

BIROn - Birkbeck Institutional Research Online

Kryshtafovych, A. and Schwede, T. and Topf, Maya and Fidelis, K. and Mout, J. (2019) Critical assessment of methods of Protein Structure Prediction (CASP) – Round XIII. *Proteins: Structure, Function, and Bioinformatics* 87 (12), pp. 1011-1020. ISSN 0887-3585.

Downloaded from: <https://eprints.bbk.ac.uk/id/eprint/29348/>

Usage Guidelines:

Please refer to usage guidelines at <https://eprints.bbk.ac.uk/policies.html>
contact lib-eprints@bbk.ac.uk.

or alternatively

1
2
3 Critical Assessment of Methods of Protein Structure Prediction (CASP) –
4

5 Round XIII
6

7 Running title: CASP13
8
9

10
11
12 Andriy Kryshtafovych¹, Torsten Schwede², Maya Topf³, Krzysztof Fidelis¹ &
13

14 John Moulton^{4*}
15

16
17 ¹Genome Center, University of California, Davis
18

19 451 Health Sciences Drive, Davis, CA 95616, USA
20

21 ²University of Basel, Biozentrum & SIB Swiss Institute of Bioinformatics, Basel,
22

23 Switzerland
24

25
26 ³Institute of Structural and Molecular Biology, Birkbeck College, University of London,
27

28 London, United Kingdom.
29

30
31 ⁴Institute for Bioscience and Biotechnology Research
32

33 9600 Gudelsky Drive, Rockville, MD 20850, USA.
34

35 and Department of Cell Biology and Molecular Genetics
36

37 University of Maryland
38
39
40
41

42 *Corresponding author:
43

44 John Moulton
45

46 tel: 240-314 6241
47

48 fax: 240-314-6255
49

50 email: jmoulton@umd.edu
51
52

53
54 Keywords: Protein Structure Prediction, Community Wide Experiment, CASP
55
56
57
58
59
60

ABSTRACT

CASP (Critical Assessment of Structure Prediction) assesses the state of the art in modeling protein structure from amino acid sequence. The most recent experiment (CASP13 held in 2018) saw dramatic progress in structure modeling without use of structural templates (historically 'ab initio' modeling). Progress was driven by the successful application of deep learning techniques to predict inter-residue distances. In turn, these results drove dramatic improvements in three-dimensional structure accuracy: With the proviso that there are an adequate number of sequences known for the protein family, the new methods essentially solve the long-standing problem of predicting the fold topology of monomeric proteins. Further, the number of sequences required in the alignment has fallen substantially. There is also substantial improvement in the accuracy of template-based models. Other areas - model refinement, accuracy estimation, and the structure of protein assemblies - have again yielded interesting results. CASP13 placed increased emphasis on the use of sparse data together with modeling and chemical crosslinking, SAXS, and NMR all yielded more mature results. This paper summarizes the key outcomes of CASP13. The special issue of *PROTEINS* contains papers describing the CASP13 assessments in each modeling category and contributions from the participants.

INTRODUCTION

CASP (Critical Assessment of Structure Prediction) is a biennial community experiment to determine the state of the art in modeling protein structure. Participants are provided with amino acid sequences of target proteins, and build models of the corresponding three-dimensional structures. Submissions are compared with experiment by independent assessors. The experiment is double blinded - participants have no access to the experimental structures and assessors do not know the identity of those making the submissions. In addition to structure models, a number of other aspects of protein modeling are assessed as well: refinement of an approximate structure closer to the experimental one, estimates of the accuracy of an overall structure model and of each residue, modeling the structure of protein oligomers, the ability to improve models using a variety of sparse data types, and the accuracy of protein structure features related to deducing aspects of function. Here we summarize the current state of the art in each of these areas as determined in the CASP13 experiment (2018). Papers in this special issue of PROTEINS provide detailed analysis by the independent assessors in each modeling area and contributions from some of the more successful participants.

In CASP13, a total of 98 research groups from 21 countries tested 185 modeling methods and submitted over 57,000 predictions in six prediction categories, maintaining the previous high level of participation. There were 90 modeling targets for tertiary structure prediction (80 assessed), and 45 for quaternary structure prediction (42 assessed), including 13 hetero-complexes (12 assessed). The 80 tertiary structure modeling targets were parsed into 111 evaluation units, which were assessed as

1
2
3 separate targets ¹. Models are solicited in two initial stages. First on a short (72 hour)
4 time scale, intended for automated model building servers, then on a three-week time
5 scale, allowing time for more complex procedures and human input (though the latter
6 now appears to be rare). Relatively easy targets are only released for the server phase.
7
8 For evaluation, targets are divided into two main categories: template based (TBM), for
9 those where one or more structural templates can be identified by sequence search,
10 and template free (FM), for targets with no sequence detectable template. Some targets
11 fall into a grey area between these categories, and are labeled TBM/FM. One significant
12 change in target composition in CASP13 was from the ongoing revolution in high
13 resolution cryo electron microscopy (EM) ². There are EM targets for a total of six
14 complexes (four heteromeric) and one protein monomer. These targets tend to be
15 considerably larger than typical in CASP, but once parsed into evaluation domains are
16 less unusual ³. A fuller account of the procedures used in CASP is available in ⁴

35 **PROGRESS IN CASP13**

36
37 The overall accuracy of models improved dramatically in CASP13, especially for the
38 more difficult targets where comparative modeling cannot be used. Figure 1 shows the
39 trends in backbone accuracy for the best models received in each CASP, as a function
40 of target difficulty (the extent to which a target or target domain is related to the
41 sequence and structure of other proteins with already known structures ⁵ -
42 **supplementary figure S1 gives more data on target difficulty**). The vertical axis shows
43 backbone accuracy in terms of GDT_TS ^{6,7}). With this measure, 100% is exact
44 agreement of the C α co-ordinates of a model with those of the experimental
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5 structure, and a random model typically has a GDT_TS of between 20 and 30%. As a
6 rule of thumb, models with values greater than about 50% have correct overall topology,
7
8 and models with values greater than ~75% have many correct atomic level details. As
9
10 the trend lines show, early CASPs saw rapid improvement, but started from very low
11
12 accuracy. Until CASP13, most recent CASPs have shown very limited overall
13
14 improvement by this measure (though more fine-grained analysis shows improvement
15
16 in specific areas ⁴). Dramatically, the CASP13 trend line, instead of plunging
17
18 downwards, continues horizontally to the most difficult targets, with a sustained
19
20 GDT_TS greater than 60. **Supplementary figure S2 shows a similar (though not quite so**
21
22 **pronounced) CASP13 trend for average GDT_TS over the six best performing groups**
23
24 **on each target, indicating that multiple groups have improved substantially.** Below, we
25
26 discuss the methodological **advances** that drove this progress.
27
28
29
30
31
32
33
34

35 **PREDICTING CONTACTS IN PROTEIN STRUCTURES**

36
37 For a quarter of a century ⁸, attempts have been made to predict three-dimensional
38
39 contacts between residues in proteins, based on correlations in amino-acid substitutions
40
41 found in protein family protein sequence alignments ⁹.
42
43
44
45

46
47 For many years, the precision of these methods as measured in CASP was stalled at
48
49 20% or a little higher. **Figure 2 summarizes progress in recent CASPs.** Starting in
50
51 CASP11 (2014), and much more successfully in CASP12, statistical methods that
52
53 consider all pairs of residues simultaneously to address transitivity effects ⁹ began to
54
55
56
57
58
59
60

1
2
3 improve accuracy, resulting a best overall precision of 47% in CASP12 - almost
4 doubling in one CASP round - one of the biggest single improvements in any metric
5 seen in any CASP. Some predictors combined the statistical models with machine
6 learning, for instance ^{10,11}. But the key algorithmic advance appears to be the proper
7 treatment of transitivity. A limitation of those methods is that at least several hundred
8 appropriate sequences are needed to produce accurate predictions.
9
10
11
12
13
14
15
16
17
18

19 In CASP13, there is another large advance in precision, to 70%, again with several
20 groups delivering similar performance. This time it is clear the improvement came from
21 the use of deep neural network methods (discussed further in a CASP13 special issue
22 paper ¹²). These techniques have of course been very effective in other areas,
23 particularly image analysis ^{13 14} and speech recognition ¹⁵. Contact prediction uses a
24 similar methodology, treating the contact matrix (an L by L matrix for a sequence length
25 L, with 1 for elements representing contacting residues pairs and 0 for non-contacting
26 ones) as an image. The network is trained on a large set of known structures, typically
27 with multiple sequence alignment information, secondary structure prediction, co-
28 evolution analysis, and related features as input and the contact matrices as output.
29 Input of information for a new protein then generates an approximate contact map.
30 These methods were already being tested in CASP12 and promising benchmarking has
31 since been published ¹⁶. But as is often the case, they took some time to mature to the
32 point where improvements in performance are clearly measurable (very clearly in this
33 instance!).
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 Although the data representation in the advanced statistical methods and deep learning
4 approaches are very different, both rely on correlations in amino acid substitutions for
5 contacting residue pairs. As a result, a limitation of both is the need for a substantial
6 depth of sequence alignment. The effect of this can be seen in figure 3, where trend
7 lines for contact precision slope upwards as a function of normalized alignment depth.
8 But this dependency is greatly reduced with the CASP13 deep learning methods,
9 resulting in higher accuracy over a wide range of alignment depths. In CASP13,
10 inclusion of metagenomics sequence data increased alignment depth for some targets.
11 For example, metagenomics data as described in ^{17 18} increases alignment depth for
12 two free modeling targets from marginally adequate (less than 1L) to greater than 2L.
13 But generally, addition of these data has had only a modest impact so far.

30 **TEMPLATE FREE MODELING**

31
32 In CASP13, the largest improvement in model accuracy is for the most difficult, free
33 modeling, targets (Figure 1, right hand side) where no structural template could be
34 detected using sequence. Figure 4 shows an example for a free modeling target where
35 a number of groups produced good models.

36
37
38 If a sufficiently reliable set of contacts are predicted, these can be used as restraints to
39 obtain more accurate three-dimensional models. Figure 5 shows the relationship
40 between main chain accuracy and normalized alignment depth for template free
41 modeling targets in the most recent CASPs. There is a strong dependency of accuracy
42 on alignment depth, consistent with the major jumps in contact performance driving the

1
2
3 3D improvement for free modeling targets. The trend line for CASP13 is well above that
4 for CASP12, consistent with the more accurate contact predictions from deep learning.
5
6 In CASP13, all FM targets with N_{eff}/L greater than 1 (effective sequence alignment
7 depth equal or greater than the length of the target) have a GDT_TS greater than 50,
8 indicating a correct topology. The majority of targets with $N_{\text{eff}}/L > 0.1$ also have GDT_TS
9
10 >50. As discussed later, a number of the less accurate models are affected by inter-
11 molecular protein interactions, something current methods are not able to handle.
12
13 (Earlier CASPs already showed a link between contacts and 3D structure accuracy ⁴,
14 but not nearly to this extent).
15
16
17
18
19
20
21
22
23
24
25

26 Part of the three-dimensional accuracy improvement in CASP13 comes from not only
27 more accurate prediction of contacts but also prediction of inter-residue distances at a
28 range of thresholds, something deep neural networks are capable of and the statistical
29 methods are not. Approaches vary ^{19,20}, but in essence, 'contact' maps are predicted for
30 each of a set of inter-residue distances - say atoms within 6, 8, 10...20... Angstroms.
31
32 Properly normalized, these predictions allow an effective potential of mean force to be
33 derived between every pair of residues in a structure (that is, up to $L \cdot L/2 - L$ potentials
34 for an L residue long sequence). These potentials can then be used to drive a structure
35 folding procedure. One group, A7D from DeepMind²⁰, appear to have very successfully
36 deployed this technique, and had the most accurate results overall. It is not fully clear
37 what current deep learning procedures are 'learning' about protein architecture. The
38 ability to predict inter-residue distance probabilities as well as contacts suggests that the
39 topology of helices and beta-sheets and inter-secondary structure packing are captured
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 in some form. But so far there is no published analysis and indeed such an analysis
4 may not be meaningfully possible. There are many potential variations on the type of
5 residual networks currently being deployed, as well as other variables that have yet to
6 be evaluated, such as the best use of dilation and dropout ²¹. This and other aspects of
7 the methods will likely be further developed and refined by the next CASP and it will
8 very interesting to see how much further improvement can be made.
9
10
11
12
13
14
15
16
17
18

19 By definition, all free modeling targets are cases where no template structure can be
20 easily detected from sequence. But there may **nevertheless** be similar folds already
21 known. An alternative approach to using predicted contacts as restraints is to survey a
22 library of known structures, assessing which, if any, are most compatible with the
23 contact set. Supplementary Figure S3 shows the dependency of backbone accuracy on
24 the nearness of structural templates. Both CASP 12 and 13 show clear dependency, but
25 it is substantially reduced in CASP13, suggesting that template searches were less
26 competitive with folding algorithms, probably because greater contact accuracy and the
27 use of more general inter-residue distance prediction made the latter approach more
28 effective.
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44

45 **As always in CASP, care is needed to make sure that apparent progress is not an**
46 **artifact of different target difficulty in successive rounds. The insert in supplementary**
47 **figure S1 shows only very small differences in average free modeling (FM) target**
48 **difficulty in the most recent three CASPs. Additionally, supplementary figure S3 shows**
49 **that the average similarities of CASP 12 and 13 FM targets to structural templates are**
50
51
52
53
54
55
56
57
58
59
60

1
2
3 also nearly identical. The figure also shows the underlying CASP12 and 13 distributions
4 of target/template similarity values are close, further supporting similar target difficulty.
5
6
7
8
9
10

11 **TEMPLATE BASED MODELING**

12
13
14 As the number of experimentally determined structures grows, so does the number of
15 sequences for which it is possible to directly use structural templates to build a model,
16 using comparative modeling techniques. Figure 6 shows the relationship between
17 backbone accuracy of best models received for the template-based modeling category
18 in the three most recent CASPs. For the easiest targets (left hand side) with a high level
19 of sequence identity to a known structure there is no apparent improvement by this
20 measure. For harder targets, CASP12 is improved over CASP11, and CASP13 is
21 substantially further improved. Given the major accuracy advance in template-free
22 modeling from improved inter-residue distance predictions, an obvious question is
23 whether those methods are contributing here too. Supplementary Figure S4 shows the
24 relationship between backbone accuracy and alignment depth for the template-based
25 targets. CASP13 shows a mild dependence of accuracy on alignment depth, suggesting
26 that contacts are also playing some role in this regime. However, as expected for these
27 targets, almost all the alignments are deep enough for good contact prediction, which
28 may obscure a larger signal. Conversely, there may be a tendency for targets with
29 deeper alignment to have more useful templates, which would also tend to contribute to
30 the signal seen in the figure. Further support for contact prediction contributing to the
31 TBM improvement comes from a post-CASP analysis comparing the performance of
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 one method with and without contact prediction included - contact information leads to
4
5 higher accuracy for a number of targets (Yang Zhang, personal communication).
6
7

8
9
10 Typically, structure templates do not provide complete coverage of a target structure,
11
12 and overall accuracy depends not only on the appropriateness of the templates but also
13
14 on modeling of regions with no template. As Figure 7 shows, by this measure, there was
15
16 modest improvement between CASP5 in 2002 and CASP12, but a much larger
17
18 improvement between CASP12 and CASP13. As we have discussed before ⁴, earlier
19
20 improvements resulted from two principal modeling strategies: identification of other
21
22 templates with the correct structure in these regions or in some sense building these
23
24 substructures from scratch. Note that one would not expect the improvement to come
25
26 from better contact prediction: by definition these are regions that are not structurally
27
28 conserved within the protein family, and contact prediction generally relies on such
29
30 conservation. Though it is possible that more accurate modeling of the structurally
31
32 conserved regions creates a more accurate context for modeling non-template regions.
33
34
35
36
37
38
39

40 **REFINEMENT**

41
42 Models generated in both the template-free and template-based modeling sections of
43
44 CASP are approximate, and there is an end-game problem of further improving
45
46 agreement with experiment. To address this challenge, CASP includes a section on
47
48 refinement, where participants are provided with an initial model and asked to submit a
49
50 more accurate version. Performance has improved enormously over the succession of
51
52 experiments, from initial attempts that marginally improved some of the targets ²² to
53
54
55
56
57
58
59
60

1
2
3 impressive examples of structure correction in recent rounds ^{23 24,25}. But it is still the
4
5 case that no single method improves every target. In the three most recent CASPs ²³
6
7 ^{24,25}, the best groups have returned improvements for 60 to 70% of the targets. One
8
9 probable reason for limited performance is that the area suffers from a serious Red
10
11 Queen problem. Refinement methods that have been shown to be useful are
12
13 increasingly incorporated in initial modeling pipelines so that the starting point structures
14
15 supplied are already partly refined. Thus, methods must improve every round just to
16
17 appear as effective as previously. This particularly affects those who participate in both
18
19 initial structure modeling and refinement, as their models may be selected as starting
20
21 structures for refinement. As a consequence, metrics of improved structure accuracy
22
23 may not be very useful for measuring refinement performance. **Nevertheless**, the three
24
25 groups who have been consistently successful in recent CASPs do show modest
26
27 improvement in performance over successive rounds ²⁵. A more qualitative judgement
28
29 of progress is to analyze the type of structural features that are corrected. A few CASPs
30
31 ago, success with minor repositioning of secondary structure elements became
32
33 common, for example ²⁶. More recently, and especially in this CASP ²⁵, larger range
34
35 corrections (for example a 7.5 Angstrom helix shift in target R0981-D4) and significant
36
37 repacking (for example in R0974s1) have been achieved.

38
39
40
41
42
43
44
45
46
47 A persistent feature of refinement performance is that some targets are more refine-able
48
49 than others, and there are always some for which no group achieves a significant
50
51 improvement (10 to 15% of targets in recent CASPs). This has been a puzzle. The most
52
53 recent assessment provides partial insight into that phenomenon, with clear examples
54
55
56
57
58
59
60

1
2
3 where the non-inclusion of interactions with other protein domains or binding partners
4
5 limits accuracy ²⁵. As will be apparent again later, the modeling field has now advanced
6
7 to the point where there is a critical need for methods that effectively include the full
8
9 molecular environment.
10

11
12
13
14 As in previous CASPs, there are differences among the most successful refinement
15
16 methods. These range from a major focus on molecular dynamics ²⁷ to hybrid Monte
17
18 Carlo/sampling method ²⁸, to methods dominated by sampling ²⁹. But overall, there is an
19
20 increasing emphasis on the importance of conformational sampling.
21
22
23
24
25

26 **ACCURACY ESTIMATION**

27
28 Although modeling methods have improved enormously, models still greatly vary in
29
30 accuracy, both globally and in different parts of a structure. For any application it is
31
32 critical to know the accuracy of a model, and so CASP includes a section on estimating
33
34 model accuracy. As detailed in ³⁰, these predictions are very useful, and have been for
35
36 some rounds of CASP. The methods roughly fall into two categories - consensus
37
38 methods that rely on the degree to which a model is similar both overall and in detail to
39
40 others, and so-called 'single model' methods that use some form of structure-based
41
42 scoring function, often together with machine learning. Both approaches have
43
44 performed well and comparably in recent CASPs. In CASP13, the assessor observed a
45
46 relationship between the reliability of single model accuracy estimates and the methods
47
48 used to generate a model, particularly for models created with high reliance on contact
49
50 prediction related methods ³¹, apparently because of **method-specific** characteristics of
51
52
53
54
55
56
57
58
59
60

1
2
3 the models. To address this, some groups are now developing method-specific
4 accuracy estimation approaches. CASP already requires that models are accompanied
5 by detailed accuracy estimates, and in future, more emphasis may be placed on these.
6
7
8
9
10

11 **PROTEIN ASSEMBLIES**

12
13
14 As noted earlier, the enormous progress in domain and monomeric protein structure
15 modeling has highlighted the next bottleneck - limitations on initial model accuracy and
16 on refinement imposed by no or inadequate inclusion of the larger molecular
17 environment. More generally, most proteins exist as part of complexes, and function is
18 often dependent on the assembly. One aspect of the problem, the ability to dock
19 subunits of proteins to each other, has been evaluated by CASP's sister organization,
20 CAPRI (<http://www.capri-docking.org/>) since 2001. In the three most recent CASPs,
21 CAPRI and CASP have worked together to engage both communities in the broader
22 problem of protein assembly, including the use of protein models. Assessment papers
23 from both organizations are included ^{32,33} in the CASP13 special issue. Participation
24 from both communities increased over CASP12, showing growing interest in this
25 important problem. The CASP13 assessor found evidence of some improvement
26 compared to CASP12 ³³.
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46

47 Assembly is most successful when there is a template for the complex, presenting an
48 assembly comparative modeling problem ³⁴, and that was again demonstrated in this
49 CASP, where the CASP13 assessor concluded that availability of good assembly
50 templates usually results in good models. Next most successful is assembly of
51
52
53
54
55
56
57
58
59
60

1
2
3 complexes where the experimental structure of all the components is known, and there
4
5 is little or no conformational change associated with assembly. 'Free' docking methods
6
7 can be effective for these ³⁵, but these targets do not generally occur in CASP.
8
9
10 Relatively few complexes do not involve significant conformational changes of at least
11
12 side chains and local regions of structure, and without an assembly template, current
13
14 methods are unable to cope with these situations. Add to this the complications of
15
16 working with approximate models for assembly components, and the problem is
17
18 daunting. As the CASP13 assembly assessor points out ³³, because of the importance
19
20 of conformational changes on assembly formation, the current standard procedure of
21
22 first building monomer structures in isolation and then attempting to dock them is
23
24 flawed. The assessor found seven of the heteromeric CASP13 assembly targets have
25
26 substantial interdependences between monomers, in a variety of ways. Figure 8 shows
27
28 an example for target H0953, an A3B1 multimer, where the trimer assembly generates
29
30 strong subunit interdependencies. In other targets (T0973, 991, 998), a helix is swapped
31
32 between subunits.
33
34
35
36
37
38
39

40 The obvious message is that successful assembly methods will have to take subunit
41
42 interdependences into account and not rely exclusively on modeling isolated subunits.
43
44 Following the emergence of powerful deep learning methods for monomers in CASP13,
45
46 there is intense interest in whether these approaches may be adaptable to the problem.
47
48 Of note in this regard is that the assembly assessor found prediction of inter-subunit
49
50 contacts in homo-assemblies was 'surprisingly successful' ³³. It's not clear how these
51
52 predictions were made, but they may be useful in identifying first, which regions are in
53
54
55
56
57
58
59
60

1
2
3 contact even in the absence of a good three-dimensional model of the monomers, and
4
5 second, possibly where there are conformational interdependencies. Deep learning
6
7 methods for image analysis have been shown to be very robust to noise¹², for example
8
9 see <https://clarifai.com/demo>. It will be very interesting to see in the next CASP whether
10
11 any examples of contact driven non-trivial assembly can be achieved, particularly with
12
13 the use of deep learning.
14
15
16
17
18

19 **DATA ASSISTED MODELING**

20
21 Even high-resolution experimental structures incorporate aspects of modeling, making
22
23 use of bond length and bond angle restrictions, avoidance of steric clashes, and
24
25 sometimes imposing reasonable electrostatic interactions. Lower resolution methods -
26
27 SAXS, chemical cross-linking coupled with mass spectrometry, and sparse NMR -
28
29 depend critically on modeling to make maximum use of the experimental data. Since
30
31 CASP11, CASP has experimented with providing these types of data to participants
32
33 after data-free models have been obtained, and assessing whether the sparse data can
34
35 be effective in increasing model accuracy. This area has great promise, but is proving
36
37 challenging to successfully implement. First, additional data must be generated by the
38
39 experimental community. A number of groups have been very co-operative and
40
41 supportive, but still, protein samples are only available for a few targets, and those
42
43 targets may not be ideal. Second, it requires specialized expertise to make optimum use
44
45 of these types of data. In spite of vigorous efforts to provide webinars and other material
46
47 in CASP13, rather few predictors have so far moved into this area. Third, because of
48
49 low participation, the newer contact prediction and deep learning methods were not
50
51
52
53
54
55
56
57
58
59
60

1
2
3 used together with the sparse data. As a result, more accurate models were obtained
4
5 without use of the experimental information.
6
7
8
9

10 *NMR*: The Gaetano Montelione and Antonio Rosato groups produced simulated sparse
11
12 NMR data for 12 proteins or protein domains, in the form of ambiguous interatomic
13
14 contact lists, chemical shifts, and RDCs . They also provided real sparse NMR data for
15
16 one CASP target, N1008. The data provided are intended be similar to that available for
17
18 large structures, and are insufficient for structure solution by standard NMR techniques
19
20 ³⁶. Nine groups took part in NMR-assisted modeling, three of whom were controls from
21
22 the Montelione lab. Generally, the models submitted are of similar accuracy to the best
23
24 unassisted models received, but for one target, N0981-D2, a model built using the
25
26 simulated data is over 30 GDT_TS units better than the best unassisted, a notable
27
28 success. The target with experimental NMR data, N1008, is a designed protein ³⁷, and
29
30 even though there are no sequence homologs, was very accurately modeled by a
31
32 number of groups, without the use of the data. As a result, the NMR assisted models
33
34 were not as accurate. That outcome illustrates how tricky it is to choose targets in which
35
36 to invest experimental effort. Of the nine groups submitting NMR assisted structures,
37
38 two (Laufer and Meiler) had markedly better results than the controls. Laufer used
39
40 molecular dynamics with a filtering technique to remove non-consistent restraints ³⁸. We
41
42 hope the encouraging results will lead to larger scale participation in this category in
43
44 CASP14.
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 SAXS: Data were generated for 11 targets in all, including 7 complexes. This was a very
4
5 impressive contribution from the experimental group (Susan Tsutakawa, Greg Hura and
6
7 John Tainer). 13 groups submitted models using these data. A number of teething
8
9 troubles that plagued the first SAXS experiment in CASP12 ³⁹ were avoided or greatly
10
11 reduced, so that a more meaningful assessment of the contribution from the SAXS data
12
13 could be made.
14
15

16
17
18
19 For no target was the best data assisted structure as accurate as the best unassisted,
20
21 although there are some examples of improved inter-subunit relationships. Again, the
22
23 issue here may be the relatively low participation, so that the results are not necessarily
24
25 representative of the newest unassisted methods. Several groups did develop
26
27 interesting pipelines incorporating SAXS data, and as is often the case in CASP, it may
28
29 take further iterations before the power of these can be properly assessed. Methods
30
31 typically take the full set of server models available for a target and evaluate the fit of
32
33 these the SAXS data, often also using additional accuracy estimations. One group also
34
35 investigated the use of normal mode driven structure changes.
36
37
38
39
40
41

42 Comparison of the SAXS envelopes with the X-ray structures suggests that for up to
43
44 half the targets there could be differences between the solution and crystal structures.
45
46 Such differences may in principle limit the accuracy that modeling can achieve using the
47
48 crystal structure as the gold standard. But there could be a number of explanations for
49
50 the discrepancies, including sequence differences between crystal and SAXS samples,
51
52 more disorder in solution, and the inherent difficulties of interpreting SAXS data.
53
54
55
56
57
58
59
60

1
2
3
4
5
6 *Chemical crosslinking and mass spectrometry:* Experimental data derived from two
7 different cross-linking chemistries were provided. Alexander Leitner and Ruedi Abersold
8 (ETH, Zurich) used a predominantly Lys primary amine-oriented (BS3) chemistry and
9 Adam Belsom and Juri Rappsilber (University of Edinburgh and Berlin Technical
10 University) employed the heterobifunctional, photoactivable cross-linking chemistry.
11
12 One data set was also provided by Marcus Hartmann, using disuccinimidyl suberate
13 (DSS) chemistry. Altogether data were collected for eight different protein samples,
14 including three hetero-multimers, two homo-multimers, and three single chain proteins.
15
16 Based on these data, five heteromeric targets and 17 single-sequence targets
17 (monomers or subunits of multimers) were released for prediction. An analysis by the
18 assessor ⁴⁰ shows that a surprisingly high fraction of cross-links appear not to be
19 compatible with the targets' X-ray structures (27-47%). In total, 14 prediction groups
20 participated. For the monomeric protein domains and subunits, there are many
21 instances where the data-assisted models are more accurate than the corresponding
22 un-assisted models from the same participants. But this comparison likely provides too
23 optimistic a view, since the groups with the greatest improvement started from scratch
24 in utilizing the cross-link data, ignoring their initial submissions, and instead making use
25 of the full set of server models available for each target. A more stringent criterion,
26 comparing the data-assisted models to the best received for each target from any group
27 generally shows all the cross-link assisted models are less accurate. For the complete
28 complexes, there is one instance of a significant improvement, for X0957, a bacterial
29 toxin/immunity protein complex, where several inter-subunit crosslinks helped select a
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 more appropriate overall assembly³³, an encouraging result. The results illustrate both
4 the promise and the challenges of using cross-link information to improve models. Many
5 cross-links are misleading in that they conflict with the corresponding X-ray structure
6 and some can be false positives. Further, the large variation in distance between
7 crosslinked residues ⁴⁰ makes the technique inherently low-resolution, and so likely best
8 suited to complexes, as the result for X0957 illustrates.
9
10
11
12
13
14
15
16
17
18

19 Pilot experiments were also conducted with FRET data on one target (generously
20 provided by Claus Seidel and Mykola Dimura) and SANS on another (provided by Anne
21 Martel). We expect to include more data of these types in CASP14.
22
23
24
25
26
27

28 **DISCUSSION**

29
30 Successful use of relatively standard deep learning techniques for predicting not only
31 three-dimensional contacts but more general inter-residue distance distributions was the
32 outstanding development of CASP13 and caused much excitement and creative
33 thinking at the CASP meeting. There is an expectation that similar approaches can be
34 applied to other areas of structure modeling, particularly improved estimates of both
35 global and local model accuracy, improved model refinement by allowing focus on
36 regions of maximum error, and recognition and prediction of protein-protein interfaces.
37 We will have to wait until CASP14 in 2020 to see which of these bears fruit. CASP14 is
38 also likely to see further progress in 3D structure modeling based on deep learning
39 approaches. Several modeling groups are developing servers that will make the new
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 methods available to broader community. It's likely that the impact on the usefulness of
4
5 modeling will be large.
6
7
8
9

10 The major progress in modeling domains and monomeric proteins without direct use of
11 a structural template is a very significant break-through: for these proteins, the long-
12 standing problem of 'protein folding' (generating a model with the correct topology) is
13 essentially solved, albeit in way that early work in the field never imagined. An
14 alignment with of at least a few dozen sequences is usually needed for the methods to
15 work, but most protein families are now that large. Success with topology prediction has
16 increased focus on the remaining problems - we are still a long way from the accuracy
17 of X-ray structures or from enabling structure-based drug design, and more complex
18 structures are the norm in biology. CASP already has well-established categories in the
19 relevant areas, particularly refinement and protein assemblies, and as already noted it
20 will be exciting to see what impact deep learning and related approaches have on
21 those. Other areas, such as conformational change in response to ligand binding and
22 environmental conditions, remain future challenges.
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41

42 CASP continues to experiment with other aspects of modeling. Of note this round was
43 the expanded number of targets for which sparse experimental data were available.
44 Although the results in terms of more accurate models are not impressive, it is clear that
45 much more development is possible, and we have already seen several groups
46 introduce methods specifically tailored to particular data types. CASP continues to
47 investigate the best ways of assessing how effectively functional information can be
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 derived from models^{41,42}, and in this round, solicited assessment comments from those
4 who provided the prediction targets⁴³. An interesting development during CASP13 was
5 the introduction of a CASP commons experiment
6
7
8
9
10 (<http://predictioncenter.org/caspcommons/>). Biologists were canvased to identify a total
11 of 35 small proteins for which structure would be particularly useful for their research.
12
13 The Montelione group cloned and expressed these, with the goal of determining which
14 are suitable for NMR structure determination, and in parallel the CASP community was
15 invited to submit models. So far one experimental structure has been obtained³⁷. A new
16 round of modeling is now beginning, using the new free modeling methods from
17 CASP13.
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40

41 We plan to hold CASP14 in 2020, with a similar timetable to previous rounds. The
42 prediction season will be spring and summer, and the conference will be at the end of
43 the year. Details will be posted on the Prediction Center web site as they become
44 available.
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

ACKNOWLEDGMENTS

CASP is only possible through the generosity and support of three groups of people: the data providers, the assessors, and the participants. A total of 36 experimental groups from 14 countries provided targets for CASP13. Nine of these groups also provided material for the collection of SAXS, SANS, or cross-linking experimental data: Owen Davies (Newcastle University); Petr Leiman (UTMB, Galveston, TX) and Matthew Dunne (ETH Zurich); Andrew Lovering (University of Birmingham); Karolina Michalska and Andrzej Joachimiak (Argonne National Lab, Chicago, IL); Anne Martel (Institute

1
2
3 Laue-Langevin, Grenoble); Jose Henrique Pereira (LBNL, Berkeley, CA); Mark van
4
5 Raaij (CNB-CSIS, Madrid); Lindsey Spiegelman (University of California, San Diego,
6
7 CA); Chin-Lin Tsai (MD Anderson Cancer Center, University of Texas), making that
8
9 aspect of the experiment possible. SAXS data were acquired for ten protein samples by
10
11 Susan Tsutakawa, Greg Hura and John Tainer (LBNL, Berkeley). Cross-link data were
12
13 acquired for seven samples by Esben Trbjerg and Alexander Leitner (Institute of
14
15 Molecular Systems Biology, ETH, Zurich), and for four samples by Adam Belsom and
16
17 Juri Rappsilber (Technische Universitat, Berlin; University of Edinburgh). Marcus
18
19 Hartmann (Max Planck Institute for Developmental Biology, Tuebingen) provided SAXS
20
21 and cross-linking data for a target they had submitted. The data for the NMR study were
22
23 generated by Antonio Rosatto (University of Florence), Janet Huang and Gaetano
24
25 Montelione (Rutgers). The material for the FRET study was produced by Claus Seidel
26
27 and Mykola Dimura (University of Dusseldorf), who also performed the FRET analysis.
28
29 We are grateful to Emily Tai (NCI/NIH) for her major contributions to the CASP
30
31 commons experiment.
32
33
34
35
36

37
38 We once again thank the assessment teams for their thorough and insightful
39
40 analyses: Randy Read and Tristan Croll (Cambridge University) for template-based
41
42 modeling and refinement assessment; Luciano Abriata and Matteo Dal Peraro (EPFL,
43
44 Lausanne) for free modeling assessment; Jose Duarte (RCSB, San Diego, CA) for
45
46 assembly assessment; Andras Fiser (Albert Einstein College of Medicine, New York, NY)
47
48 for contact and cross-link assisted assessment; Greg Hura and Susan Tsutakawa (LBNL,
49
50 Berkeley, CA) for SAXS-assisted assessment; Gaetano Montelione (Rutgers, NJ) for
51
52 NMR-assisted assessment; Chaok Seok (Seoul National University, South Korea) for
53
54
55
56
57
58
59
60

1
2
3 model accuracy assessment; and Lisa Kinch for her work on target structure analysis. As
4
5 always, for participants, it takes courage to expose their methods to such intense and
6
7 public scrutiny. We greatly appreciate the 98 research groups who submitted their work
8
9 to this CASP. We again thank *PROTEINS* for providing a mechanism for peer reviewed
10
11 publication of the outcome of the experiment.
12
13

14 The CASP Prediction Center at UC Davis is supported by a grant from the US
15
16 National Institute of General Medical Sciences (NIGMS/NIH), R01GM100482 to KF.
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

REFERENCES

1. Kinch LN, Kryshatafovych A, Monastyrskyy B, Grishin NV. CASP13 target classification into tertiary structure prediction categories. *Proteins* 2019.
2. Egelman EH. The Current Revolution in Cryo-EM. *Biophys J* 2016;110(5):1008-1012.
3. Kryshatafovych A, al. e. Cryo-EM targets in CASP13: overview and evaluation of results. [CASP13 special issue] 2019.
4. Moulton J, Fidelis K, Kryshatafovych A, Schwede T, Tramontano A. Critical assessment of methods of protein structure prediction (CASP)-Round XII. *Proteins* 2018;86 Suppl 1:7-15.
5. Kryshatafovych A, Fidelis K, Moulton J. CASP10 results compared to those of previous CASP experiments. *Proteins* 2014;82 Suppl 2:164-174.
6. Zemla A, Venclovas, Moulton J, Fidelis K. Processing and evaluation of predictions in CASP4. *Proteins* 2001;Suppl 5:13-21.
7. Zemla A. LGA: A method for finding 3D similarities in protein structures. *Nucleic Acids Res* 2003;31(13):3370-3374.
8. Gobel U, Sander C, Schneider R, Valencia A. Correlated mutations and residue contacts in proteins. *Proteins* 1994;18(4):309-317.
9. de Juan D, Pazos F, Valencia A. Emerging methods in protein co-evolution. *Nat Rev Genet* 2013;14(4):249-261.

10. Koscioliek T, Jones DT. Accurate contact predictions using covariation techniques and machine learning. *Proteins* 2016;84 Suppl 1:145-151.
11. Wang S, Sun S, Xu J. Analysis of deep learning methods for blind protein contact prediction in CASP12. *Proteins* 2018;86 Suppl 1:67-77.
12. Kadathil S, Greener J, Jones D. Recent developments in deep learning applied to protein structure prediction. [CASP13 special issue] 2019.
13. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak J, van Ginneken B, Sanchez CI. A survey on deep learning in medical image analysis. *Med Image Anal* 2017;42:60-88.
14. Biswas M, Kuppili V, Saba L, Edla DR, Suri HS, Cuadrado-Godia E, Laird JR, Marinho RT, Sanches JM, Nicolaides A, Suri JS. State-of-the-art review on deep learning in medical imaging. *Front Biosci (Landmark Ed)* 2019;24:392-426.
15. Afouras T, Chung J, Senior A, Vinyals O, Zisserman A. Deep audio-visual speech recognition. *IEEE transactions on pattern analysis and machine intelligence* 2018.
16. Wang S, Sun S, Li Z, Zhang R, Xu J. Accurate De Novo Prediction of Protein Contact Map by Ultra-Deep Learning Model. *PLoS Comput Biol* 2017;13(1):e1005324.
17. Steinegger M, Mirdita M, Soding J. Protein-level assembly increases protein sequence recovery from metagenomic samples manifold. *Nat Methods* 2019;16(7):603-606.

18. Kandathil SM, Greener JG, Jones DT. Prediction of interresidue contacts with DeepMetaPSICOV in CASP13. *Proteins* 2019.
19. Xu J. Distance-based protein folding powered by deep learning. *Proc Natl Acad Sci U S A* 2019.
20. Senior A, et al. Protein structure prediction using multiple deep neural networks in CASP13. *PROTEINS 2019*;[CASP13 special issue].
21. Adhikari B. DEEPCON: Protein Contact Prediction using Dilated Convolutional Neural Networks with Dropout. *bioRxiv* 2019.
22. MacCallum JL, Hua L, Schnieders MJ, Pande VS, Jacobson MP, Dill KA. Assessment of the protein-structure refinement category in CASP8. *Proteins* 2009;77 Suppl 9:66-80.
23. Modi V, Dunbrack RL, Jr. Assessment of refinement of template-based models in CASP11. *Proteins* 2016;84 Suppl 1:260-281.
24. Hovan L, Oleinikovas V, Yalinca H, Kryshtafovych A, Saladino G, Gervasio FL. Assessment of the model refinement category in CASP12. *Proteins* 2018;86 Suppl 1:152-167.
25. Read RJ, Sammito MD, Kryshtafovych A, Croll TI. Evaluation of model refinement in CASP13. *Proteins* 2019.
26. Nugent T, Cozzetto D, Jones DT. Evaluation of predictions in the CASP10 model refinement category. *Proteins* 2014;82 Suppl 2:98-111.
27. Heo L, Arbour CF, Feig M. Driven to near-experimental accuracy by refinement via molecular dynamics simulations. *Proteins* 2019.

- 1
2
3 28. Park H, Lee GR, Kim DE, Anishchenko I, Cong Q, Baker D. High-
4 accuracy refinement using Rosetta in CASP13. *Proteins* 2019.
5
6
7
- 8 29. Lee GR, Won J, Heo L, Seok C. GalaxyRefine2: simultaneous
9 refinement of inaccurate local regions and overall protein structure.
10
11
12
13
14
15
16
17
18
19
20
21
22
23 30. Kryshtafovych A, Monastyrskyy B, Fidelis K, Schwede T, Tramontano
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
31. Won J, Baek M, Monastyrskyy B, Kryshtafovych A, Seok C. Assessment
of Protein Model Structure Accuracy Estimation in CASP13: Challenges
in the Era of Deep Learning. *PROTEINS* 2019;[CASP13 special issue].
32. M. L, al. e. Blind prediction of homo- and hetero-protein complexes:
The CASP13-CAPRI experiment. *PROTEINS* 2019;[CASP13 special
issue].
33. Guzenko D, Lafita A, Monastyrskyy B, Kryshtafovych A, Duarte JM.
Assessment of protein assembly prediction in CASP13. *Proteins* 2019.
34. Anishchenko I, Kundrotas PJ, Vakser IA. Modeling complexes of
modeled proteins. *Proteins* 2017;85(3):470-478.
35. Vakser IA. Protein-protein docking: from interaction to interactome.
Biophys J 2014;107(8):1785-1793.

- 1
2
3 36. Montelione G, Others. Assessment of Guided Protein Structure
4 Prediction using Sparse NMR Data in CASP13. [CASP13 special issue]
5 2019.
6
7
8
9
10 37. Koepnick B, Flatten J, Husain T, Ford A, Silva DA, Bick MJ, Bauer A, Liu
11 G, Ishida Y, Boykov A, Estep RD, Kleinfelter S, Norgard-Solano T, Wei
12 L, Players F, Montelione GT, DiMaio F, Popovic Z, Khatib F, Cooper S,
13 Baker D. De novo protein design by citizen scientists. *Nature*
14 2019;570(7761):390-394.
15
16
17
18
19
20
21
22 38. Robertson JC, Nassar R, Liu C, Brini E, Dill KA, Perez A. NMR-assisted
23 protein structure prediction with MELDxMD. *Proteins* 2019.
24
25
26
27 39. Ogorzalek TL, Hura GL, Belsom A, Burnett KH, Kryshtafovych A, Tainer
28 JA, Rappsilber J, Tsutakawa SE, Fidelis K. Small angle X-ray scattering
29 and cross-linking for data assisted protein structure prediction in CASP
30 12 with prospects for improved accuracy. *Proteins* 2018;86 Suppl
31 1:202-214.
32
33
34
35
36
37
38
39 40. Fajardo J, al. e. Assessment of chemical-crosslink-assisted protein
40 structure modeling in CASP13. [CASP13 special issue] 2019.
41
42
43
44 41. Liu T, Ish-Shalom S, Torng W, Lafita A, Bock C, Mort M, Cooper DN,
45 Bliven S, Capitani G, Mooney SD, Altman RB. Biological and functional
46 relevance of CASP predictions. *Proteins* 2018;86 Suppl 1:374-386.
47
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3 42. Huwe PJ, Xu Q, Shapovalov MV, Modi V, Andrade MD, Dunbrack RL, Jr.
4
5 Biological function derived from predicted structures in CASP11.
6
7 Proteins 2016;84 Suppl 1:370-391.
8
9
10 43. Lepore A, others. Target highlights in CASP13: experimental target
11
12 structures through the eyes of their authors. [CASP13 special issue]
13
14 2019.
15
16
17 44. Monastyrskyy B, D'Andrea D, Fidelis K, Tramontano A, Kryshtafovych
18
19 A. New encouraging developments in contact prediction: Assessment
20
21 of the CASP11 results. Proteins 2016;84 Suppl 1:131-144.
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

FIGURE CAPTIONS

Figure 1:

Trend lines of backbone accuracy for the best models in each of the 13 CASP experiments. Individual target points are shown for the two most recent experiments. The accuracy metric, GDT_TS, is a multiscale indicator of the closeness of the C α atoms in a model to those in the corresponding experimental structure. Target difficulty is based on sequence and structure similarity to other proteins with known experimental structures (see ⁵ for details). There is a striking improvement in model accuracy in CASP13 (top black line), particularly for the more difficult targets.

Figure 2:

Best contact prediction precision in recent CASPs. CASPs 9 and 10 continued a long trend of low precision. CASP11 shows a small advance, while the two most recent, CASP12 and 13, show dramatic improvements. In CASPs 11 and 12 progress is the result of more sophisticated statistical models, together with largely conventional machine learning. The further jump in CASP13 is the result of the effective deployment of deep learning methods. (Average fraction of correctly predicted contacts for the most confidently predicted L/5 contacts 24 or more residues apart in the sequence, where L is target length. Free modeling targets, average for the best performing group in each CASP. Contacting residue pairs defined as those with less than 8 Angstroms between C β atoms).

1
2
3 Figure 3:
4

5 Contact prediction precision trend lines as a function of sequence alignment depth and
6 target length. In CASP13 there is a reduced dependency on alignment depth, resulting
7 in more accurate results for shallow alignments as well as higher precision overall.

8 Strikingly, for ten out of the 31 free modeling targets, the best predictions achieved
9 100% precision for this subset of contacts (see figure 2 for definitions). The effective
10 alignment depth, N_{eff} , includes metagenomic sequences compiled as described in ^{17 18}.

11 N_{eff} was calculated using a 90% sequence identity cutoff and a minimum of 60%
12 sequence coverage (details in ⁴⁴).
13
14
15
16
17
18
19
20
21
22
23
24
25

26 Figure 4:
27

28 Crystal structure of a 354 residue domain of a free modeling target (T0969-D1),
29 ESKIMO 1, a probable xylan acetyltransferase, PDB 6CCI (left panel) and the most
30 accurate CASP model (right panel). Most of the structure core is modeled to a $C\alpha$
31 accuracy of better than 1 (cyan) or 2 Angstroms (green). Irregular loop regions are less
32 accurate (yellow, better than 4 Angstroms or orange, up to 8 Angstroms error. Some
33 residues (red) in external loops have larger errors.
34
35
36
37
38
39
40
41
42
43
44

45 Figure 5:
46

47 Best model main chain accuracy (GDT_TS) as a function of sequence alignment depth
48 and target length for CASPs 12 and 13. Accuracy depends on alignment depth, as is
49 expected if the result is dominated by contact prediction accuracy and related
50
51
52
53
54
55
56
57
58
59
60

1
2
3 advances. Across all alignment depths, CASP13 models are on average more accurate
4
5 than those in CASP12.
6
7
8
9

10 Figure 6:

11
12 Best model backbone accuracy (GDT_TS) as a function of target difficult for template-
13
14 based models in recent CASPs. CASP13 shows a marked improvement in accuracy
15
16 compared to previous CASP. Targets are those where there is clear sequence
17
18 relationship to a known structure (termed TBM) and those with a marginal relationship
19
20 (TBM/FM).
21
22
23
24
25

26 Figure 7:

27
28 Trend lines for the fraction of non-principal template ('loop') residues correctly modeled.
29
30 There is a substantial improvement in CASP13. (Best models received for each target,
31
32 3.8 Angstrom C α atom agreement or better considered correct, TBM and TBM/FM
33
34 targets).
35
36
37
38
39

40 Figure 8:

41
42 Part of the experimental structure of target H0953 (PDB 6F45), the adhesin tip complex
43
44 of a bacteriophage tail fiber, illustrating subunit structure interdependence. One of the
45
46 two protein chains contributing to this assembly forms a trimer (colored red, green and
47
48 blue), with the N terminal five strand beta sheets of the three monomers packing against
49
50 each other. The C terminal three beta strands of each monomer inter-digitate with each
51
52 other. The C terminal stands also form an interface with the helical end of another
53
54
55
56
57
58
59
60

1
2
3 subunit (green). Impressively, in spite of the apparent interdependency of the five-strand
4
5 beta-sheets, accurate models were returned for that part of the structure. But failure to
6
7 consider the even more intimate subunit interactions of the three N terminal strands
8
9 resulted in incorrect models for that subdomain.
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

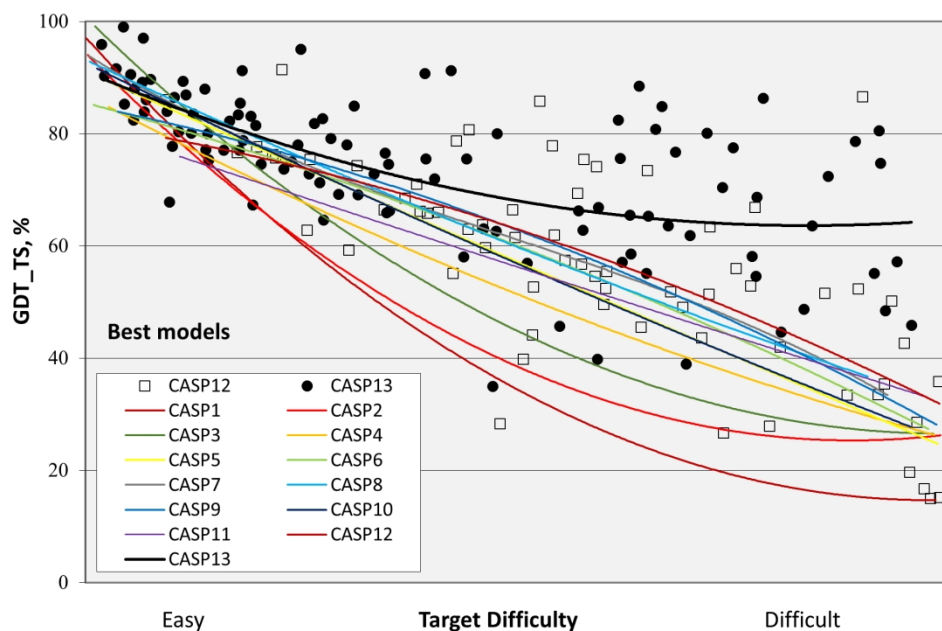


Figure 1: Trend lines of backbone accuracy for the best models in each of the 13 CASP experiments. Individual target points are shown for the two most recent experiments. The accuracy metric, GDT_TS, is a multiscale indicator of the closeness of the C α atoms in a model to those in the corresponding experimental structure. Target difficulty is based on sequence and structure similarity to other proteins with known experimental structures (see 5 for details). There is a striking improvement in model accuracy in CASP13 (top black line), particularly for the more difficult targets.

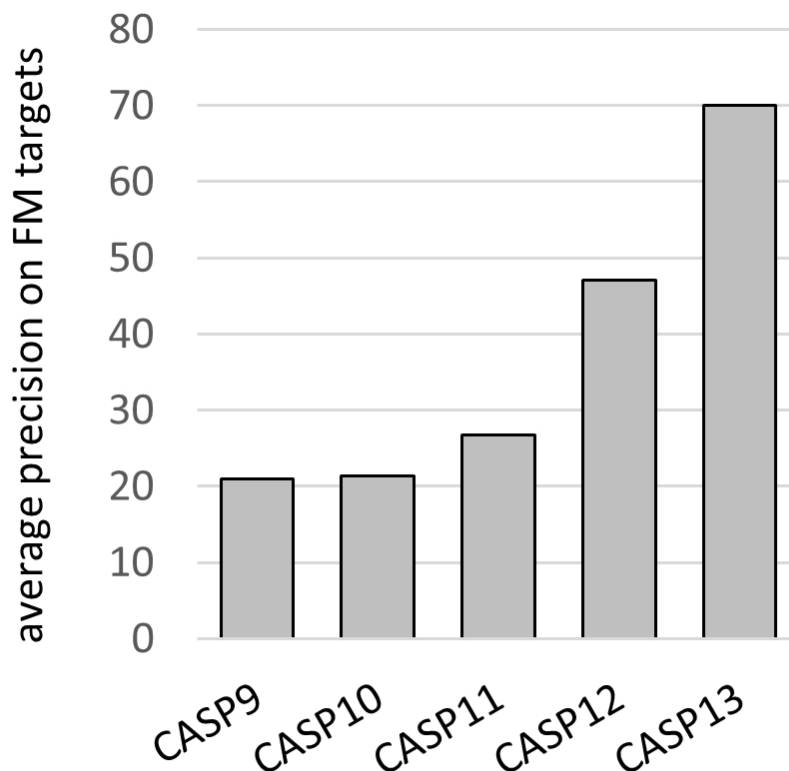


Figure 2: Best contact prediction precision in recent CASPs. CASPs 9 and 10 continued a long trend of low precision. CASP11 shows a small advance, while the two most recent, CASP12 and 13, show dramatic improvements. In CASPs 11 and 12 progress is the result of more sophisticated statistical models, together with largely conventional machine learning. The further jump in CASP13 is the result of the effective deployment of deep learning methods. (Average fraction of correctly predicted contacts for the most confidently predicted $L/5$ contacts 24 or more residues apart in the sequence, where L is target length. Free modeling targets, average for the best performing group in each CASP. Contacting residue pairs defined as those with less than 8 Angstroms between $C\alpha$ atoms).

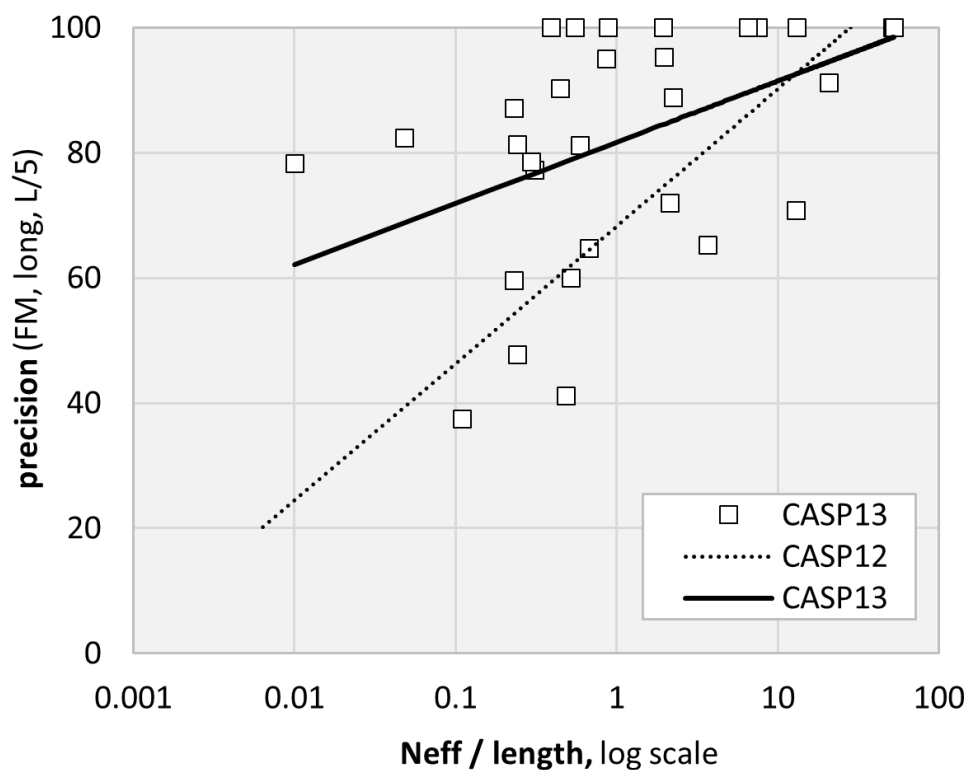


Figure 3: Contact prediction precision trend lines as a function of sequence alignment depth and target length. In CASP13 there is a reduced dependency on alignment depth, resulting in more accurate results for shallow alignments as well as higher precision overall. Strikingly, for ten out of the 31 free modeling targets, the best predictions achieved 100% precision for this subset of contacts (see figure 2 for definitions). The effective alignment depth, Neff, includes metagenomic sequences compiled as described in 17 18. Neff was calculated using a 90% sequence identity cutoff and a minimum of 60% sequence coverage (details in 44).

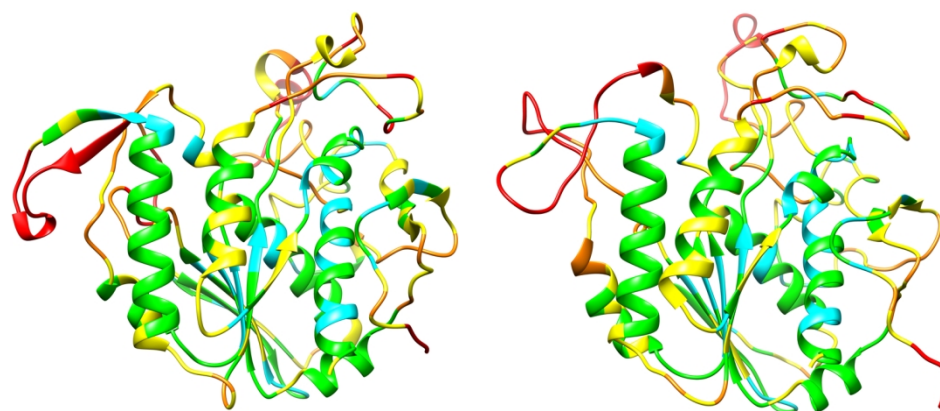


Figure 4: Crystal structure of a 354 residue domain of a free modeling target (T0969-D1), ESKIMO 1, a probable xylan acetyltransferase, PDB 6CCI (left panel) and the most accurate CASP model (right panel). Most of the structure core is modeled to a C α accuracy of better than 1 (cyan) or 2 Angstroms (green). Irregular loop regions are less accurate (yellow, better than 4 Angstroms or orange, up to 8 Angstroms error. Some residues (red) in external loops have larger errors.

153x70mm (300 x 300 DPI)

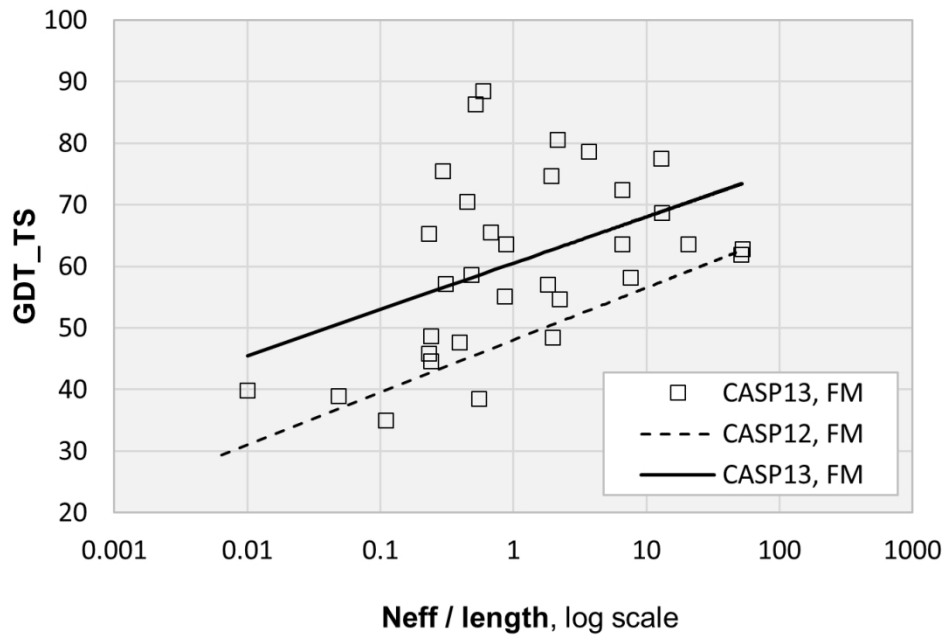


Figure 5: Best model main chain accuracy (GDT_TS) as a function of sequence alignment depth and target length for CASPs 12 and 13. Accuracy depends on alignment depth, as is expected if the result is dominated by contact prediction accuracy and related advances. Across all alignment depths, CASP13 models are on average more accurate than those in CASP12.

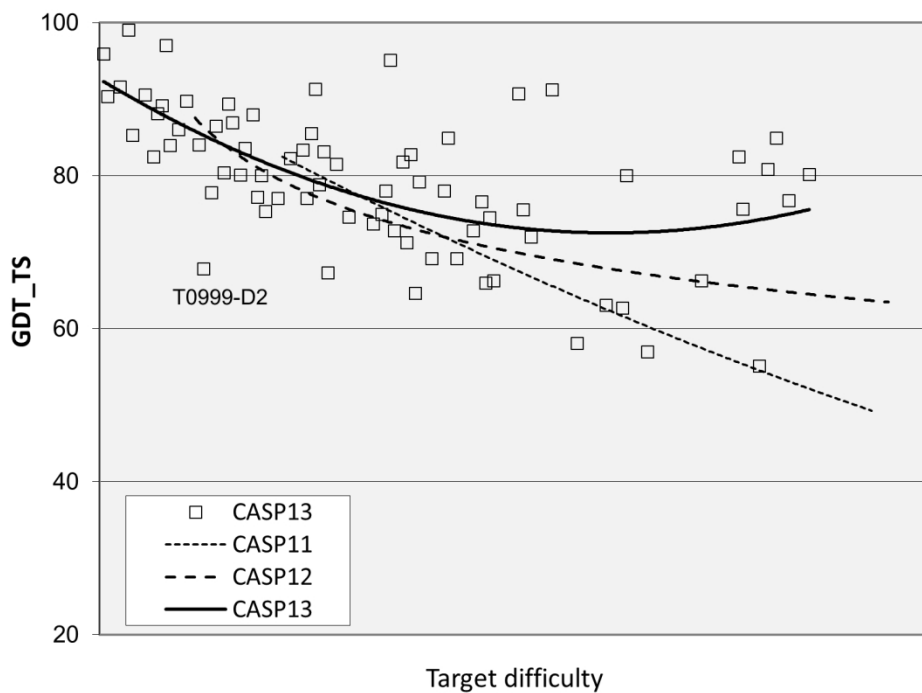


Figure 6: Best model backbone accuracy (GDT_TS) as a function of target difficulty for template-based models in recent CASPs. CASP13 shows a marked improvement in accuracy compared to previous CASP. Targets are those where there is clear sequence relationship to a known structure (termed TBM) and those with a marginal relationship (TBM/FM).

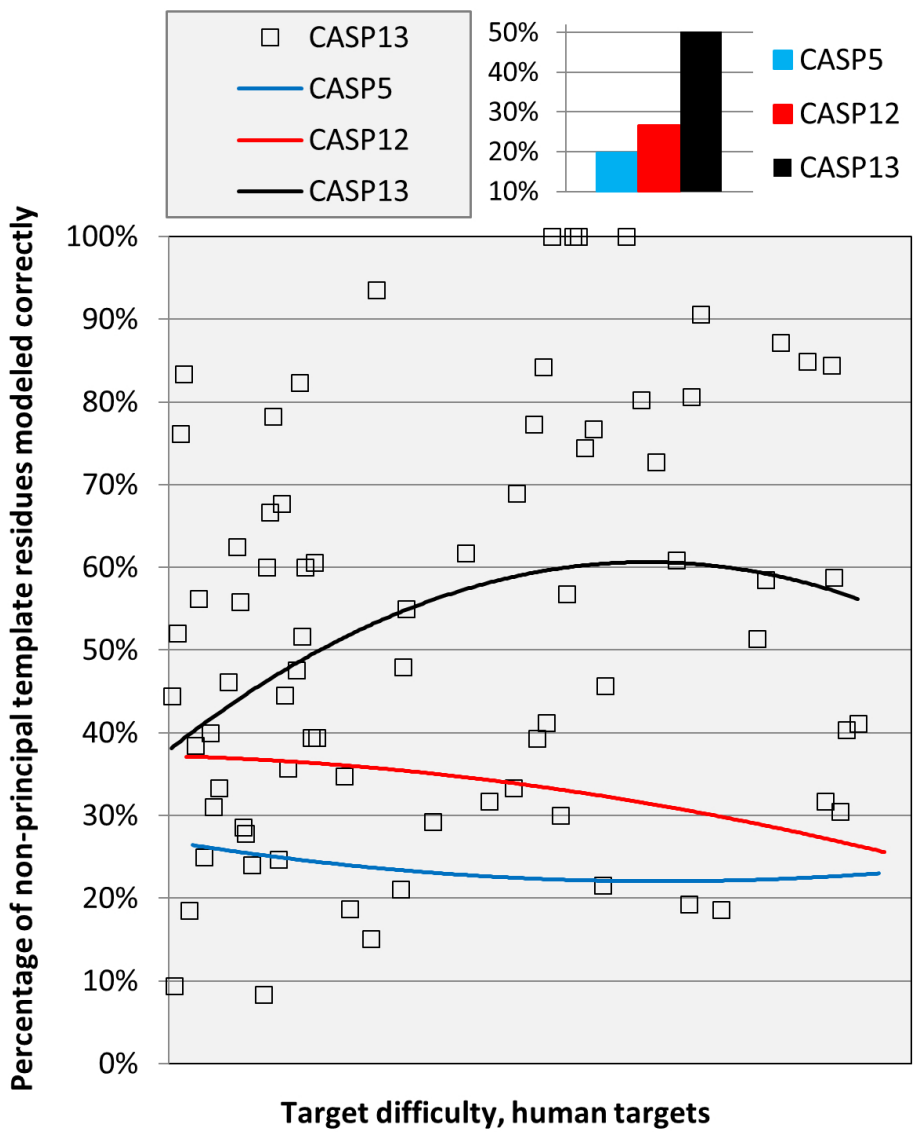


Figure 7: Trend lines for the fraction of non-principal template ('loop') residues correctly modeled. There is a substantial improvement in CASP13. (Best models received for each target, 3.8 Angstrom Ca atom agreement or better considered correct, TBM and TBM/FM targets.)

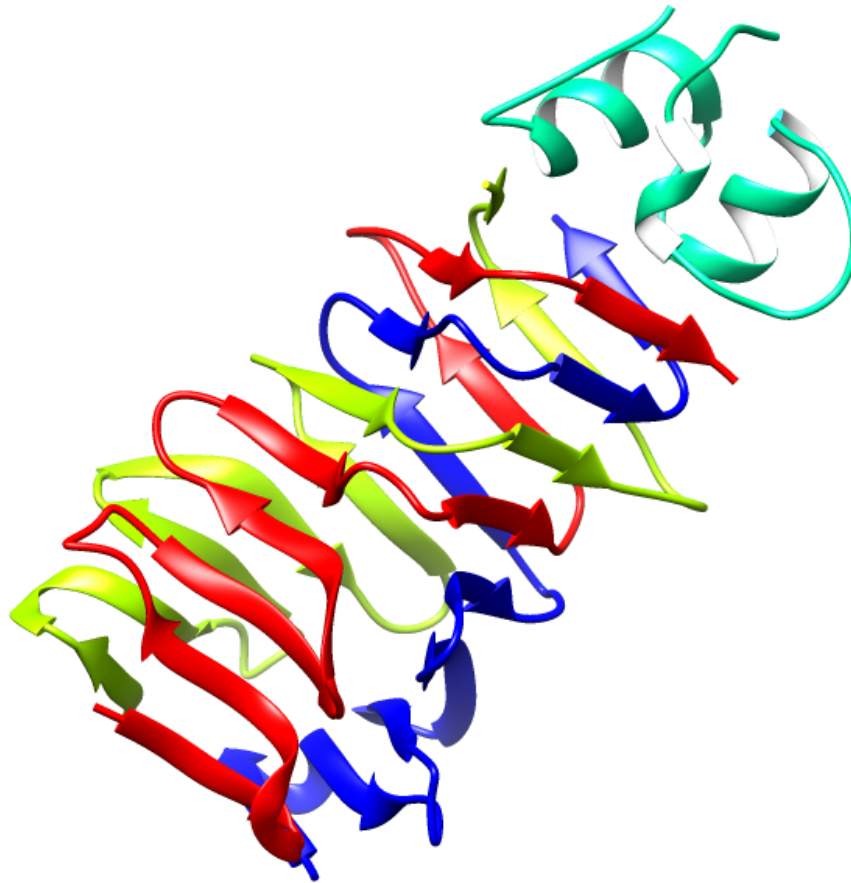


Figure 8: Part of the experimental structure of target H0953 (PDB 6F45), the adhesin tip complex of a bacteriophage tail fiber, illustrating subunit structure interdependence. One of the two protein chains contributing to this assembly forms a trimer (colored red, green and blue), with the N terminal five strand beta sheets of the three monomers packing against each other. The C terminal three beta strands of each monomer inter-digitate with each other. The C terminal stands also form an interface with the helical end of another subunit (green). Impressively, in spite of the apparent interdependency of the five-strand beta-sheets, accurate models were returned for that part of the structure. But failure to consider the even more intimate subunit interactions of the three N terminal strands resulted in incorrect models for that subdomain.