PROTEINS: Structure, Function, and Genetics Suppl 5:163–170 (2001)
DOI 10.1002/prot.10053

Comparison of Performance in Successive CASP Experiments

Česlovas Venclovas,[1,2] Adam Zemla,[1] Krzysztof Fidelis,[1] and John Moult[3]*
[1]*Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore, California*
[2]*Institute of Biotechnology, Vilnius, Lithuania*
[3]*Center for Advanced Research in Biotechnology, University of Maryland Biotechnology Institute, Rockville, Maryland*

ABSTRACT **As the number of completed CASP (Critical Assessment of Protein Structure Prediction) experiments grows, so does the need for stable, standard methods for comparing performance in successive experiments. It is critical to develop methods for determining the areas in which there is progress and in which areas are static. We have added an analysis of the CASP4 results to that previously published for CASPs 1, 2, and 3. We again use a unified difficulty scale to permit comparison of performance as a function of target difficulty in the different CASPs. The scale is used to compare performance in aligning target sequences to a structural template. There was a clear improvement in alignment quality between CASP1 (1994) and CASP2 (1996). No change is apparent between CASP2 and CASP3 (1998). There is a small barely detectable improvement between CASP3 and the latest experiment (CASP4, 2000). Alignment remains the major source of error in all models based on less than about 30% sequence identity. Comparison of performance in the new fold modeling regime is complicated by issues in devising an objective target difficulty scale. We have found limited numerical support for significant progress between CASP3 and CASP4 in this area. More subjectively, most observers are convinced that there has been substantial progress. Progress is dominated by a single group. Proteins 2001;Suppl 5:163–170.** © 2002 Wiley-Liss, Inc.

Key words: protein structure prediction; communitywide experiment; CASP

INTRODUCTION

How can the quality of CASP4 best be compared with that in earlier CASPs? Aspects of this question are addressed in articles by the assessors in this issue of *Proteins*. In the present study, we attempt a more global view, focusing on performance across CASPs, rather than across individuals, and using a small set of numerical evaluation tools. These tools are similar but not identical to those in the corresponding CASP3 paper.[1]

GENERAL CONSIDERATIONS

Choice of Models to Evaluate

In CASPs 3 and 4, up to five models were submitted by a prediction group on each target. More models were permitted in earlier CASPs. We have used two selections from the available models. One allows measurement of the very best performance, and is the best model from any group for a given target, irrespective of whether it was ranked as the expected best by that group. The second selection is the average over the best models from the six best-performing groups, and facilitates measurement of the extent to which the best results are dominated by a single group, or generally represent the state of the art.

Relative Target Difficulty

Not all protein structures are equally difficult to model. At one end of the spectrum, targets with a high level of sequence identity to a known structure can be modeled with relatively small errors (typically <1 Å for Cα atoms at $>60\%$ sequence identity), while at the other, many new folds are still very hard to predict, and all models of these folds may be close to random. Any attempt to compare performance over different CASPs must therefore begin with establishing difficulty scales. In the comparative modeling and fold recognition regimes, the difficulty of constructing an accurate model has been shown to depend primarily on two factors,[2] (1) the level of sequence identity between the target protein and that of the nearest protein of known structure, and (2) the extent to which the corresponding structures can be superimposed. In the previous analysis,[1] we used the DALI algorithm[3] to measure structure superposability. In the present work, the local global alignment (LGA) algorithm[4] has been used. LGA uses a rigid structure superposition, as opposed to the contact superposition method of DALI, and generally finds a solution with slightly more corresponding residues under a given threshold. Superposability is defined as the fraction of Cα pairs that are closer than 5 Å in the LGA superposition. The ranking of target difficulty has not been affected significantly by this change in the method of superposition. Sequence identity is taken to be the fraction

Grant sponsor: Department of Energy; Grant number: DE-FG02-96ER-6227; Grant sponsor: University of California/Lawrence Livermore National Laboratory; Grant number: W-7405-Eng-48.

*Correspondence to: John Moult, Center for Advanced Research in Biotechnology, University of Maryland Biotechnology Institute, 9600 Gudelsky Drive, Rockville, MD 20850. E-mail: jmoult@tunc.org

Received 20 July 2001; Accepted 29 October 2001

of identical residue pairs aligned in the LGA superposition.

Choice of Template Structures

In assessing target difficulty, it is necessary to use the best template structure available at the time of the appropriate CASP experiment, not the best one currently available. For CASPs 1 and 2, lists of the potential template structures to each target were generated using DALI.[3] For CASPs 3 and 4, these lists were generated using PROSUP.[5] For the present analysis, template–target superpositions were generated for all the structures in these lists using LGA, and the template with the highest number of residues matched to a target chosen. In a few instances, the most structurally similar templates have a substantially lower sequence identity to the corresponding target than one of the competing templates. In these cases (one in CASP4, 10 over all CASPs), the template with the higher sequence identity was selected. In practice, the difference in achievable model quality is little affected by whether the highest sequence or highest structure identity template is selected for these targets.

The assessor analyses in CASP4 make use of more extensive parsing of structures into domains than was done in previous CASPs. That parsing permits more effective isolation and assessment of the features of models. We are more concerned with the view of a predictor. It is often not possible to build models based on domains, because of the difficulties of identifying the boundaries from sequence. For the comparative modeling and fold recognition target analysis, only three targets have been divided into domains: T0090, T0116, and T0121. Two of these have obvious domain divisions apparent from sequence comparisons, and the third, T0116, is a very large structure, obviously multidomain, although domain boundaries are hard to identify from sequence. For these targets, whole structures, as well as the domains, appear in the plots.

DISTRIBUTION OF TARGET DIFFICULTY

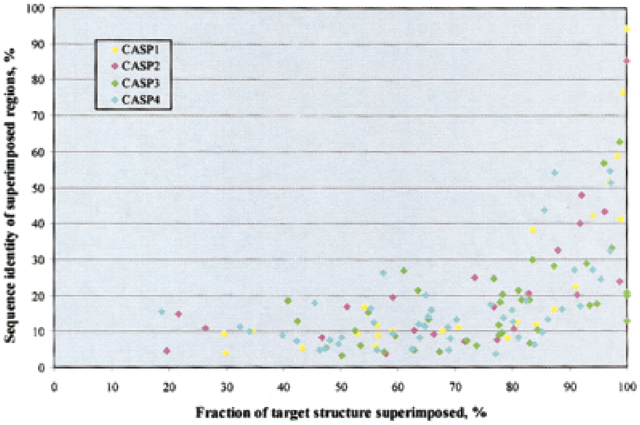

Fig. 1. Relative Difficulty of targets in the four CASP experiments, displayed as a function of the percentage sequence identity between the target and the best available template (vertical axis) and the fraction of the target structure that can be superposed on the template (horizontal axis). Comparative modeling targets tend to cluster at the right-hand side of the plot, with easiest (high sequence identity) at the top. Fold recognition targets are spread out from right to left, with the most difficult (usually analogous relationships) at the far left. Targets from each CASP span the full range of difficulty.

Figure 1 shows the distribution of sequence identity and superposability for most of the targets in all four CASP experiments (some very short new fold targets are omitted). The distribution of difficulty in CASP4 is similar to that of the earlier CASPs. Some features are worth noting. As in CASP3, there are few targets with a sequence identity greater than 60% to a known structure. Two targets at relatively high sequence identity (~44% and ~54%) have a rather low fraction of superposable residues (~84%, as opposed to a more normal >90% for this level of sequence identity). These are T0123, pig β-lactoglobin, and T0099, a designed SH3 protein. The structure of T0099 was determined by nuclear magnetic resonance (NMR), and the average structure may be less accurate than a typical x-ray structure. The β-lactoglobin has an unusually large number of local structures different from the bovine relative of known structure. There is also a set of targets within the 80–90% superposable range with a relatively low degree of sequence identity. The far left CASP4 point represents the full T0116 structure. The full structure of T0121 is the relatively high sequence identity CASP4 point at about 58% superposability.

Analysis of performance requires the projection of the data in Figure 1 into a one-dimensional ranking of difficulty. As in the previous analysis, target difficulty is ranked by a combination of the fraction sequence identity of superposable residues and the fraction of total residues that can be superposed. That is, the difficulty of each target is expressed as a linear combination of ranks by structure superposability and sequence identity:

$$(\mathrm{RANK\_STR\_ALN}+\mathrm{RANK\_SEQ\_ID})/2$$

where RANK_STR_ALN is the rank of the target along the horizontal axis of Figure 1, and RANK_SEQ_ID is the rank along the vertical axis.

ALIGNMENT ACCURACY

A critical factor dominating model quality throughout the comparative modeling and fold recognition regimes is the accuracy of alignment of the target sequence onto a structural template.[6] An error of one residue along the chain results in a Cα error of 3.8 Å, an error of four positions, up to ~12 Å. Such large errors overwhelm other factors, and so detection of improvement in alignment accuracy is a key performance measure. Figure 2 shows the fraction of correctly aligned residues in the best model for each target, ordered by target difficulty, for the four CASPs. Residues are considered to be correctly aligned if the corresponding Cα atoms fall within 3.8 Å of one another in the LGA superposition between a model and the experimental structure, and no other Cα is closer. The solid bars show the fraction of residues correctly aligned,

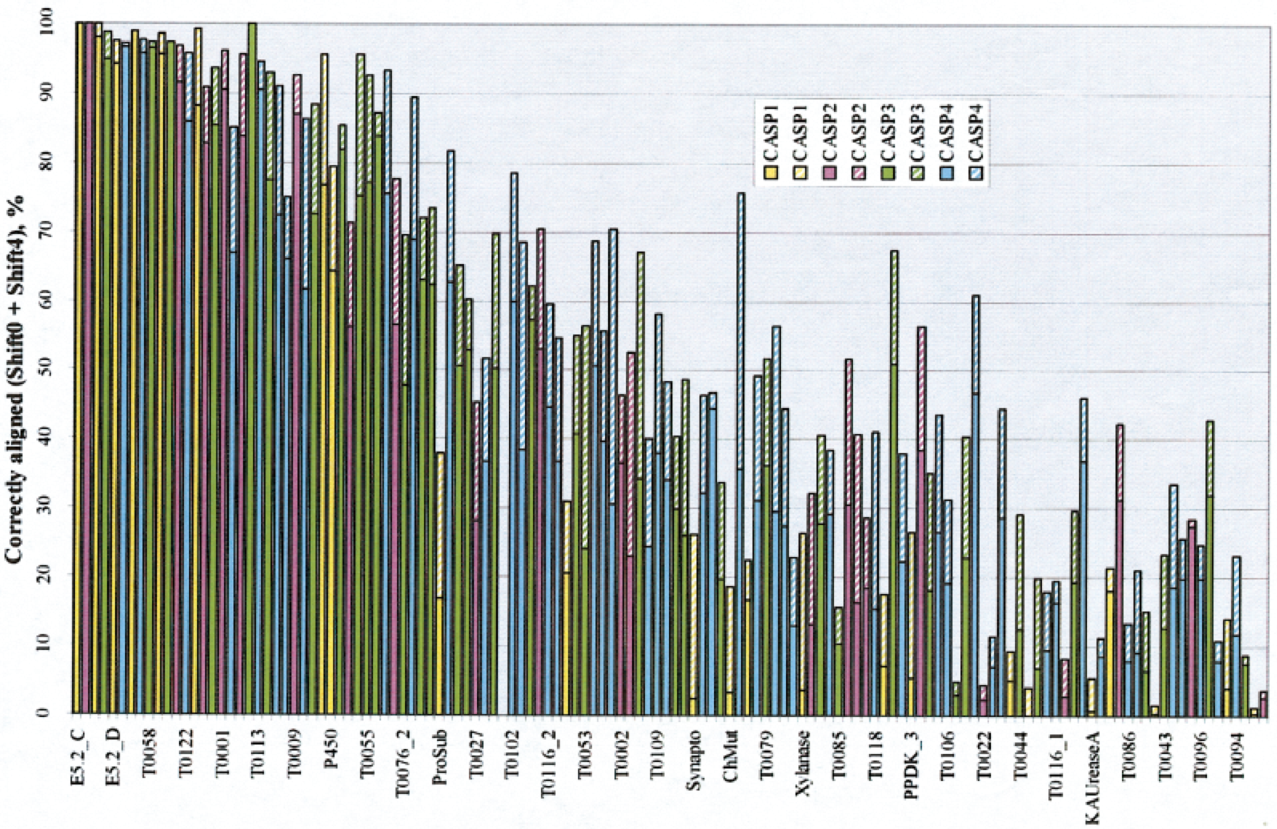

Fig. 2. Fraction of residues correctly aligned between the target structure and the best model for each target. Yellow, CASP1; red, CASP2; green, CASP3, blue, CASP4. Full bars represent the fraction of correctly aligned residues, and hatched bars the additional fraction of residues in error by not more than four residue positions. The targets are arranged top to bottom, starting with the least difficult. Alignment accuracy falls steadily with increasing difficulty of targets. For the more difficult, targets, it is clear that performance in later CASPs is superior to CASP1, but there is no easily discernable difference between CASPs 2, 3, and 4. The data corresponding to this plot are available at the CASP website (predictioncenter.llnl.gov).

and the hashed bars show the additional fraction of residues aligned to within ±4 residues.

At the top of the plot, the comparative modeling targets with a high level of sequence identity (down to ~30%) to a known structure have essentially perfect alignments. (Perfect accuracy is rarely 100%, since even at very high sequence identity, some regions are not alignable, e.g., sequence equivalent loops with different conformations.) The first serious misalignment is of a helix in target T0122, at 32% identity. Below this target, as the difficulty increases, the alignment quality deteriorates progressively and rapidly. The first fold recognition targets begin at approximately T0114. Thereafter, the order is approximately homologous fold recognition targets, followed by analogous folds, and then by new folds (see ref. 7, for discussion of these terms). In CASP4, in all but two cases, the best fold recognition models are based on a correct template but, as a general rule, they are no more than 40% correctly aligned, and often considerably worse. New folds targets, typically at the far right of the plot, usually have low-quality models, and this results in a poor alignment score. Alignment is a less useful metric for this class of models, and an alternative measure is discussed later.

It is clear from Figure 2 that there is considerable variation in the accuracy of the alignment of the best model in a given region of target difficulty, making it hard to see trends over the different CASPs. Figure 3 shows the same data, averaged over sets of five consecutive targets to produce a smoother plot. Figure 3(a) presents the data for fraction of best models correctly aligned (corresponding to the solid bars in Fig. 2). Figure 3(b) shows the fraction aligned within ±4 residues. The most striking feature of both plots is the improvement in alignment between CASP1 and the subsequent experiments. One must look more closely to see any improvements thereafter. In the correct alignment plot [Fig. 3(a)], the CASP4 predictions are undistinguished from those of CASPs 2 and 3, except perhaps in two short regions of the difficulty span, around targets T0109 and T0106. In the plot allowing ≤4 residue alignment errors, there is a distinct but small improvement in CASP4, over a wide range of predominantly fold recognition targets, from around T0055 to T0079, and then again around T0118. Although this detectable improvement is better than none at all, it makes a dismayingly small dent in the problem. It is very clear that in the distant (<30% sequence identity) comparative modeling

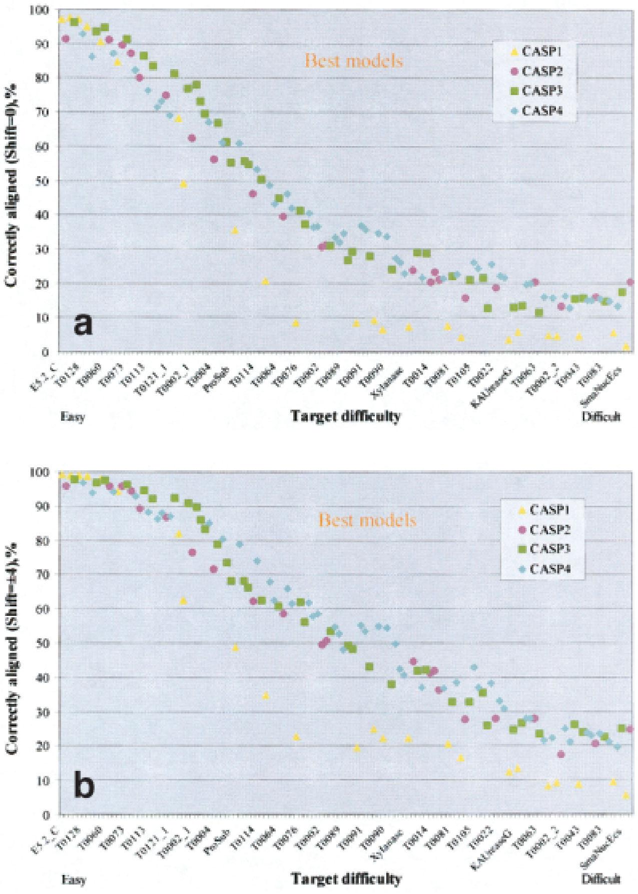

Fig. 3. Fraction of residues correctly aligned between the target structure and the best model for each target (**a**) and aligned with an error of ≤4 residue positions (**b**). To make the trends more visible, each data point in Figure 2 has been smoothed by averaging over itself and the two neighboring points on each side. Both plots show that performance improved substantially between CASP1 and CASP2, but not detectably between CASP2 and CASP3. A small improvement from CASP3 to CASP4 is apparent in a few regions. The data corresponding to this plot are available at the CASP web site (http://predictioncenter.llnl.gov).

and fold recognition regimes alignment quality remains THE bottleneck to improving the quality of the model.

FACTORS AFFECTING EVALUATION IN THE NEW FOLD MODELING REGIME

A number of factors must be taken into account when comparing performance in the new fold regime:

Target difficulty: Different considerations apply than for other classes of prediction. First, as discussed below, some classes of fold are substantially easier to predict than others, for example all alpha structures versus all β. Second, many new fold methods depend on identifying motifs, ranging in size from a few residues to full domains, of similar conformation in known structures. The extent to which such motifs exist for a particular target is therefore very relevant, and for larger targets there is a higher probability that appropriate motifs will be available for a significant number of residues. Third, Baker et al.[8] recently identified contact order as a critical variable determining predictability of fold. That is, folds in which the average number of residues along the chain between contacting residues is small are significantly easier to model than are cases in which that quantity is large.

Choice of evaluation metric: Comparison of performance is also complicated by the fact that the quality of the models is rarely very high. Experience over the CASPs has shown that it is difficult to find evaluation criteria that capture in a quantitative way the few good features of such models. We have again used the global distance test (GDT),[9] introduced in CASP3. The algorithm finds the maximum number of residues where the distance between the target and corresponding model Cα is less than some threshold, in a sequence-dependent superposition. In looking at the results, it should be born in mind that a distance threshold is a stricter criterion than an root-mean-square deviation (RMSD), in the sense that the RMSD of a set of residues is usually substantially less than the distance threshold used to define the set. (See the GDT data on the CASP website, http://predictioncenter.11nl.gov, for examples of the relationship between a distance threshold and RMSD—factors of ≤2 are not uncommon.) We consider distance thresholds of 1, 2, 4, and 8 Å.

Choice of targets: In addition, the choice of what exactly should be considered a target for new fold methods remains unclear. In CASP4, only four domains were categorized as strictly new folds by the assessors. A further 12 domains are in a gray area between a new fold and an analogous fold. In practice, fold recognition methods work poorly on this latter class of target, and new fold methods generally, but not always, do better. So we have included all 16 domains in the analysis. In the previous analysis, only the seven CASP3 targets considered difficult or impossible fold recognition targets, and <120 residues, were included. In CASP4, it is more clear that interesting results are also being obtained for longer targets. We address this difference by considering both the seven short targets for CASP3 and CASP4 and the full set of targets in both CASPs.

Role of luck: Finally, a further complication of low model quality is that luck may play a more significant role. We address that below by considering consistency of performance.

In spite of all these complications, we do consider the comparison across CASPs still worth making. Real substantial progress is easy to spot. As always, trouble with noise is an indication of a weak signal.

NEW FOLDS MODELING PERFORMANCE

Figure 4(a) demonstrates the relative performance on the new fold targets in the different CASPs, using the GDT criterion, and using the best model for each target. For each target, the number of residues superposable below an inter-Cα distance threshold of 1, 2, 4, and 8 Å is shown. A simple way of looking at these data is to choose a thresh-

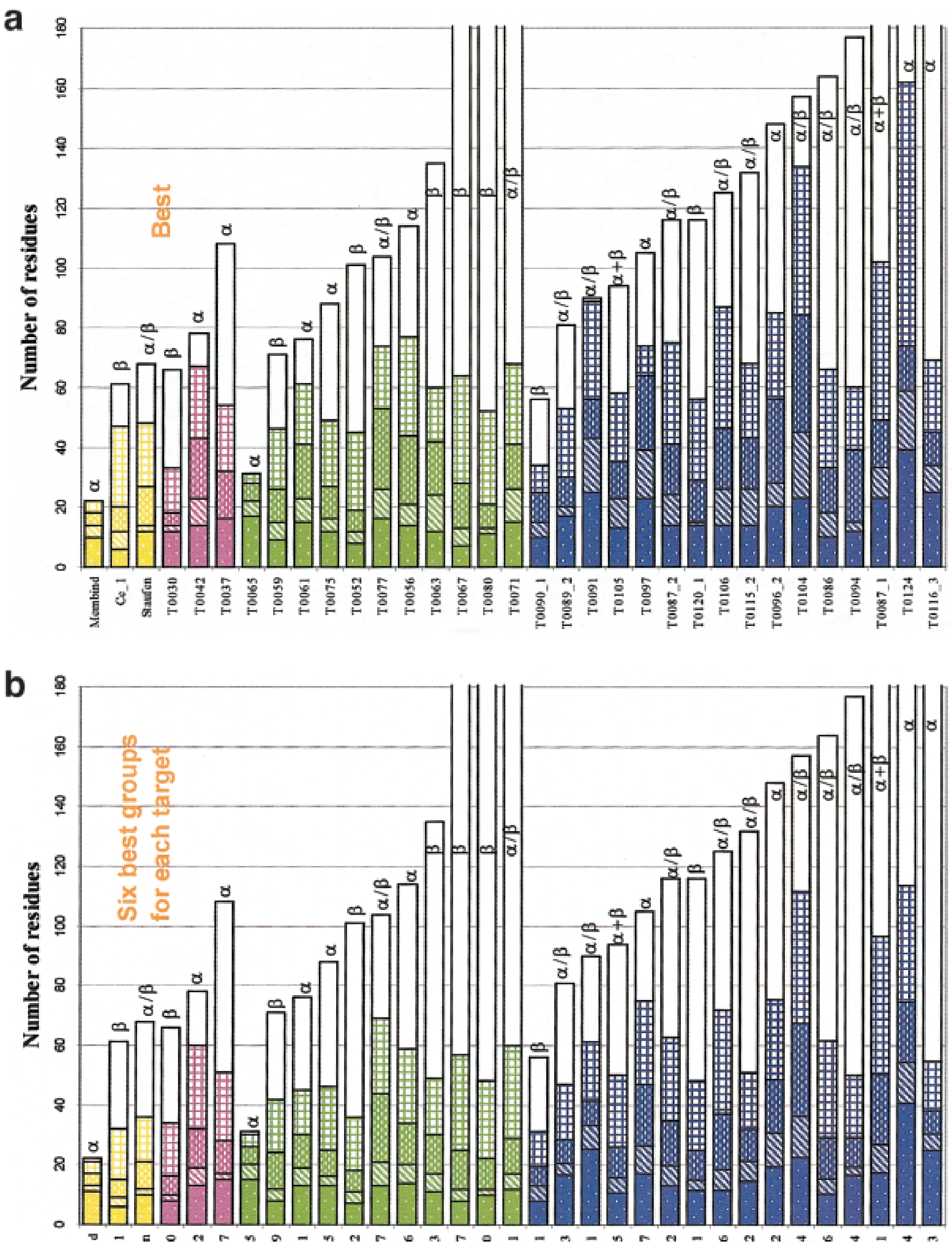

Fig. 4. Comparison of new fold modeling performance in the four CASPs. **a:** The best prediction on each target. **b:** Average over the best predictions from up to six top groups for each target. The number of residues closer than 1, 2, 4, 8, and >8 Å to the equivalent residues in the target structure are shown in each bar. Greek letters indicate the predominant topology class of each target. Targets are ordered by size, for each CASP. Yellow, CASP1; red, CASP2; green, CASP3; blue, CASP4. In CASP1, no prediction approached an accuracy of 40 residues superposed to <4 Å. In CASP2, one target met this criterion, in CASP3, five did so, and in CASP4, 10. Although this trend in encouraging, differences in the number of targets and the types of topology reduce the significance of the signal. **b:** Performance is less good, suggesting that only a few groups produce the highest-quality models.

old, and to see whether any prediction on a target exceeds that value. Following the new fold assessors in recent CASPs, we ask whether there are any predictions with more than 40 residues superposable below a 4Å threshold (a 40/4 criterion). In CASP1, no models pass this threshold, in CASP2 there is one (T0042). For CASPs 3 and 4, first consider the sets of seven shortest targets. In both CASP3 and CASP4, three of these meet the 40/4 criterion (T0061, T0077, and T0056 in CASP3 and T0091, T0097, and T0087-2 in CASP4). One CASP3 target is <40 residues, so cannot meet the threshold. In both CASP3 and CASP4, two of these targets are helical, and one is an α/β structure.

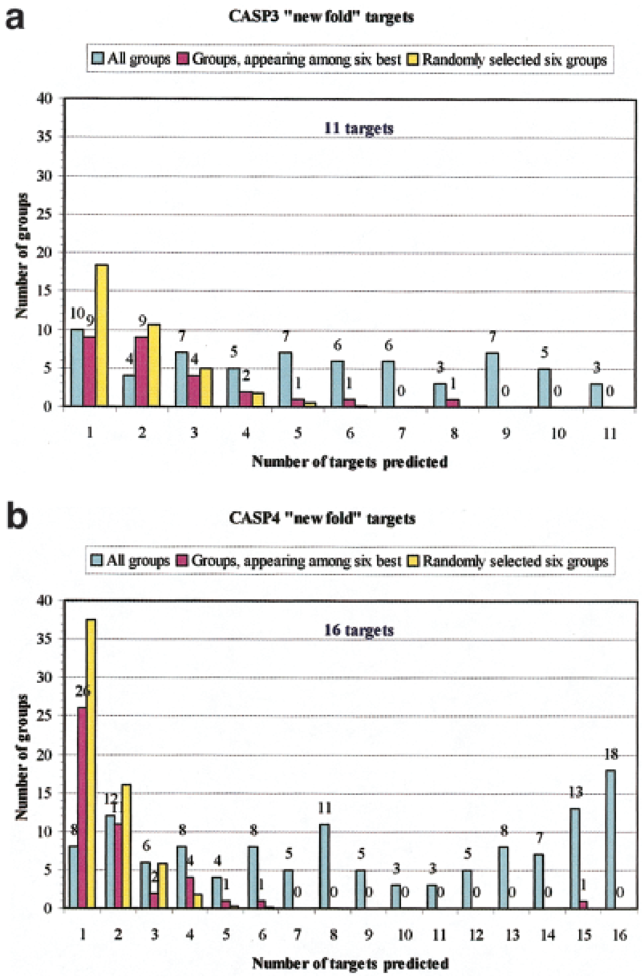

Fig. 5. Measure of sustained performance in new fold modeling. **a:** CASP3. **b:** CASP4. Data cover the new fold targets in each case. Blue bars show the number of groups submitting models for one, two, up to the total number of the targets. In all, almost twice as many groups took part in this category in CASP4 than in CASP3, and they tended to attempt more of the targets. The brown bars show the number of groups scoring among the six best for one target, two targets, three, and so on. In both CASPs, there is a tendency for each group to score well on a very small number of targets. Most obviously in CASP4, where 26 groups were among the best six only one time out of a possible 16. The yellow bars show the expected number of groups in the top six performers for one, two, and so on targets if the results were completely random. The actual and random distributions are similar, particularly in CASP4. These data suggest that few groups had any level of sustained performance, and that apparent improvements in CASP4 may be partly a consequence of more groups taking part. A very notable exception is provided by the performance of a single group, who were among the top 6, 8 times out of 11 in CASP3, and an impressive 15 out of 16 in CASP4.

Thus, by this measure, there is an improvement between CASP1 and CASP2, and between CASP2 and CASP3, but no obvious difference in performance between CASP3 and CASP4.

The use of additional thresholds does suggest that there is some improvement in CASP4. One CASP4 target has 40 residues of under a 2-Å threshold (T0091), and a second has 60 residues of under a 4 Å (T0097). T0091 also has almost all the residues of under a 8 Å, corresponding to an overall RMSD of 5 Å. This is a clear new fold and a protein of unknown function, so there is little possibility of any other information other than the strength of new fold prediction methods leading to this result. These levels of performance have never seen before in CASP and, although limited, are very encouraging.

Inclusion of the longer targets also leads to a stronger impression of progress between CASP3 and CASP4. In CASP4, seven out of nine longer targets have at least one prediction with 40 residues of under a 4-Å threshold (T0106, T0115-2, T0096-2, T0104, T0124, T0087-1, and T116-3). One of these has 60 residues of under 4, and one has 80 residues of under 4. The 80 residues of under 4 performance was obtained using fold recognition methods, and presumably reflects the fact that this structure contains a subdomain with a known fold. But the best predictions on the other six targets were all achieved using new fold methods. Five of these well-predicted structures are helical, and the remaining two are α/β proteins—the same pattern as for the short more successfully predicted targets in both CASP3 and CASP4. In CASP3, there are four additional longer targets. Performance on two of these targets meets the 40/4 criterion (T0063 and T0071), but one of those results (for T0071) was obtained using fold recognition methods.

The comparison is complicated by the fact both the well predicted longer CASP3 targets are predominantly antiparallel β topologies. These are the only cases of new fold fragments of reasonable size that were correctly predicted for this topology in any CASP. All six CASP4 targets for which no model reached the 40/4 threshold are predominantly antiparallel β structures. Evidently, extensive β structure is still very difficult to predict. It is worth noting that there were additional reasons why two of the CASP4 targets are hard: one (T0089) is an extra domain that is very difficult to identify as such, without structural information. The second (T0090-1) is a domain that is very noncompact in the monomer, but forms part of the dimer interface in the biological unit. Modeling a dimer without information about the contacting regions is beyond the scope of current methods.

On balance, in spite of the complication with the effect of topology and the number of targets, the results do seem to represent real progress. In particular, predictions meeting the 40/4 threshold were obtained for more than one-half the full set of CASP4 new fold targets. Further, a few targets had longer regions under this threshold or had more accurate predictions for 40 residues.

CONSISTENCY PERFORMANCE IN THE NEW FOLD REGIME

Are one or very few groups performing well, or are many people able to make predictions of approximately the same quality in the new folds regime? We assess that in two ways. First, Figure 4(b) shows the same set of targets as in Figure 3(a), with the average performance over the top best-performing six groups for each target, instead of the best performance for each. The single target in CASP2 (T0042) no longer meets the 40/4 threshold, three out of the four targets in CASP3 that met the criterion do not do so, and only 4 out of the previous 10 in CASP4 meet the criterion. The smaller decrease in GDT scores in CASP4 compared with the other CASPs suggest an improved general level of performance.

Other factors need to be taken into account when looking at general performance. In particular, the number of groups making new fold category predictions almost doubled between CASP3 and CASP4: 124 attempting one or more of these targets in CASP4 versus 63 in CASP3. In the limit, if the predictions are random, more significant looking predictions would result. We address this effect by considering how many groups consistently fall into the top six performers across multiple targets. Figure 4 shows these data for CASP3 [Fig. 4(a)] and CASP4 [Fig. 4(b)]. The blue bars show the number of groups who predicted one, two, up to the total number of targets in each CASP. In CASP4, the largest number of groups (18) attempted all targets. In CASP3, the most popular number of targets to attempt was only one. So, more predictors attempted more targets in CASP4. The brown bars show the number of groups who were among the six best scoring for one, two, up to all targets. The yellow bars show the number of groups falling in the top six for one, two and so on targets, derived from randomly picking six best groups from those that attempted each target. In both CASPs, the random and actual distributions are similar, with a large number of groups scoring in the six best a small number of times. The resemblance to random is significantly stronger for CASP4. In CASP3, nine groups scored among the six best only once, and nine did so twice. In CASP4, an astonishing 26 groups only scored in the best 6 once, and 11 did so twice. The tail of groups scoring in the six best three, four, five and so on times is similar in the two CASPs. This does suggest that the role of luck and the increased number of predictors has contributed to the apparent improvement in CASP4. Consideration of the subset of targets for which one or more groups achieved the 4/40 threshold shows a similar picture (data not shown). There is an outstandingly consistent performance by the same single group, who appear in the top six predictions eight (out of nine targets attempted) out of 11 times in CASP3 and an extraordinary 15 out of 16 times in CASP4. Particularly in CASP4, this group appears to be well ahead of everyone else, in terms of sustained performance.

CONCLUSIONS

On the whole, we find these results rather disappointing. Alignment quality, the factor dominating quality in the comparative modeling and fold recognition regimes, has improved very little since CASP2. There is some evidence of limited improvement in the new fold modeling regime, but it is dominated by a single group. Many different factors influencing performance in this category—different topologies, different number of predictors, the continuing blurring of the boundary between fold recognition methods and new fold methods—make it very difficult to draw firm conclusions. Still, if there had been a really large increase in model quality, it would easily stand out. With one or two exceptions, the best predictions still extend only to part of a structure. Several facts do point to likely progress between CASP3 and CASP4: more targets meeting the 40/4 criterion for some nonrandom significant fragment of structure, and targets meeting higher thresholds for the first time, including an overall RMSD to experiment of 5 Å on a complete structure of 90 residues. Also, almost everyone who has inspected the results for each CASP is convinced of progress. Unfortunately, we cannot find a way of demonstrating that in a statistically respectable manner. The best thing we can do is see whether there is again an apparent improvement at CASP5, as has been the case in the new fold category for every CASP so far.

Improvement in any category between CASPs does not necessarily reflect improvements in the methods. Some part will be attributable to the increasing size of the database of known sequences and structures on which to draw for predictions. Larger sequence families mean that bigger profiles can be built, and this should improve alignment performance and secondary structure prediction,[10] as well as aid in the choice of fragments for new fold construction. A larger base of known structures is particularly helpful in constructing motif libraries on all length scales for use in new fold building. Increased computer power allows longer and more folding trajectories to be run, as well as the use of more sophisticated and computationally intensive algorithms in other modeling areas, such as alignment and model refinement.

There are a number of areas of modeling for which we have not attempted comparison over the CASPs. In some areas, such as contact prediction,[11] numerical evaluation criteria are not mature enough to encourage this. We have also not devised a consistent numerical measure for the success of fold recognition, although this is probably possible. In other areas, it appears clear that any recent progress, if it exists, is too small to reliably measure. Such areas are secondary structure prediction, and detailed comparative modeling, including side chain construction, loop building and refinement. Emerging large-scale benchmarking procedures, both by individual groups,[12] and on a communitywide basis, using automatic servers with no human intervention,[13] are starting to provide more statistically robust data than can be obtained through CASP, and will play an increasingly important role in measuring progress.

ACKNOWLEDGMENTS

This work was supported in part by DOE grant DE-FG02-96ER 62271 (to J.M.). Part of the work was performed under the auspices of the U.S. Department of Energy by the University of California, Lawrence Livermore National Laboratory under contract W-7405-Eng-48.

REFERENCES

1. Venclovas C, Zemla A, Fidelis K, Moult J. Some measures of comparative performance in the three CASPs. Proteins 1999; Suppl 3:231–237.
2. Marchler-Bauer A, Bryant SH. Measures of threading specificity and accuracy. Proteins 1997;29S:74–82.
3. Holm L, Sander C. Dali: a network tool for protein structure comparison. Trends Biochem Sci 1995;20:478–480.
4. Cristobal S, Zemla A, Fischer D, Rychlewski L, Elofsson A. A study of quality for protein threading models. BMC Bioinformatics 2001;2:5.
5. Lackner P, Koppensteiner WA, Domingues FS, Sippl MJ. Automated large scale evaluation of protein structure predictions. Proteins 1999;37:7–14.
6. Martin AC, MacArthur MW, Thornton JM. Assessment of comparative modeling in CASP2. Proteins 1997;29S:14–28.
7. Murzin A, Hubbard TJP. Prediction targets of CASP4. Proteins 2001; this issue.
8. Bonneau R, et al. Rosetta in CASP4: Progress in ab initio protein structure prediction. Proteins 2001; Suppl 5:119–127.
9. Zemla A, Venclovas C, Moult J, Fidelis K. Processing and evaluation of predictions in CASP4. Proteins 2001; Suppl 5:13–21.
10. Rost B, Sander C. Third generation prediction of secondary structures. Methods Mol Biol 2000;143:71–95.
11. Lesk AL, Lo Conte L, Hubbard TJP. Assessment of novel fold targets in CASP4: predictions of three-dimensional structures, secondary structures, and interresidue contacts. Proteins 2001; Suppl 5:98–118.
12. Sauder JM, Arthur JW, Dunbrack RL Jr. Large-scale comparison of protein sequence alignment algorithms with structure alignments. Proteins 2000;40:6–22.
13. Bujnicki JM, Elofsson A, Fischer D, Rychlewski L. LiveBench-1: continuous benchmarking of protein structure prediction servers. Protein Sci 2001;10:352–361.