequences, with a geometric decline of inconsistency with gap distance.

We also determined whether rates of inconsistency differed for different structural types. Interestingly, there were large differences between structural types in the overall degree to which their positions were involved in inconsistent alignments (Table S2). Overall, domains in the alpha class contained many more inconsistently aligned residues than the other classes (mean 58%, SD 29% compared with mean 30%, SD 12% for class *c*, the next highest). However, the most inconsistently aligned domains were from the single-stranded beta helix fold, for which no residues were consistently aligned by any method. Alpha/alpha superhelices, alpha/alpha toroids and 7-bladed propellors also featured as highly inconsistent.

As might be expected given the above, the folds displaying the lowest level of inconsistency were less repetitive folds from class *d* (thioesterase, DNA clamp, CBS-domain pair) and beta barrel folds (SH3, PH-domains). Other complex folds such as the beta trefoil and double-stranded beta helix were also found to be much more consistently aligned.

To test the idea that fold complexity is related to the inconsistency of the alignments we plotted the total inconsistency against a Fourier-transform-based symmetry measure (Taylor *et al*., 2002). For all domains with >4 neighbours at the top threshold. This plot appears as Figure S1. The trend has a significant Spearman rank correlation of 0.53 ($P < 0.01$; $N = 210$), suggesting that there is a relationship between structural complexity and alignment consistency.

3.4 Assessment by geometric measures

It is apparent that the level of inconsistency in structural alignments varies substantially between methods and that it can be partially

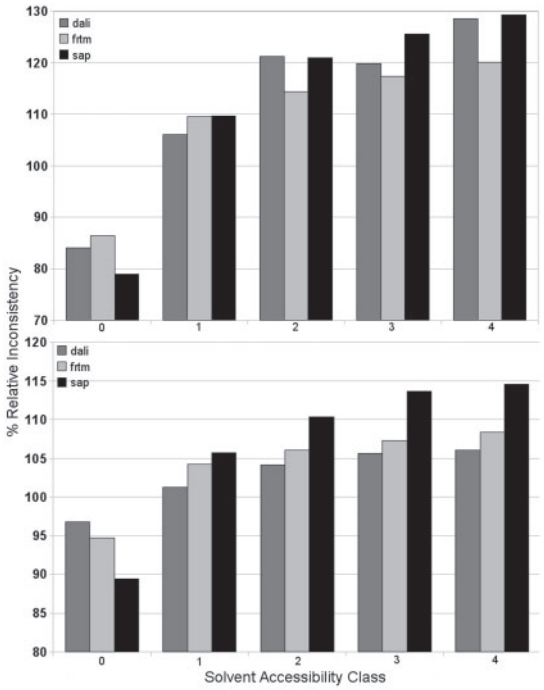

Fig. 5. Relative inconsistency for three methods in relation to solvent accessibility. Solvent accessibility was split into classes in bins of 20% with 0 being the lowest. Panels are arranged as in Figure 4.

Table 1. Geometric assessment of structural alignments

Method	fTM	SI	MI	SAS
2235				
TM-align	2.83 (2)	2.90 (1)	2.88 (1)	2.90 (1)
FATCAT (flexible)	2.77 (1)	3.18 (2)	3.10 (2)	3.18 (2)
Fr-TM-align	2.89 (3)	3.25 (3)	3.19 (3)	3.25 (3)
CE	4.28 (5)	3.91 (4)	3.93 (4)	3.91 (4)
FATCAT	4.08 (4)	4.28 (5)	4.14 (5)	4.28 (5)
DALI	5.17 (6)	4.46 (6)	4.63 (6)	4.46 (6)
MAMMOTH	6.94 (7)	6.86 (7)	7.11 (8)	6.86 (7)
SAP	7.03 (8)	7.15 (8)	7.03 (7)	7.15 (8)
114477				
FATCAT (flexible)	1.44 (1)	1.37 (1)	1.38 (1)	1.37 (1)
Fr-TM-align	2.39 (2)	2.66 (2)	2.54 (2)	2.66 (2)
TM-align	2.98 (3)	3.18 (3)	3.12 (3)	3.18 (3)
FATCAT	4.58 (5)	4.52 (4)	4.42 (4)	4.52 (4)
CE	4.45 (4)	4.58 (5)	4.54 (5)	4.58 (5)
DALI	6.13 (6)	5.80 (6)	6.05 (6)	5.80 (6)
MAMMOTH	6.86 (7)	6.82 (7)	7.25 (8)	6.82 (7)
SAP	7.17 (8)	7.06 (8)	6.71 (7)	7.06 (8)

The mean rank for each method by each score over the top 99.9% (upper) and 98.5% (lower) of alignments is shown. The number in the top left denotes the number of alignment in each set.

explained by the structural complexity of the fold. It is then interesting to consider how the level of inconsistency relates to more traditional methods for assessing structural alignments—first to put the results into context and second as a complementary method of assessment.

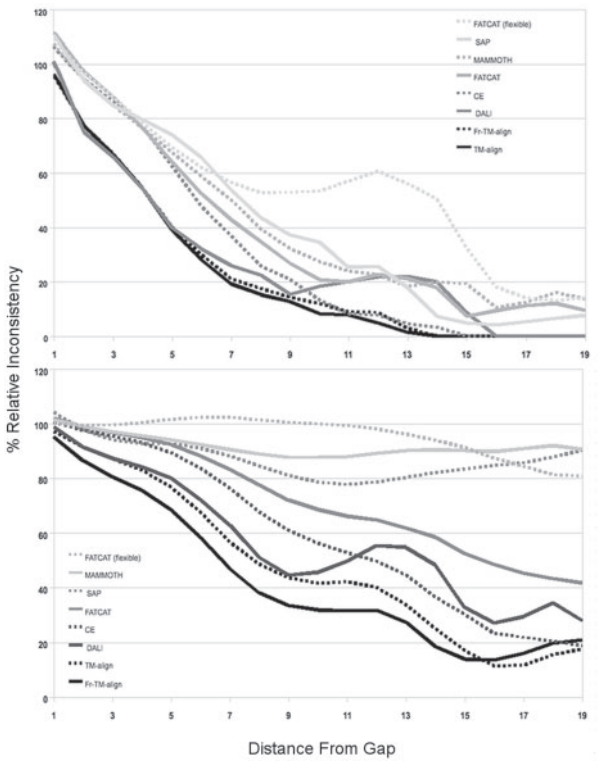

Fig. 6. Relative inconsistency as a function of gap distance. Panels are arranged as in Figure 4.

Following Kolodny *et al.* (2005), we determined several geometric measures of structural alignment quality: the SI, MI and SAS measures defined by Kolodny were all used to assess the eight methods, additionally, we considered the approximate TM-score (as defined in Section 2). Table 1 shows the results at low and high structural distance as judged by having a minimum TM score of 0.61 and 0.42, respectively.

There were 2235 pairs of alignments in the closely similar set and 114 477 pairs in the more distantly similar set. The results are quite similar for both structural distances: TM-align, FATCAT (flexible) and Fr-TM-align are the best three methods in all cases regardless of the metric used. Similarly, SAP and MAMMOTH both rank as worst by all metrics. The SI and SAS metrics produce identical rankings in almost all cases.

These results show that good alignments by geometric standards are not necessarily consistent, suggesting that there may be some overfitting to geometric scores when considered from the point of view of assigning homology. Conversely SAP, which is highly consistent, produces poor geometric scores by the four metrics used here. However, since SAP tends to produce longer alignments than some of the other methods this could be the result of trimming the alignment differently.

4 DISCUSSION

We have assessed seven popular methods for pairwise structural alignment using geometric measures and inconsistency (as a proxy for homology) and found a wide range of independent variation on these measures: some methods (e.g. FATCAT) produce good

structural scores but are highly inconsistent. Others (e.g. SAP) are highly consistent but do poorly with measures of geometric accuracy. On average, TM-align is both consistent and geometrically sensitive for structurally similar domains, whereas Fr-TM-align is a better choice for less similar domains, showing that it is possible to perform well at both simultaneously. However even for the most consistent methods the level of inconsistency is very high.

What meaning does the substantial level of inconsistency that we have identified have for structural alignment? One important point is that none of the methods have (to our knowledge) been devised with consistency as an aim, which is likely to be at least partially responsible. The degree of consistency represents an upper-bound on the information a method can provide about homologous relationships: consistency is necessary but not sufficient for alignments to be accurate in this regard.

This poses an interesting question: is it possible in principle for structural alignment methods to be completely accurate in determining homology between positions? (a similar question was previously asked by Godzik (1996)) If equivalences are allowed to occur in an arbitrary order then the optimal alignment frequently involves substantial non-monotonicities even for quite similar structures (Shih and Hwang, 2004). This is highly unlikely from the point of view of sequence evolution and so it suggests a substantial tension exists between finding optimal structural similarities and finding positional homologous relationships. In the 1D–1D mapping established by sequence alignments it is not possible to have many-one relationships, for example, however once the distance cutoff exceeds a certain level it is possible for the optimal set of equivalences to contain many-one pairings since several regions might be equally related via duplication. In this situation it might be necessary to use a different scoring system, such as that proposed by Schulz (1977).

The most significant contributory factor to inconsistent structural alignments is the treatment of gaps. Since indels are responsible for much of the structural change which occurs through evolution, it is clearly necessary to develop a more accurate gap score for structural alignment. Generally it may be recognized that finding optimal structural similarity and finding homologous residues are not necessarily one and the same goal and that it might therefore be necessary to create methods designed for each purpose.

Another important issue is that optimization of structural similarity is not in all cases the ideal strategy for identifying homology: functional motions of the kind which flexible alignment aims to capture are one obvious example where homologous relationships would not in the case of rigid body alignment be captured correctly since homologous residues are no longer structurally equivalent in such a case. However, in general it is well known that structural alignments are very useful for improving multiple sequence alignments, suggesting that they do provide a useful guide to positional homologies (Armougom *et al*., 2006; O'Sullivan *et al.,* 2004; Pei *et al*., 2008). These methods all use consistency as the basis for incorporating constraints derived from structural alignments into their multiple sequence alignments. However, if the underlying structural alignments are themselves inconsistent this will either lead to conflicting restraints (in the case of several structures) or unacknowledged errors (in the case of only a single pair of structures). The potential for multiple sequence alignments to benefit from this is limited by the quality of the underlying methods. Improvements in the consistency of pairwise structural alignments would therefore have great benefits for these methods.

Our results suggest that, at least for FATCAT, allowing multiple rotational frames leads to better structural similarity scores (as it should) at the expense of greater inconsistency. Flexible alignment is correctly seen as an important innovation in aligning protein structures, however our results demonstrate that it is not a panacea: introducing flexibility into alignments greatly enlarges the search space, exacerbating problems in generating high quality alignments for more distantly related proteins. Although it is obviously necessary to use flexible alignments in certain circumstances, such as where functional motions have occurred, it seems prudent to use them only when it is known beforehand that their use is justified in a particular case, and manual supervision of the final results is advisable.

Overall, the principal theme underlying the inconsistencies identified is repetition: as with sequence alignment the existence of strongly periodic structures creates ambiguities for alignment. Thus the least consistently aligned domains are the repeats such as beta-helices and the least consistently aligned elements are generally helices, since these are often quite regular and may therefore present similar problems to repetitive folds. It seems likely that elements of repetitious regularity are not properly accounted for in decoy sets used to generate background distributions. Inclusion of this feature may be significant in improving our understanding of the statistical distribution of protein structural similarity. In particular this might be of use in developing tools for comparing repeat structures.

Another possibility for improving the results of large-scale pairwise alignments (e.g. in database search or when using large datasets) is to realign significantly similar structures using a consistency criterion (as used by, e.g. T-COFFEE; Notredame *et al*., 2000 and MULTAL; Taylor *et al*., 2000) to arrive at an improved set of pairwise alignments. This could be done without requiring full multiple alignment of the set of structures, which is impractical if the number of structures is large. This suggests reevaluating the problem from a strictly pairwise basis to a one-versus-many basis. In general, however the results we have shown demonstrate that there are still significant improvements to be made in pairwise structural alignment and have suggested several possibilities for improvement.

ACKNOWLEDGEMENTS

We thank Alessandro Pandini for a critical reading of the manuscript and the anonymous reviewers for their helpful suggestions. Also we thank the staff at the RCSB PDB for providing source code and class libraries for CE and FATCAT.

Funding: MRC (UK): Grant number U117581331.

Conflict of Interest: none declared.

REFERENCES

Alesker,V. *et al*. (1996) Detection of non-topological motifs in protein structures. *Protein Eng.*, **9**, 1103–1119.

Andreeva,A. *et al.* (2008) Data growth and its impact on the SCOP database: new developments. *Nucleic Acids Res.*, **36**, D419–D425.

Armougom,F. *et al.* (2006) Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. *Nucleic Acids Res.*, **34** (Suppl. 2), W604–W608.

Birzele,F. *et al.* (2007) Vorolign-fast structural alignment using Voronoi contacts. *Bioinformatics*, **23**, E205–E211.

Brenner,S.E. *et al.* (2000) The ASTRAL compendium for protein structure and sequence analysis. *Nucleic Acids Res.*, **28**, 254–256.

Budowski-Tal,I. *et al.* (2010) FragBag, an accurate representation of protein structure, retrieves structural neighbours from the entire PDB quickly and accurately. *Proc. Natl Acad. Sci.,* **107**, 3481–3486.

Carugo,O. (2007) Recent progress in measuring structural similarity between proteins. *Curr. Protein Pept. Sci.*, **8**, 219–241.

Chen,Y. and Crippen,G.M. (2005) A novel approach to structural alignment using realistic structural and environmental information. *Protein Sci.*, **14**, 2935–2946.

Gibrat,J.F. *et al.* (1996) Suprising similarities in structure comparison. *Curr. Opin. Struct. Biol.*, **6**, 377–385.

Godzik,A. (1996) The structural alignment between two proteins: is there a unique answer? *Protein Sci.*, **5**, 1325–1338.

Holm,L. and Sander,C. (1993) Protein-structure comparison by alignment of distance matrices. *J. Mol. Biol.*, **233**, 123–138.

Hollup,S.M. *et al.* (2011) Exploring the limits of fold discrimination by structural alignment: a large scale benchmark using decoys of known fold. *Comput. Biol. Chem.*, **35**, 174–188.

Kabsch,W. and Sander,C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22,** 2577–2637.

Kifer,I. *et al.* (2011) GOSSIP: a method for fast and accurate global alignment of protein structures. *Bioinformatics*, **27**, 925–932.

Kolodny,R. *et al.* (2005) Comprehensive evaluation of protein structure alignment methods: scoring by geometric measures. *J. Mol. Biol.*, **346**, 1173–1188.

Krissinel,E. and Henrick,K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D.*, **60**, 2256–2268.

Lackner,P. *et al.* (2000) ProSup: a refined tool for protein structure alignment. *Protein Eng.*, **13**, 745–752.

Mayr,G. *et al.* (2007) Comparative analysis of protein structure alignments. *BMC Struct. Biol.*, **7,** 50.

Morais,D.A.D.L. *et al.* (2011) Superfamily 1.75 including a domain-centric gene ontology method. *Nucleic Acids Res.*, **39**, D427-D434.

Mosca,R. *et al.* (2008) Alignment of protein structures in the presence of domain motions. *BMC Bioinformatics*, **9**, 352.

Notredame,C. *et al.* (2000) T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.*, **302**, 205–217.

Novosad,T. *et al.* (2010) Searching protein 3D structures for optimal structure alignment using intelligent algorithms and data structures. *IEEE Trans. Inf. Technol. Biomed.*, **14**, 1378–1386.

O'Sullivan,O. *et al.* (2004) 3DCoffee: Combining protein sequences and structures within multiple sequence alignments. *J. Mol. Biol.*, **340**, 385–395.

Ortiz,A.R. *et al.* (2002) MAMMOTH (matching molecular models obtained from theory): an automated method for model comparison. *Protein Sci.*, **21**, 3255–3263.

Pandit,S.B. and Skolnick,J. (2008) Fr-TM-align: a new protein structural alingment methods based on fragment alignments and the TM-score. *BMC Bioinformatics*, **9**, 531.

Pei,J.M. *et al.* (2008) PROMALS3D web server for accurate multiple protein sequence and structure alignments. *Nucleic Acids Res.*, **36**, W30–W34

Porter,C.T. *et al.* (2004) The Catalytic Site Atlas: a resource of catalytic sites and residues identified in enzymes using structural data. *Nucleic Acids Res.*, **32**, D129–D133.

Rocha,J. *et al.* (2009) Flexible structural protein alignment by a sequence of local transformations. *Bioinformatics*, **25**, 1625–1631.

Sadowski,M.I. and Taylor,W.R. (2010a) Protein structures, folds and fold spaces. *J. Phys. Condens. Matter*, **22**, 033103.

Sadowski,M.I. and Taylor,W.R. (2010b) On the evolutionary origins of "fold space continuity": a study of topological convergence and divergence in mixed alpha-beta domains. *J. Struct. Biol.*, **172**, 244–252.

Salem,S. *et al.* (2010) FlexSnap: flexible non-sequential protein structure alignment. *Algorithm. Mol. Biol.*, **5**, 12.

Schulz,G.E. (1977) Recognition of phylogenetic relationships from polypeptide chain fold similarities. *J. Mol. Evol.*, **9**, 339–342.

Shatsky,M. *et al.* (2004) FlexProt: alignment of flexible protein structures without a predefinition of hinge regions. *J. Comput. Biol.*, **11**, 83–106.

Shibberu,Y. and Holder,A. (2011) A spectral approach to protein structure alignment. *IEEE Trans. Comput. Biol. Bioinf.*, **8**, 867–875.

Shih,E.S.C. and Hwang,M.J. (2004) Alternative alignments from comparison of protein structures. *Proteins*, **56**, 519–527.

Shindyalov,I.N. and Bourne,P.E. (1998) Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Prot. Eng.*, **11**, 739–747.

Taylor,W.R. (1999) Protein structure comparison using iterated double dynamic programming. *Protein Sci.*, **8**, 654–665.

Taylor,W.R. and Sadowski,M.I. (2010) Protein products of tandem gene duplication: a structural view. In Dittmar,K. and Liberles,D. (eds) *Evolution After Gene Duplication*. Wiley-Blackwell, Hoboken, NJ, pp. 133–163.

Taylor,W.R. *et al.* (2000) Multiple protein sequence alignment using double-dynamic programming. *Comput. Chem.*, **24**, 3–12.

Taylor,W.R. *et al.* (2002) A Fourier analysis of symmetry in protein structure. *Prot. Eng.*, **15**, 79–89.

Veeramalai,M. *et al.* (2009) TOPS++FATCAT: fast flexible structural alignment using constraints derived from TOPS+ strings model. *BMC Bioinformatics*, **9**, 358.

Vesterstrom,J. and Taylor,W.R. (2006) Flexible secondary structure based protein structure comparison applied to the detection of circular permutation. *J. Comput. Biol.*, **13**, 43–62.

Ye,Y.Z. and Godzik,A. (2003) Flexible structure alignment by chaining aligned fragment pairs allowing twists. *Bioinformatics*, **19**, ii246–ii255.

Zhang,Y. and Skolnick,J. (2004) Scoring function for the assessment of protein structure template quality. *Proteins*, **52**, 702–710.

Zhang,Y. and Skolnick,J. (2005) TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.*, **33**, 2302–2309.