

The published manuscript is available at EurekaSelect via <http://www.eurekaselect.com/openurl/content.php?genre=article&doi=10.2174/1574893615999200504103643>

# Rosetta and the journey to predict proteins' structures, 20 years on

**Jad Abbass<sup>1\*</sup> and Jean-Christophe Nebel<sup>2</sup>**

<sup>1</sup>*Department of Computer Science, Lebanese International University, Bekaa, Lebanon*

<sup>2</sup>*Faculty of Science, Engineering and Computing, Kingston University, London, KT1 2EE, UK*

*\*Correspondence to: Jad Abbass, Department of Computer Science, Lebanese International University, Bekaa, Lebanon. T: +961 3 939370. F: +961 8 640871. E-mail: Jad.abbas@liu.edu.lb*

## Abstract

For two decades, Rosetta has consistently been at the forefront of protein structure prediction. While it has become a very large package comprising programs, scripts, and tools, for different types of macromolecular modelling such as ligand docking, protein-protein docking, protein design, and loop modelling, it started as the implementation of an algorithm for ab initio protein structure prediction. The term 'Rosetta' appeared for the first time twenty years ago in the literature to describe that algorithm and its contribution to the third edition of the community wide Critical Assessment of techniques for protein Structure Prediction (CASP3). Similar to the Rosetta stone that allowed deciphering the ancient Egyptian civilisation, David Baker and his co-workers have been contributing to deciphering 'the second half of the genetic code'. Although the focus of Baker's team has expanded to de novo protein design in the past few years, Rosetta's 'fame' is associated with its fragment-assembly protein structure prediction approach. Following a presentation of the main concepts underpinning its foundation, especially sequence-structure correlation and usage of fragments, we review the main stages of its developments and highlight the milestones it has achieved in terms of protein structure prediction, particularly in CASP.

## Keywords

Rosetta, Review, Protein structure prediction, fragment assembly, CASP

## 1. Introduction

Currently, Rosetta is a very large open source software suite encompassing several algorithms for different types of macromolecular modelling and protein structures analysis such as ligand docking [1], protein-protein docking [2], de novo protein design [3], and loop modelling [4]. Rosetta has moved from being a property of the University of Washington to 'RosettaCommons', i.e. a collaborative project that includes more than 150 developers in 23 different countries (rosettacommons.org). However, its first version was initially an implementation of a single algorithm for ab initio protein structure prediction written in FORTRAN [5]. In 2005, Rosetta 2 was launched using C++. However, the automatic translation process adopted then to move from FORTRAN to C++ made it unsuitable for further development. Consequently, to ensure its future, Rosetta had to be rewritten from scratch as a fully object-oriented C++ suite known as Rosetta 3 that was launched in 2009 [6] – its latest sub-version is 3.11. Different programs have been gradually added and improvements have taken place during the course of development, which has led to the current Rosetta software suite. Although the Baker lab has been focusing on de novo protein design in the past few years leading to prominent findings [7–13], predicting the structure of proteins has remained their utmost goal.

'Rosetta' was first mentioned in the literature in 1999 in a paper entitled "Ab initio Protein Structure Prediction of CASP III Targets Using ROSETTA" [14]. While they had used variable-length fragments during the pre-Rosetta phase (discussed later), they eventually decided to adopt fragments of fixed size; 9-mers represented the core of the building process, whereas 3-mers played a refinement role. Since then, those fragments' lengths have been continuously adopted in Rosetta. Results were truly encouraging as the group was ranked the best in the ab initio PSP category [15] in CASP3.

Herein, we review Rosetta as a fragment-based PSP describing its main components, such as the fragments picker, energy functions and fragments assembly, and related projects such as Robetta, Foldit and Rosetta@home. Rosetta's contributions of to the scientific community are thoroughly investigated by analysing its performance and breakthroughs in each CASP's round from the second to the thirteenth. However, before focusing on Rosetta, we present a short overview of fragment assembly PSP methodologies, including Rosetta, and the preliminary studies during the 1995-99 period that contributed to its birth. For the reader with an interest in the overall PSP and protein folding problems, excellent and recent reviews are available in the literature [16–25]

## 2. Overview of fragment-based PSP and Rosetta methodology

Motivated by the fact there is a strong correlation between sequence and structure at the local level, fragment-based protein structure prediction methods were first proposed in 1994 by Bowie and Eisenberg [26]. They rely on the concatenation of short rigid fragments excised from actual protein structures to construct putative protein models. Still, unlike homology and threading modelling, fragment-based predictors are able to handle template-free modelling (FM) targets; sometimes with very high accuracy, especially for small proteins [27]. Fragment-based protein structure prediction packages can be seen as offering a 'compromise' between ab initio and fold recognition modelling [28]. Methods such as FRAGFOLD [29], Rosetta [6], I-TASSER [30], and QUARK [31] have demonstrated the strength of such approach [32]. Regardless of the fragments' length used by those methods, their popularity is supported by five main points. First, since the smallest element considered in computation is a set of amino acids instead of a single one, the entropy of conformational search space is decreased in a dramatic way. Second, short sub-sequences converge towards a relatively limited number of sub-structures. Third, usage of Monte Carlo simulations instead of Molecular Dynamics allows making those methods much faster than pure physics-based ones. Fourth, the fragments that are used are already of low-energy, therefore, local interactions need not to be calculated within the fragments after each substitution; this feature makes such approach much less expensive than its competitors. Fifth, from a short fragment perspective, a structure can be built from fragments of other structures that belong to totally different architectures, folds and structures.

Early studies conducted by Rosetta's creators during the pre-Rosetta phase highlighted local sequence-structure relationships [33] suggesting that methods built on Bowie and Eisenberg's principles should only consider short fragments. As a result, the design of Rosetta relied on the assembly of short fragments (3-mers and 9-mers) excised from high resolution protein structures. Using the target's sequence, for each position, the best 9-mers and 3-mers are selected. This is performed, not only using sequence similarity and the sequence profile, but also by considering secondary structure prediction information generated from several sources as well as Ramachandran map probabilities. Then, the process of building conformations is conducted using two levels of search and refinement: coarse and fine grained associated with their respective energy functions. In the first level, low-resolution conformations are generated by representing the chain by heavy atoms of the backbone besides a single centroid for the side chains, whereas in the second one, all atoms are modelled. In addition to keeping the fragments rigid during the simulation as most methods do, Rosetta maintains bond angles and length at some ideal values to reduce the search space. Accordingly, the sole degrees of freedom in the coarse-grained search are the backbone torsion angles, whereas side chains are only taken into account in the fine-grained stage [27]. A noteworthy observation concerning the force fields type used in both scoring functions is the usage of both physics and knowledge-based terms [34].

During the coarse-grained search and refinement and in order to generate a conformation's backbone along with its side chain centroids, Rosetta operates in two main steps: first, 9-mer fragments are inserted within the initial fully extended conformation; second, insertions of 3-mer fragments are used to refine the structure previously generated. 9-mers and 3-mers are protein fragments extracted for each amino acid - except for the protein C terminus - of the protein of interest from a template database according to some similarity criteria. Eventually, Rosetta converts the coarse-grained conformation into an all-atom representation by adding all missing atoms using knowledge-based information extracted from known structures [27].

Figure 1 shows a high-level Rosetta's timeline that reveals the lifetime of the main phases, versions, scoring functions and additional tools, i.e. extensions (discussed later).

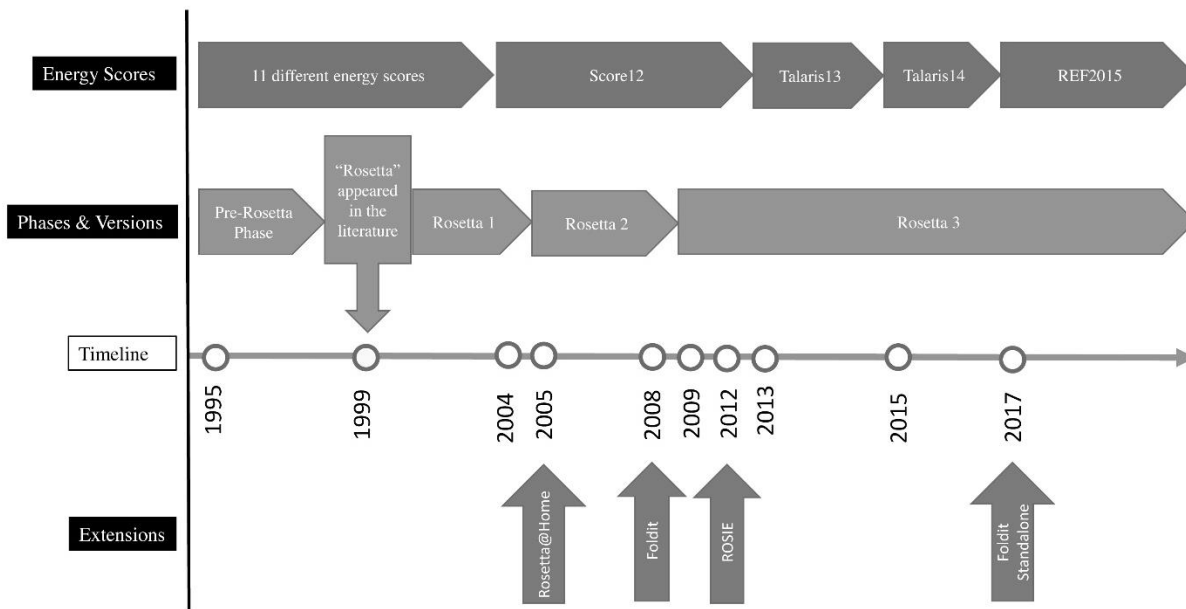


Figure 1: Rosetta's timeline since 1995.

### 3. Preliminary studies during the pre-Rosetta phase

Before the launch of an *ab initio* protein structure prediction based on fragment-assembly, David Baker and his co-workers had been investigating some ideas related the conformational conservation of the short sequences found in different proteins [35], the recurring short sequence motifs used to identify protein family borders [36] and the strong correlations between local sequences and structures [37]. Regardless of a specific secondary structure, their thorough study illustrated that the sequence-structure correlation shows dramatically increasing relative entropy as the length goes from 3 to 8 amino acids, then a slow increase till 10 amino acids (the peak of relative entropy value) followed by a slow decrease till 15 amino acids. More specifically, the authors investigated some inconsistencies researchers had faced to reach an accurate local sequence-structure mapping. Taking into account fragments of length range of 3 to 15 amino acids, another study was conducted to investigate all types of fragments including those that lie in a transition region between different secondary structures; the latter was considered new as previous papers had excluded such parts. Some 'ideal' sequence lengths were found as follows: 13 and 15 amino acids for helix caps (helix-turn-helix motifs), 7 to 11 for helices, 3 and 5 for  $\beta$ -strands, loops and turns. As for the transition fragments results were not as accurate as pure secondary structure regions, however sequences of length 7 and 9 residues showed some level of successful mapping for turn-to-sheet. Furthermore, another study reveals that local sequence motifs that are likely to recur within a protein family. They concluded that local interactions favour a limited number of substructures which in turn dramatically decreases the search space, i.e. reduces its entropy. The University of Washington's researchers suggested that most probably this is what happens *in vivo* giving a

reasonable explanation of the fast folding process. All those findings had paved the way for the Baker research group to adopt a simple hypothesis: a protein structure can be constructed from a set of short substructures.

In 1997, a paper was published by the Baker lab describing a relatively simple algorithm for predicting the tertiary structure of a protein using a fragment assembly approach aided by simulated annealing and a Bayesian scoring function [38]. Although some concepts have been changed in the subsequent releases, one can consider this paper as the first step towards Rosetta. Once fragments are determined based on sequence similarity a purely knowledge-based scoring function is employed for measuring non-local interactions to build the final conformation. An important particularity of this scoring function is exploitation of Bayes' statistical theorem using a large database where the sequences have known structures: it allows identifying the most likely structure that could be associated with a given sequence.

$$P(\text{Structure} | \text{Sequence}) = P(\text{Structure}) \times P(\text{Sequence} | \text{Structure}) / P(\text{Sequence})$$

Less than two years later, a follow-up paper was published suggesting an improvement to the scoring function [39]. It introduced some sequence-independent terms, such as the properties of packing  $\beta$ -strands to form  $\beta$ -sheets, besides the old sequence-dependent ones; yet using the same Bayes' statistical theorem. Eventually, based on their previous work previously described, David Baker and Christopher Bystroff contributed to CASP2 in the 'ab initio' category (8 targets) under the group 'BAKER'; the fragment library was called "I-Sites" and comprised fragments of 3 to 15 residues length [40]. Although, overall, their results were generally not satisfactory, they succeeded to reach a reasonable accuracy for one of the targets.

#### 4. Energy functions

Rosetta's energy function, which combines knowledge-based and physics-based terms, has passed through four main phases: *Score12* [34], *Talaris13* [41], *Talaris14* [42] and *REF2015* [43].

The *Score12* force field lasted for around 10 years as the default energy function of Rosetta3 [34]. It has been considered to be the 'gold standard' as during those years Rosetta achieved many milestones such as reaching native-like conformations for small proteins [27] as well as some of CASP's targets' for high accurate predictions [44–47]. The *Score12* energy function comprises two versions: coarse-grained for low resolution, where a residue is represented by the backbone's heavy atoms besides the centroid of the side chain, and fine-grained for high resolution, that is, for all-atom representation. Low resolution terms include (i) secondary structure pairing terms – a knowledge-based score to evaluate the favourable hydrogen bonding value between any couple of strands and helix-strand packing, (ii) radius of gyration, also known as packing density, which is used to favour compact folds using van der Waals attraction forces, (iii) van der Waals repulsion term, (iv) solvation term [48] that includes both a bonus and penalty value and (v) pair-interaction electrostatic forces for up to 12 Å distance of separation. High resolution terms include, in addition to the last three terms above: (i) a Hydrogen bond score [49] (ii) Ramachandran and torsion angles ( $\phi$  and  $\psi$ ) preferences, (iii) *Dunbrack* rotamer energy – a knowledge-based term to assess the likelihood of a certain rotamer to exist [50, 51] and (iv) the reference energy for each residue type in its unfolded state.

Relying mainly on a new *Dunbrack* rotamer library [52], a wider range of experimentally high resolution conformations in the PDB, and a thorough optimisation process to adjust weights, *Talaris13* was introduced instead of *Score12*. *Talaris14* was simply an error-corrected version of *Talaris13* as only a hydrogen bond's weight was changed and the remaining weights were adjusted accordingly. Although the paper giving details of *Talaris14* paper was published in 2014, its widespread usage commenced in 2016. *REF2015* has been Rosetta's official energy function since July 2017; it includes some updates such as optimised electrostatic parameters and additional terms using the Lennard-Jones potential for hydrogen atoms. It is worth mentioning that the four versions of the energy functions comprise weighted-based terms and none of the changes that took place could be considered as 'major'. Moreover, the unit of all energy functions is Rosetta energy unit (REU); a Rosetta-specific metric that cannot be converted into standard physical units such as kilocalories per mole (kcal/mol).

## 5. Fragment picking

The process of creating libraries of fragments (9-mers and 3-mers) is treated as an independent task that is performed prior to the execution of the main ab initio Rosetta PSP. In 2011, a new fragment picking tool called 'picker' was introduced to replace the previous one – 'nnmake' (written in FORTRAN) - and has been in use since then [53]. The main enhancement provided by 'picker' is that restraints such as distances between some atoms were introduced in the protocol used to build fragments for the ab initio prediction of proteins. In addition, it was written in an object-oriented modular approach.

The latest protein database file, where fragments are excised from, comprises 16,801 high resolution template structures of average size of 257 amino acids. The "picker" tool comprises three protocols: best fragments, quota and flexible loop design protocols. Whereas the last one is used for protein design, the first two are used for fragment picking. The Quota protocol is the one that is dedicated primarily for ab initio protein structure prediction. As its name implies, it applies the principle of "quota" for secondary structure prediction taken from at least two different resources (explained further below).

The scoring function, on which the selection of candidate fragments is based, is evaluated at each position in the sequence in question (except the last 8 and 2 positions in case of 9-mer and 3-mers respectively) typically to generate 25 and 200 9-mers and 3-mers respectively. More specifically, the fragments are chosen based on: secondary structure predictions from at least two resources such as PsiPred [54], Jufo [55], SAM [56] or Porter [57], their corresponding scores in the Ramachandran map, and the sequence profile by PSI-BLAST [58]. Note that the secondary structure prediction of the middle residue is the one that determines the 'overall' secondary structure of the fragment. For the sake of further explanation of the 'quota' protocol, we will assume that three secondary structure predictors are used: PsiPred, Jufo and SAM. It is worth noting that due to the overall higher accuracy achieved by PsiPred, predictions produced by PsiPred, Jufo and SAM could be associated to the following weights: 0.6, 0.2 and 0.2 respectively. In this regard, the Quota protocol works as follows. Since none of these three predictors is optimal, an approach using a unique total score based on which all 9-mers and 3-mers will be selected would be biased. For instance, if a fragment's middle residue is predicted with a confidence to be 50% a helix, 30% a strand and 20% loop, all fragments could be chosen as helix. Instead, the Quota protocol guarantees that the corresponding percentages of fragments will be generated from each pool. In such a case, amongst the 200 3-mers, 100 fragments will be helical and taken as follows: 60 from PsiPred, 20 from Jufo and 20 from SAM pools. The following three terms: profile, secondary structure prediction of the middle residue and the Ramachandran map probability value of the middle residue constitutes the scoring function based on which the fragments from the corresponding pool are picked. The default weights of each term are as follows: 1, 1 and 2 respectively.

## 6. Fragment assembly

Starting from a fully extended chain, the fragment assembly process takes place via a Monte Carlo search; a sequence window of length 9 is randomly selected and one of the available 25 candidate fragments in its turn is randomly selected. Once the torsion angles of that window are replaced by those of the chosen fragment's, the coarse-grained energy score is calculated; the minimisation process is performed using Simulated Annealing [59]. Therefore, if the energy score after an insertion is smaller than that of the previous conformation, it will be accepted, otherwise, the Metropolis criterion [60], for the sake of avoiding getting trapped local minima, may also accept it with lower probability for larger energy increases. In short, probability of accepting 'bad' moves, whose formula is shown below, decreases exponentially with the  $\Delta E$ , which describes how worse the energy increases.

$$P = e^{-\Delta E/kT}$$

Whilst  $k$  is constant known as the Boltzmann constant,  $T$  represents the temperature, a parameter that plays a key role in the Metropolis criterion. In natural annealing, temperature is first set to a high value, then, gradually, decreases until the material reaches the shape needed. In simulated annealing, temperature changes in the same context; it is first set to a certain high value, then decreased, consequently decreases the probability of accepting a

'bad' move (in Rosetta, it is fragment insertion that results in increasing the energy of the conformation). In both natural and simulated annealing, the heating and gradual cooling cycle can be repeated several times. Indeed, once a fragment replacement is accepted, the temperature is reset to its initial value, i.e. the acceptance probability is back to its default value.

The 9-mer insertion phase involves 28,000 insertion attempts, however terms of the coarse-grained energy score are added gradually. For instance, in the first 2,000 attempts, only steric overlaps, i.e. van der Waals terms, are considered, whereas in the last 4,000 insertion attempts, the complete energy function is estimated. Once the 9-mer insertion phase is completed, 8,000 insertion attempts using fragments of size 3 are performed, taking into account the whole coarse-grained energy function. After the overall 36,000 insertion attempts, the simulation produces a conformation with heavy backbone atoms only. Optionally, all additional atoms are then added using ideal values, and fine-tuned using an all-atom energy score, also known as a fine-grained energy function. The technical term of this phase in Rosetta is called 'relax'.

## 7. Rosetta's extensions

From a hierarchical perspective, Robetta protein structure prediction server [61] is considered the 'mother' of Rosetta whenever it participates in CASP. In the 'abstracts' book (found at <http://www.predictioncenter.org/casp12/index.cgi>) where participants describe their method, Rosetta's team explain that targets' sequences are firstly passed through Robetta. Robetta is a fully automated server that, when fed with a sequence of interest, follows two different routes: first, a comparative modelling one and second, a de novo approach using Rosetta. While domains and/or regions with high sequence similarity score are forwarded to the Rosetta-based tool for homology modelling called RosettaCM [62], regions with low homologs detection (mainly long loops) are modelled using the standard de novo Rosetta. Once all those regions have been built independently, they are combined by an iterative domain assembly method [63] where fragments are inserted into the areas connecting those regions using de novo Rosetta's scoring function. Side chains are then added following the same standard as in de novo modelling. Accordingly, whenever an FM target is submitted, Robetta will simply act as de novo Rosetta. Robetta's initial key step is determining domains' start and ending and candidate templates for easy regions. Such a procedure is achieved using highly ranked sequence alignment tools for homology and threading modelling: HHSearch [64], Sparks [65], RaptorX [66], BLAST [67], PSI-BLAST [58], FFAS03 [68, 69] or 3D-Jury [70]. Furthermore, if GREMLIN [71], a state-of-the-art contact prediction method that employs meta-genome sequences [72] provides accurate results, those results are fed to Rosetta as restraints for both sampling and refinement phases. For the sake of quality assessment phase – where top model(s) should be chosen amongst a large set of decoys (up to 300,000) – ProQ2 is carried out [73]. The steps described in this section correspond to those of the pipeline followed by "BAKER-ROSETTASERVER" group that is mentioned in the next section. In 2003, Robetta went online – Robetta Server - allowing users to submit their sequences for prediction using either comparative modelling or ab initio. Moreover, users could build Rosetta's fragments online for local execution of Rosetta [61].

In 2005, Rosetta@Home, a distributed grid computing system that uses volunteers' idle processors all around the world to execute processes related to different Rosetta applications and studies including CASP predictions, was introduced [46, 74, 75]. Besides Rosetta@Home, another initiative was taken to involve the public mainly in protein structure prediction: a game called 'Foldit' was launched in 2008 [76]. Outcomes have been beyond the developers' expectations as valuable findings were produced and published using this gaming environment [77–80]. A highly improved version called "Foldit StandAlone" was made available three years ago [81]. In 2012, a much bigger server than Robetta called ROSIE (the Rosetta Online Server that Includes Everyone) was launched [49]. Currently it includes 18 Rosetta applications that can be executed without usage of local computers. It is worth noting that Rosetta@Home is used to run a large amount of processes being submitted every day to ROSIE.

## 8. CASP Contributions and Results

The state of the field of PSP has been monitored and quantitatively evaluated since 1994 by the biennial CASP event. This community-wide experiment has grown significantly from a set of 33 targets which attracted 27 contributing groups who submitted 186 models (CASP1, 1994) to a set of 194 targets which led to the submission of around 57,000 models from 185 participating groups (CASP13, 2018). Since the second CASP round, Rosetta has continuously contributed to that competition achieving notable results comparing to other groups. Herein, we first introduce the main concepts of CASP such as the different types of targets, evaluation metrics and domain definition, then we review the contribution of Rosetta for each round.

### 8.1 CASP Overview

Typically, a set of protein sequences are released gradually across a couple of months during which research groups from around the world attempt to predict their 3D structures by submitting putative models (up to 5 per target). Once a target's submission deadline has passed, determination of its native structure is conducted *in vitro*. If successful, a thorough evaluation is performed on the submitted models. In the first 6 rounds – that is, till CASP6 – targets were classified into three categories: ‘Comparative modelling’, ‘fold recognition’ and ‘ab initio’; the latter was renamed to ‘native fold’ then ‘new fold’ in CASP4 and CASP5 respectively. Afterwards, targets released by CASP have usually been classified into two main categories: template-based modelling (TBM) and Free modelling (FM). In addition, since CASP8, domains have been widely adopted, leading targets to split, mainly large ones, for fairer assessment. However, interestingly, both the domain determination and classification (FM/TBM) operations are carried out after the completion of submission phase. This procedure relies mainly on two sources: first, the availability of templates for specific units and second, comparison between the scores of servers of the whole targets versus those of the split domains. There are several reasons that led to taking into consideration domains rather than the whole conformation for evaluation purposes. First, some targets are relatively long and therefore contain several domains; evaluating the whole prediction may seem unfair especially that a global structural superimposition is likely to yield bad scores and therefore may not reveal the prediction accuracy of ‘independent’ regions, namely domains. Second, some targets contain domains of different nature: whereas some of them may be classified as FM, the other(s) might be classified as TBM. One should also note that even in single-domain targets, evaluation of only a subset of the structure might take place. Indeed, as structures of tails are often classified as coils, i.e. they have no defined secondary structure according to the dictionary of secondary structures of proteins (DSSP) [82], their prediction is not relevant as their inclusion would most likely only lead to a lowering of model quality [83–87]. Since CASP4, the only information about targets during the pre-submission phase is if they have been released as either ‘server only’ or ‘all groups’. While models for the ‘server only’ targets must be produced automatically, those for the ‘all groups’ targets may include manual interventions, such as bending, that are carried out by field experts. Typically, those tunings are performed on models produced by the automated server used by each group. Consequently, the period for submitting models following the release of a target's sequence is much longer for ‘all groups’ than for ‘server only’ targets.

As mentioned earlier, each group can submit up to five models per target, one of them should be designated as *first model*; taking into account the best of the five models is referred to as *best model*. Up to CASP4, in addition to visual inspection, accuracy of submitted models was assessed according to the root mean square deviation (RMSD) of the longest reasonable well-predicted continuous subset of the target. Since CASP5, Global Distance Test Total Score (GDT\_TS) [88, 89] started to emerge as a better metric as it is less sensible to outliers. It has eventually become the main evaluation criterion. GDT\_TS is the average of residues whose superimpositions with the native structure is below 1, 2, 4 and 8Å. The range is [0-100] where 100 is achieved by two completely identical structures. We will refer to it as “GDT” in the text. In order to rank groups, the community organisers have adopted the cumulative z-score of the GDT of the *first model* after removing the outliers where Z-score is below a certain threshold (typically -2.0). Z-score range is typically from -3 to +3. The rationale behind using the z-score is to avoid penalising groups for very bad predictions.



## 8.2 Rosetta in CASP

In the next sections, we summarise the achievements of Rosetta in each round of CASP since the second one. We highlight the targets for which the Baker group achieved significantly better results than its competitors. To this end, we have chosen for each round one target where Rosetta made a relatively breakthrough, showing their best model along with the corresponding native structure in Figures 1, 2 and 3. Coordinates of the submitted models were retrieved from the archive of CASP ([http://predictioncenter.org/download\\_area/](http://predictioncenter.org/download_area/)) and those of the native conformations from the PDB; visualisations were carried out using PyMol [90]. Although some models shown in Figures 2, 3, 4 and 6 may seem quite different from the native structures, they actually displayed remarkable accuracy relatively to either the difficulty of the target or to the models produced by the competing groups. Since the main aim of this review is to assess Rosetta as an ab initio PSP, the focus will be on FM and ‘server only’ targets. In the same context, the *best model* will be considered besides the *first model* since protein Quality Assessment – the set of techniques used to select the most accurate decoy out of a pool – may be considered as a research area on its own; it was even included as a separate category in some rounds since CASP7 [91, 92].

### 8.2.1 CASP2: first attempt

In CASP2, the Baker group was called “I-Sites”, in reference to the name of the fragments library; the library contained 82 sequence patterns with their corresponding structures, distributed amongst 13 motifs whose length ranges from 3 to 15 [40, 93]. Baker and Bystroff contributed to 9 targets; in terms of performance only a small part of a very long target (Target 22) is worth mentioning: out of the whole 592 amino acid structure, the sub-structure ranging between positions 200 and 300 – in particular between 216 and 276 – showed remarkable accuracy. Taking into account a window of 40 amino acids, ‘good’ superimpositions were recorded where the lowest RMSD achieved was 4.9 Å, see Figure 2. It is important to note that no successful prediction was recorded by any participating group in the tertiary structure category of CASP2.

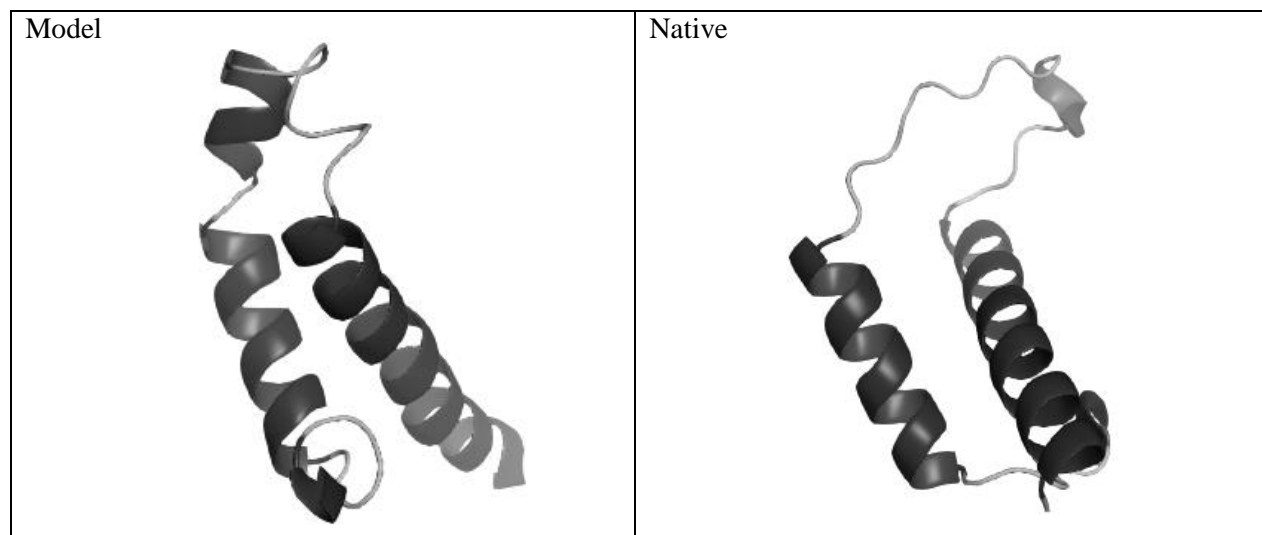


Figure 2: Short fragment (216 – 276) Structure of the model submitted by ‘Baker and Bystroff’ group (I-Sites) in CASP2 for Target T0022 (592 amino acids in total) (left) and the corresponding native structure fragment – PDB ID: 1FUI (right).

### 8.2.2 From CASP3 to CASP6: world domination

In CASP3, that took place in 1999, ‘Rosetta’ made its formal entrance topping the 61 competing groups in the Ab initio category for 15 different targets [14, 15]. 4 targets, respectively 4 and 7, were classified as ‘easy’, ‘medium’ and ‘hard’ based on their structural class (mainly-beta sheets proteins are considered to be particularly hard for ab initio predictions), architecture or simply according to the accuracy of their predictions. In one ‘medium’ target

(T0056), the Baker group was able to predict 69 residues with an RMSD of 4.1 Å, see Figure 3 (a). Moreover, only Rosetta was able to predict the correct fold of a hard target (Target 83 – C-terminal domain). Taking into consideration fragments superposition and correct architecture, Rosetta ranked the best group overall; it was able to hit 7 medium/hard targets with at least one fragment (length  $\geq 30$ ) with an RMSD less than or equal to 3.0 Å. However, out of the 12 ab initio targets for which Rosetta submitted predictions, only the overall models of four targets could be classified as ‘good’.

By the time Rosetta contributed in CASP4, 3-mers and 9-mers became the standard and fixed-length sizes of the fragments it uses. As both methodologies that employ pure physics laws and some knowledge-based ones demonstrated the ability to some extent to predict new folds, usage of the term ‘ab initio’ did not seem appropriate anymore to the CASP community. Thus, it was replaced by ‘Novel Fold’ (NF) [94]. Rosetta demonstrated significant performance in relatively long fragments in 17 out of the 24 CASP4 NF targets. In 14 out of those 17 targets - in particular T091, T106 and T116, Rosetta submitted the best models amongst the 55 competitors [95]. Baker’s team was able to produce fragments whose length ranges from 53 to 139 amino acids with an average RMSD of 4.8 Å. Similar to the previous version, Rosetta was the best group for the ‘Novel Fold’ category. Figure 3 (b) shows the prediction of T106 versus the native structure in which a fragment of length 106 was modelled with an RMSD of 6.0 Å.

For the third consecutive times, the Baker group was assessed as the best one in the NF category that comprised 13 targets in CASP5, according to the z score-based evaluation of the GDT [89]. Amongst those targets, 5 were classified as ‘pure new fold’ whilst the remaining 8 were considered on the ‘edge’ between ‘new fold’ and ‘fold recognition’. Four out of those ‘hard’ 5 targets were alpha-beta proteins. Target T0129 contained seven alpha-helices; as a whole, Rosetta was the only group that was able to predict it correctly, in particular its model 4. In addition, although other groups were able to produce accurate models of the ‘easy’ target T0170, small 4-helix protein, the Baker group still achieved the best results. Perhaps, the most interesting achievement made by Rosetta was with target T0187 – a pure beta sheets domain of 36 amino acids. Although it was categorised on the edge of ‘comparative modelling’ and ‘fold recognition’, no other group in any category was able to predict the correct fold at any ‘acceptable’ level of accuracy. See Figure 3 (c) that shows the first model and the corresponding native structure. It is worth noting that during CASP5: the separation between ‘server’ and ‘human/server’ groups started to emerge, and ‘Robetta’ started to contribute, however, as an ‘auxiliary’ plan for TBM targets [96].

In CASP6, among the 17 targets that belonged to the NF category, Rosetta was able to submit the best model for 7 of them. Overall, once more, it was ranked the best group in that category [44, 97]. Its more notable achievements were observed in the targets: T0198, T0212 and T0272. In T0198 – a six-helix bundle - the RMSD over 210 residues reached 4.0 Å. The average z-score was higher by 80% compared to the group that ranked second. However, taking all targets into accounts, the average of the GDT of the Baker group of the best and first models was 34.4 and 31.8 respectively, and for the 6 targets that were categorised as ‘new fold hard’, results were a bit disappointing. By then, reviewers had become confident that fragment assembly methods like Rosetta and FragFold were the most successful approaches. We show in Figure 4 – (a), the submitted model of the first domain of T0272 (85 residues) along with the native structure.

Figure 5 presents a pictorial summary of Rosetta’s achievements from CASP2 to CASP6, where we highlight for each round the ‘most remarkable’ superimposition in terms of the longest fragments within ‘acceptable’ RMSD value ( $<5\text{Å}$ ). The figure clearly shows that Rosetta’s performance has improved continuously during the period. While the fragment selected for CASP4 has a RMSD slightly higher than the one used for CASP3, its length is significantly larger, which is a better achievement. Comparison between performance at CASP3, CASP5 and CASP6 is much simpler as, by chance, all fragments are of length 69 for those three rounds.

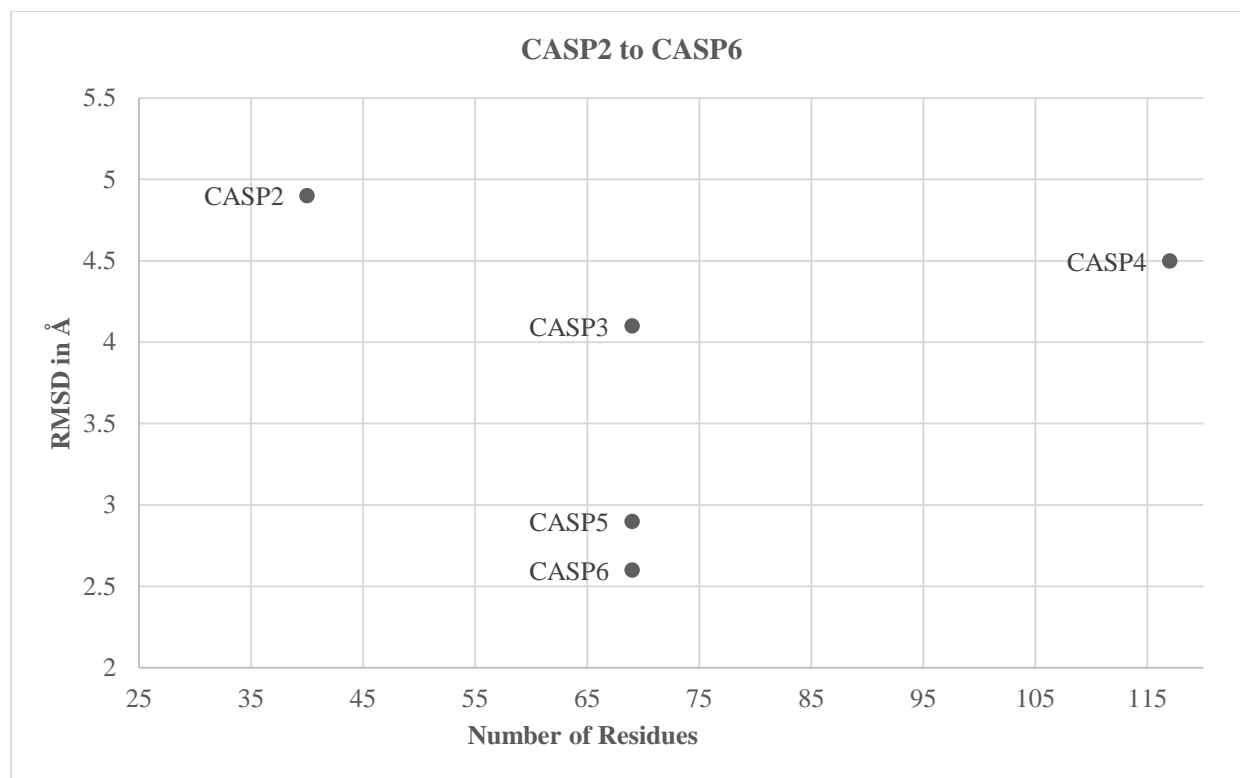


Figure 5: Longest fragments that Rosetta was able to predict with RMSD less than or equal to 5Å during the CASP2 to CASP6 period

### 8.2.3 CASP7 and CASP8: competition with I-Tasser

With CASP7, three notable changes took place: first, the terms FM and TBM were adopted; second, I-TASSER, although requiring fewer computational resources, became a serious competitor for Rosetta, even if the Baker group still reached the first place for FM targets; and third, Rosetta started to use Rosetta@home to generate its CASP predictions [46, 98]. In this round, 15 FM and 4 FM/TBM targets were released. Target T0319 was considered a hard one due to the low GDT scores recorded by all groups. However, Rosetta succeeded to show the ‘essential features’ of both subdomains despite their incorrect orientation. Although Rosetta’s model of the second domain of T0321 was ranked 24<sup>th</sup> in terms of GDT, it was judged as the best model by visual inspection. Figure 4 (b) shows the Baker’s 4<sup>th</sup> model of the second domain of target T0347; an overall GDT of 52.8 – evaluated as ‘moderate’ quality - was achieved thanks to the usage of extensive all-atom refinement performed using the grid computing system Rosetta@home.

In the 8<sup>th</sup> round of CASP, 102 groups competed in the FM category that comprised 13 targets; out of the 13, 10 were classified as FM and 3 as FM/TBM [99]. Three of the 10 FM were classified under ‘server only predictions’ and the remaining under ‘human/server predictions’. For 4 targets, no group was able to submit any model with a ‘satisfactory’ level of accuracy. Overall and taking into consideration the number of *best models*, the human/server and server groups of Rosetta – ‘DBAKER’ and ‘BAKER-ROBETTA’ – ranked first and third respectively. The second place was taken by MUFOLD-MD server group [100]. Figure 4 (c) shows a high-accurate model submitted by the server group of Rosetta for the second domain of target T0513. In addition, it is worth noting that the ‘Robetta’ server achieved remarkable results for the 64 TBM domains of CASP8 [47].

#### 8.2.4 CASP9 and CASP10: a duel between “Baker lab” and “Zhang lab”

In the ninth version of CASP where around 17,000 models were submitted, visual inspection became secondary and was only used for selected targets, when groups achieved similar performances. In CASP9, 30 domains were classified as FM, 4 of them were reserved for ‘server only’ [101]. ‘ROSETTASERVER’ ranked third whilst the first and second places were achieved by ‘Zhang-server’ (that is, I-TASSER) and ‘Quark’; a similar outcome was obtained in the human/server groups where the ‘Zhang’ and ‘Baker’ groups ranked first and second respectively. Due to the low number of FM targets, in particular in the ‘server only’ category, one could conclude that differences between their statistical-based results might lack ‘significance’. Interestingly, the assessment committee highlighted the ‘ROSETTASERVER’ predictions of target T0581 (Model 4) as the ‘winner’ among the CASP9 FM targets. The architecture of that target is a sandwich consisting of 2 helices and 4 beta sheets (see Figure 6 (a)). Rosetta was not only able to predict correctly the core components, but also to model an unusual helix that consists of two perpendicular helix extensions. Another remarkable achievement was the correct modelling of the four beta sheets of target T0581, whereas 89% of the models submitted by the other servers modelled those beta sheets as an entirely helical part. Even PSI PRED - one of the best secondary structure predictors then - modelled that part to be 3 helices and one beta strand. The way Rosetta dealt with that problem was by using two additional resources for secondary structure predictions (this feature was presented earlier). However, credits of the success of T0581 that captivated the assessment committee of CASP9 should also be shared with ‘Foldit’: one of the accurate models produced by Rosetta@home was provided to Foldit game players as a starting conformation where it was further improved before its submission to CASP [79].

CASP10 was quite similar to CASP9 in terms of the low number of FM targets and group ranking. There were 20 FM targets (one of them was FM/TBM) that attracted more than 9k models from 147 prediction groups (68 were ‘server only’) [102]. ‘BAKER-ROSETTASERVER’ ranked third whilst ‘Quark’ and ‘Zhang-server’ ranked first and second respectively. Remarkable achievements of Rosetta include targets: T0651-D2 - See Figure 6 (b) - and T0695 hit by models 3 and 4 respectively. In terms of TBM targets, Rosetta team had developed an amended version of Rosetta that can hit TBM targets using comparative modelling called RosettaCM [62]. Their results in CASP10 were noteworthy in terms of the accuracy of the backbone and side chain atoms, due mainly to the all-atom refinement stage.

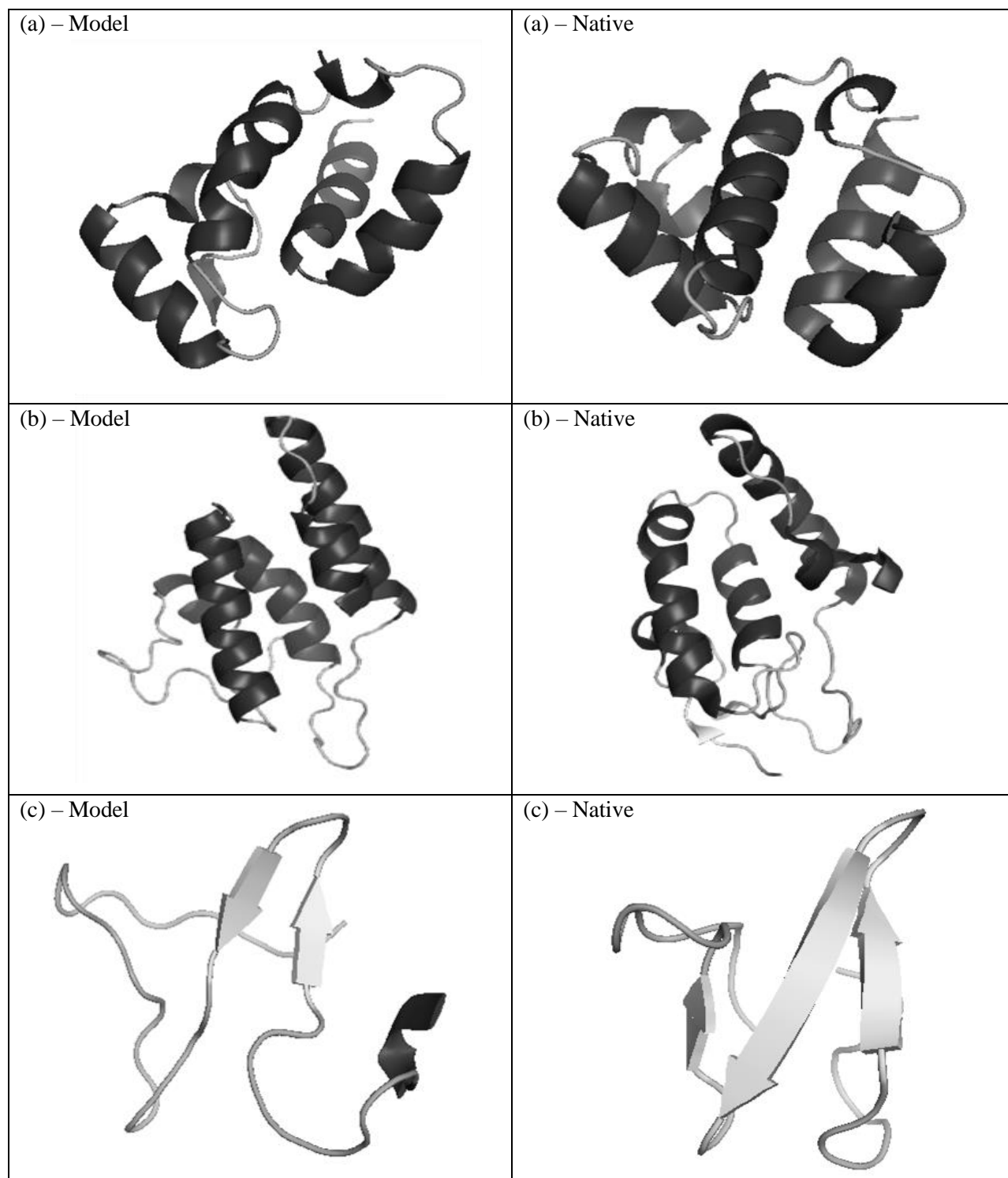


Figure 3 (a): Structure of part (23-136) of the fifth model submitted by the ‘Baker’ group in CASP3 for Target 56 (left) and the corresponding part of the native structure – PDB ID: 1EH2 (right). Taking into account a window of 69, an RMSD of 4.1 Å was recorded for the best fragment. (b): Structure of the first model submitted by the ‘Baker’ group for Target 106 (left) in CASP4 and the native structure – PDB ID: 1IJX (right). (c): Structure of the first model submitted by the ‘Baker’ group in CASP5 for Target 186 (left) and the native structure – PDB ID: 1O12 (right). Except the length of the second beta strand, the whole topology was correctly predicted.

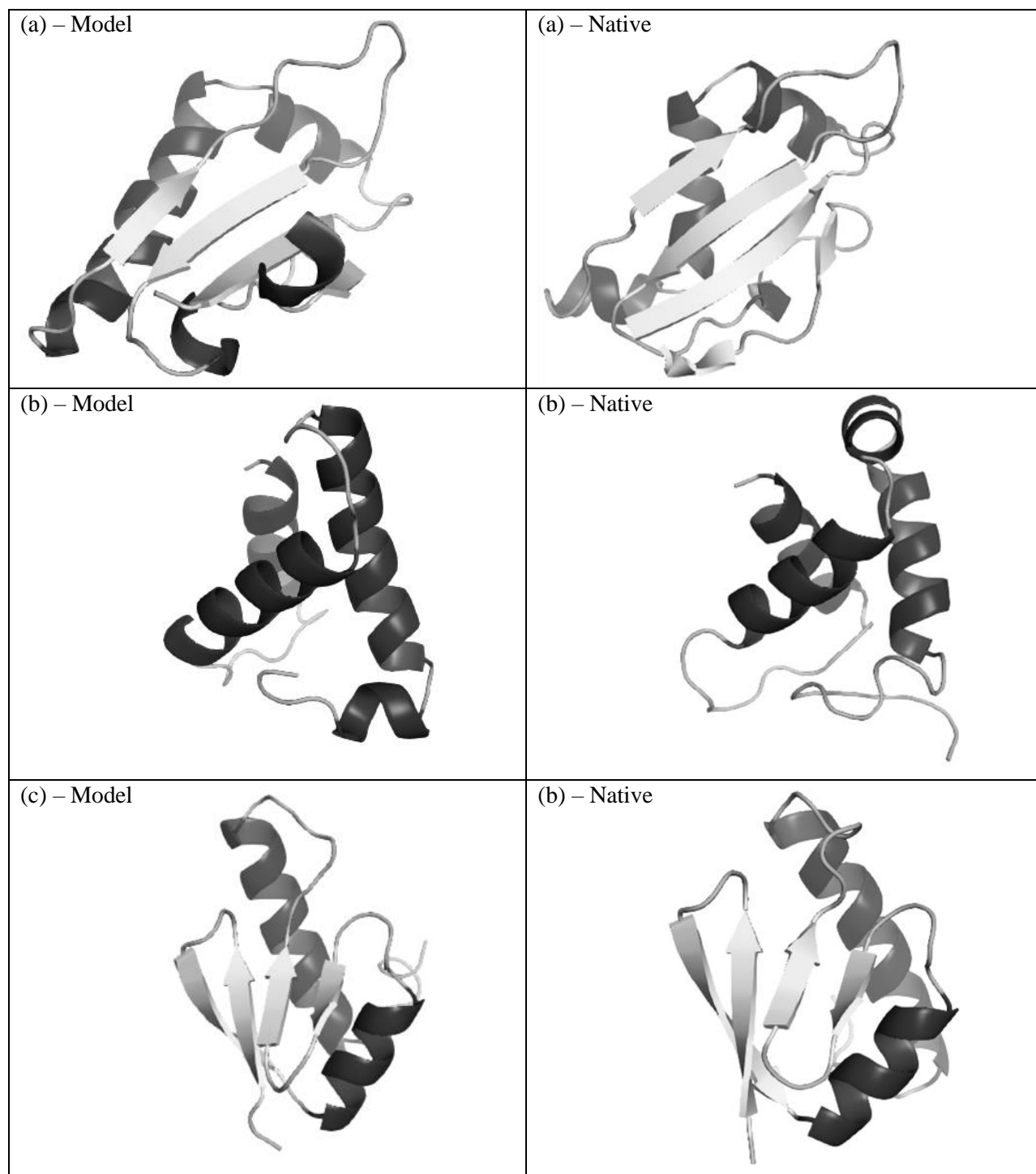


Figure 4 (a): Structure of the first model in CASP6 for Target T0272 (left) and the native structure – PDB ID: 1WJ9 (right). A GDT score of 58.53 was recorded. (b): Structure of the fourth model in CASP7 for Target T0347 (left) and the native structure – PDB ID: 2HWJ (right). Except for the N-terminal end, the overall shape was predicted with high accuracy (c): Structure of the first model submitted in CASP8 for Target T0513 –D2 (left) and the native structure – PDB ID: 3DUP (right) of length 69. GDT is 70.65

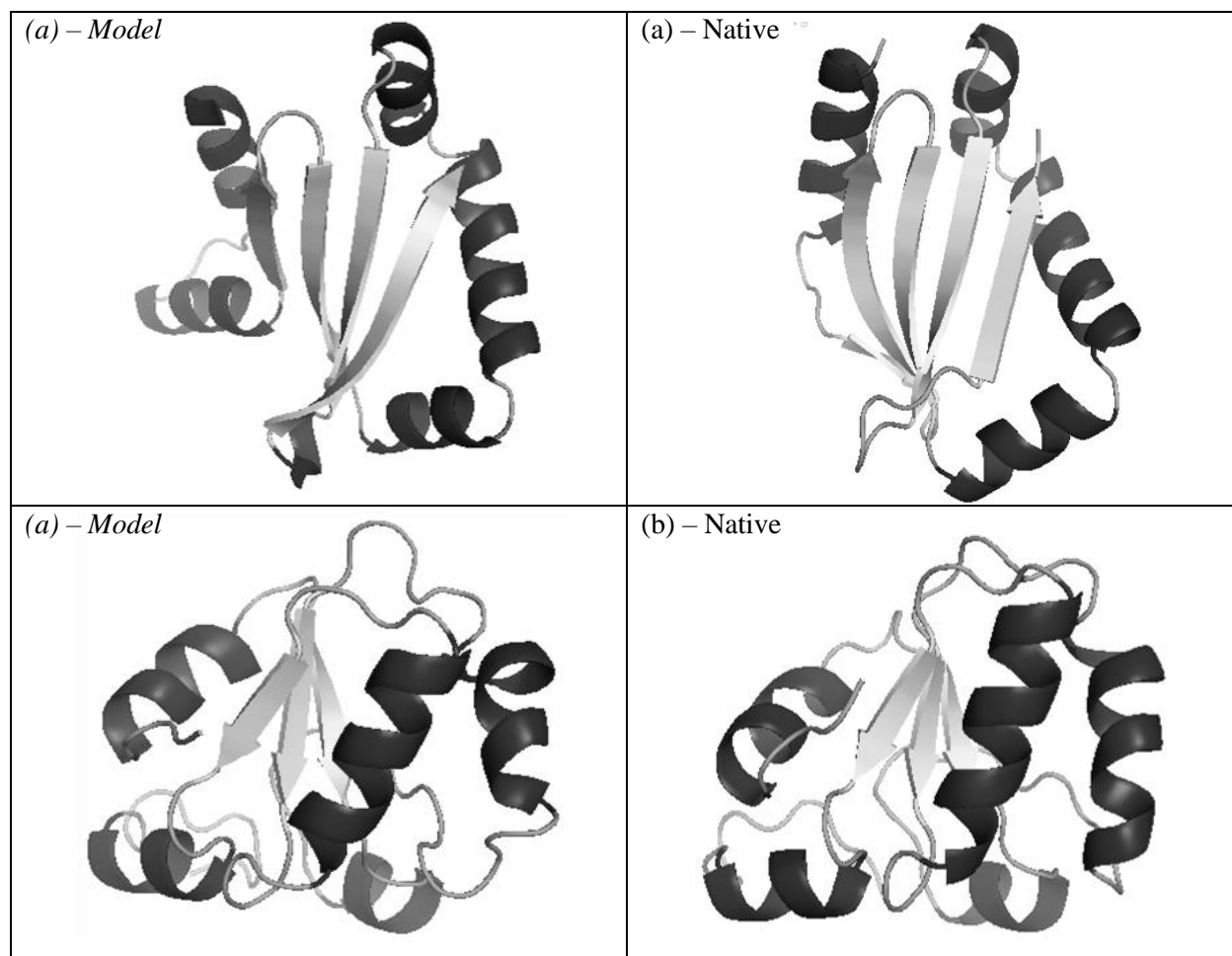


Figure 6 (a): Structure of the fourth model submitted by the Rosetta server group in CASP9 for Target T0581 (left) and the native structure – PDB ID: 3NPD (right). A GDT score of 67.86 was recorded. (b): Structure of the third model in CASP10 for the second domain of Target T0651 (left) and the native structure – PDB ID: 4F67 (right). GDT is 74.10.

### 8.2.5 CASP11, CASP ROLL and CASP12: back to the top

Due to the relatively low number of FM targets in CASP9 and CASP10, questions arose regarding the significance and fairness of their assessment and consequently the evaluation of the predictor’s methodologies and participating groups. Accordingly, the CASP community decided to address that shortage in CASP11 and CASP ROLL. The latter was a special continuing version for a one-year period that was dedicated to FM targets [84]. CASP11 and ROLL comprised 45 and 38 FM targets respectively; the largest number of ‘hard’ targets that had ever been released [103]. Although it was expected that CASP ROLL would release a much larger number of targets, the unexpected low number of participating groups led the organising committee to move some targets to the traditional version, i.e. CASP11. Nevertheless, as the total number of targets was unprecedented in the history of CASP, CASP11 and CASP ROLL were considered a ‘crucial occasion’ to clarify the status of ab initio PSP and reveal the strengths and weaknesses of the competing research groups. Whilst, I-TASSER, QUARK and RBO\_Aleph ranked the second, third and fourth best server groups by offering some breakthroughs, ‘BAKER-ROSETTASERVER’ stood out as the winner in both CASP11 and CASP ROLL [104–106]; in CASP ROLL, only “Zhang\_ab\_initio” seems to remain a close competitor. Notable accuracy was achieved for large parts of T0790-D2, T0761-D2 and R0021, see Figure 7 (a). On the human/server side, the following groups: ‘BAKER’, ‘Kiharalab’ and ‘LEER’ ranked the best three indistinguishable predictors [106, 107]. A remarkable achievement of the human/server Rosetta group in CASP11

was the target T0806-D1: Rosetta was able to predict a 256-residue domain with absence of any templates. Such a success was considered an outstanding accomplishment as it was the longest FM domain ever predicted. Constraints during sampling and refinement like residue-residue contact information – also known as contact maps – greatly contributed in the success of many targets [71, 106, 108]. It is important to mention that due to the availability of Rosetta’s source code, some successful groups – outside the Baker group – contributed in CASP11 with either an amended version of Rosetta or using a Rosetta protocol, such as energy functions or refinement, within their prediction pipeline. Those groups are: “COHVaidehi” [109], “FALCON\_TOPO” [110], “FLOUDAS\_SERVER” [111], “Handl” [112], “MeilerLab” [113], “MULTICOM-CLUSTER” [114], “NEFILIM” [115], “Pareto” [116], “RosEda” [117], “HomREM”, and “Rosetta\_at\_Kingston” [118, 119]. For instance, the server group “Rosetta\_at\_Kingston” that uses structural class prediction for fragment picking was able to perform even better than both official Rosetta groups in 6 out of 14 domains. It showed major improvement in the FM domain T0804-D2, where it achieved a GDT TS score of 44.6, whereas Rosetta’s first model only reached 30.4. The same group predicted a TBM domain (T0773 – GDT: 93.3) by a first model more accurately than, not only Rosetta groups’, but also all “Zhang-lab” related groups’.

CASP12 comprised 96 domains/targets classified as follows: 39 FM, 38 TBM and 19 FM/TBM. Overall, results for FM targets were substantially improved over the previous round. This is mainly due to the progress that had been made in contact maps predictions since many methodologies employed them as restraints while sampling the search space. Prominent results were detected with targets whose length exceed 100; ‘very accurate’ models were submitted for 16 out of 32 of those medium-sized targets [120]. Whilst in the overall results of all targets, both I-TASSER and QUARK performed better than Rosetta due to some failures in the FM/TBM domain category, Rosetta captured the first place in the hardest category, FM targets and interestingly, in the TBM category as well. In CASP12, Rosetta used contact predictions in two new ways: by guiding first, the selection of local fragments; second, the human intervention for the human/server targets [121]. In the latter case, improvements - sometimes dramatic - were shown for the majority of targets, even if a few models saw their accuracy decreased compared to those initially generated by the server. In the case of target T0942-D1, this strategy proved particularly beneficial as complicated usage of co-evolution data raised GDT from 35 to 77. Figure 7 (b) shows the first model of ‘BAKER-ROSETTASERVER’ for the server-only domain T0860-D1 of 136-residue length and the corresponding native conformation. Although the target was TBM, superposition with the native conformation resulted in GDT of 81.80; the highest amongst all methodologies including comparative modelling and fold recognition (for such targets, credits should be awarded to Robetta rather than ‘pure Rosetta’).

Since the competition amongst the server groups ‘Zhang-Server’, ‘Quark’ and ‘BAKER-ROSETTASERVER’ has become quite tight in the past 3 rounds, including CASP12, we have conducted a detailed analysis as an attempt to reveal strengths and weaknesses of each of those groups. Further data have been collected from CASP12 archive regarding those top three contributors by extracting the GDT of the submitted models in all categories. The average of the GDT of each group in each category is calculated and summarised in Table 1. It is worth noting that we have collected all data without taking into consideration Z-score values; therefore, Rosetta appears in the third place for FM whereas it leads the official Z-score based ranking.

**Table 1: CASP12’s Top Three Servers’ Detailed GDT Scores**

		<b>All - 96</b>	<b>FM – 39</b>	<b>TBM - 38</b>	<b>FM/TBM - 19</b>
<b>Average</b> (Std. dev.)	Zhang-Server	<b>52.0</b> (22.3)	<b>30.9</b> (12.6)	<b>72.4</b> (12.5)	<b>54.4</b> (11.3)
	QUARK	<b>51.3</b> (22.5)	<b>30.7</b> (12.1)	<b>71.8</b> (12.9)	<b>52.6</b> (14.5)
	BAKER-ROSETTASERVER	<b>50.4</b> (23.9)	<b>30.5</b> (14.8)	<b>72.3</b> (14.2)	<b>47.3</b> (16.1)
Min - Max	Zhang-Server	9.2 – 96.6	9.2 -54.7	50.9 – 96.6	39.2 – 80.1
	QUARK	9.4 – 96.3	9.4 - 57.0	49.5 – 96.3	23.3 - 79.3
	BAKER-ROSETTASERVER	12.1 – 97.6	12.1 – 77.1	34.1 – 97.6	20.3 - 76.5



Table 1 shows that the best group reached 52/100 as an average of GDT score of 96 targets which is considered as of 'moderate' quality, i.e. within the [40 – 59] scale. However, the status is much worse in FM domains, where on average, the best three pipelines deliver models in the 'poor' category [0 - 39]. Regarding the TBM category, results are quite good; surprisingly, 'pure' comparative and threading modelling contributing groups were not able to beat those three pipelines (Note that in such category credits should be awarded to Robetta rather than Rosetta). One would explain this by the following reason: reliance on several homologs to select fragments from leads probably to better conformations rather than usage of one or a very limited number of homologs.

Regarding the quality of the near-native structure, a study suggests that short fragments are likely to produce higher quality; on the other hand, multiple fragment lengths are able to generate overall better decoys as long fragments have a positive effect in the early stages of simulations [122].

Table 1 reveals that Rosetta presents the largest standard deviation amongst the three competitors. This may be explained by the facts that (i) whenever Rosetta succeeds to reach a 'good region' in the conformational space, its ability to explore that region is higher than the others due to its usage of much shorter fragments (maximum value of GTD in FM category is by far better than the two remaining competitors), and (ii) whenever simulations tend to reach a 'bad region', final structures' accuracy is relatively low due to the absence of long fragments.

### 8.2.6 CASP13: the start of the deep learning era?

In the latest round of CASP, i.e. the thirteenth, the overall FM results demonstrated 'unprecedented success' [123]. The GDT averages of the best models submitted for FM targets in CASP12 and CASP13 increased from 52.9 to 65.7 respectively. The main reason was the employment of deep learning techniques to improve inter-residues distance predictions that were used as restraints for tertiary structure prediction [124]. More accurate contact predictions were mainly due to the dramatic fall of the number of template structures needed in the target's family to predict the contact maps. Although many methods employed deep learning techniques and showed remarkable results, AlphaFold (Group name: A7D) has shown some outstanding and unprecedented accuracies for many targets [125]. This system is the outcome of a collaborative London-based project between DeepMind, The Francis Crick Institute and University College London. Amongst the 112 assessed domains, taken from 80 targets, 32 were classified as FM, 13 as FM/TBM and 22 as TBM-Hard [87]. Taking into account all those 'hard' 67 targets and server groups only "BAKER-ROSETTASERVER" ranked fifth, while "Zhang-Server", "Quark", "RaptorX-DeepModeller" and "RaptorX-Contact" ranked first, second, third and fourth respectively. However, in terms of "research lab", "Baker-lab" was the third, whilst "Zhang-lab" and "RaptorX-lab" were first and second respectively. The same results were shown when only considering the 32 FM targets. On the other hand, it is important to mention that taking into account "all groups", A7D – registered as human/server group - stood out by far as the best group outperforming significantly all its competitors : while the sum of A7D's z-scores was 94.7, the group that ranked second -"Zhang"- only achieved 67.2.

Compared to its competitors, Rosetta server produced remarkable achievements for the following targets: T0953s1-D1 (67 residues, GDT 48.88), T0960-D2 (84 residues, GDT 56.55), T0968s1-D1 (118 residues, GDT 66.74), T0968s2-D1 (115 residues, GDT 71.30, Figure 7 (c)), T0975-D1 (281 residues, GDT 63.61- highest), T0978-D1 (413 residues, GDT 47.34), T0986s1-D1 (92 residues, GDT 68.21), T0992-D1 (107 residues, GDT 81.78), , T1000-D2 (368 residues, GDT 58.83), T1001-D1 (139 residues, GDT 73.20), T1005-D1 (326 residues, GDT 55.83), T1015s1-D1 (88 residues, GDT 57.67) and T1019s1-D1 (58 residues, GDT 83.62). In addition, the human/server group "BAKER" submitted two additional notable models: T0986s1-D1 (92 residues, GDT 79.08) and T1000-D2 (368 residues, GDT 63.18).

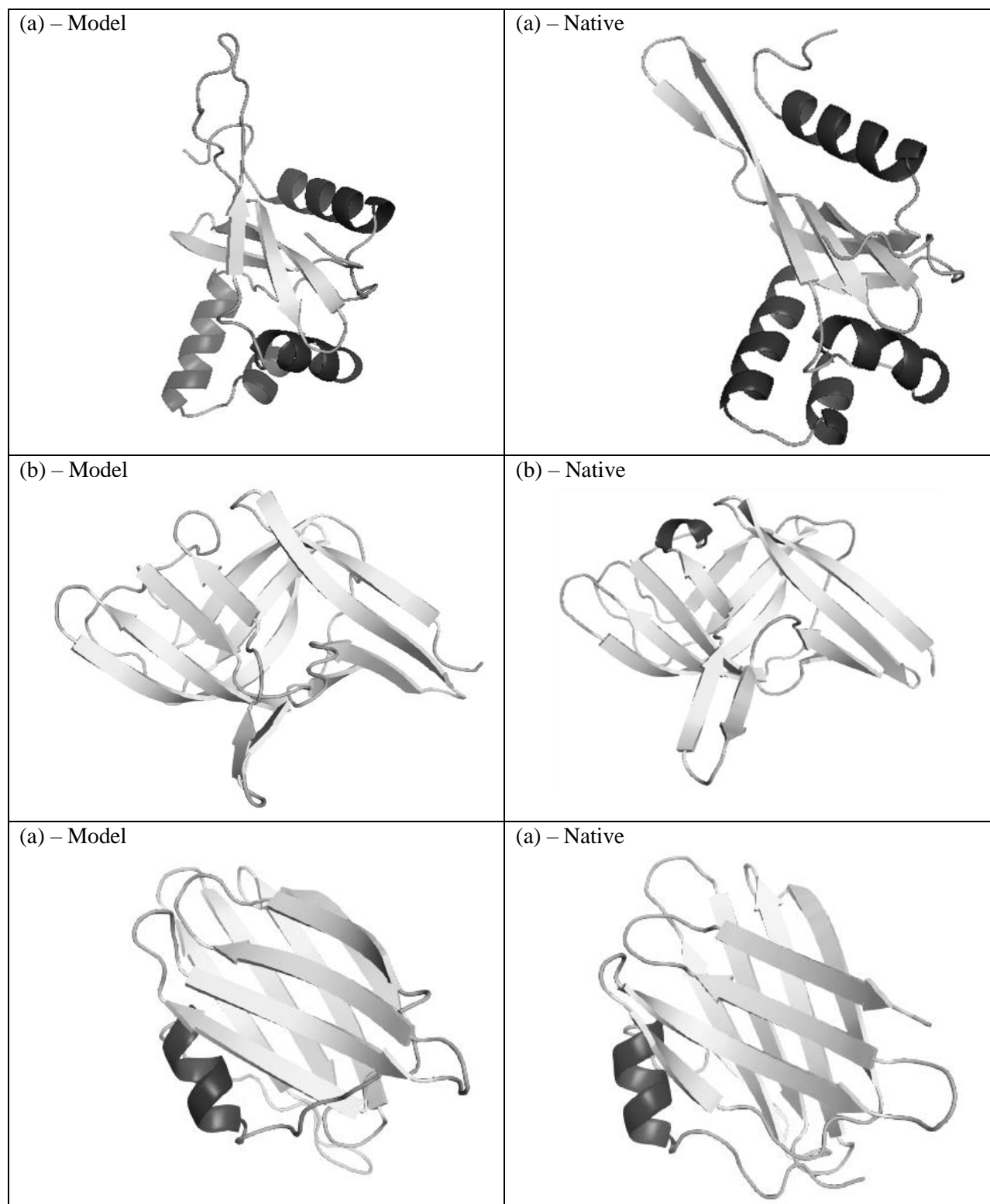


Figure 7 (a): Structure of the third model submitted by the Rosetta sever group in CASP11 for Target T079-D2 (left) and the native structure – PDB ID: 4L4W (right). Out of the 130 residues, 92 recorded an RMSD of 3.9 Å. (b) Structure of the first model by the Rosetta server in CASP12 for target T0860-D1 (left) and its corresponding native structure - PDB ID: 5FJL (right). (c) Structure of the third model by the Rosetta server in CASP13 for target T09680- D1 (left) and its corresponding native structure – PDB ID: 6CP9 (right).

Since CASP7, RMSD of the longest fragment has been no longer adopted, instead relatively small domains have been used as evaluation units, and GDT has become the official metric for ranking. Consequently, the pictorial summary from CASP7 to CASP13 – shown in Figure 8 and Figure 9 – takes into consideration the GDT that the Rosetta server group was able to achieve in FM targets. Figure 7 shows the longest target Rosetta was able to predict a model with GDT greater than 40 (since above that threshold, the overall fold is believed to be correct). In addition, Figure 8 reveals the highest GDT score Rosetta was able to achieve.

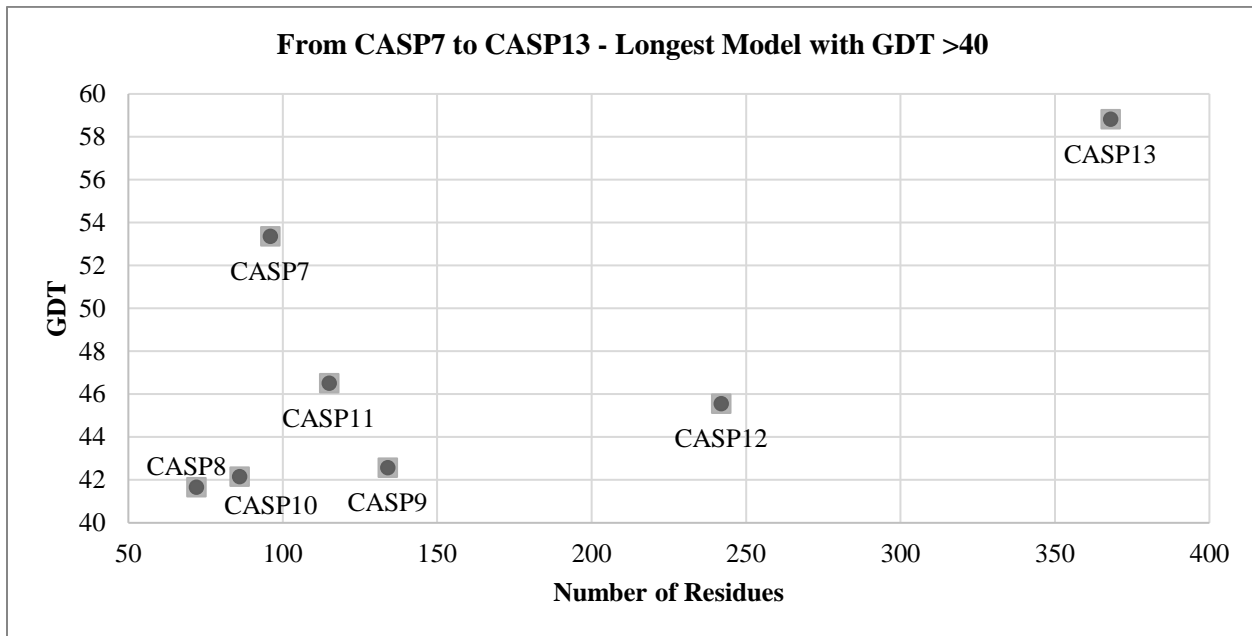


Figure 8: Achievements of Rosetta Server group from CASP7 to CASP13 in terms of the longest FM target where one of the models submitted recorded a GDT greater than 40.

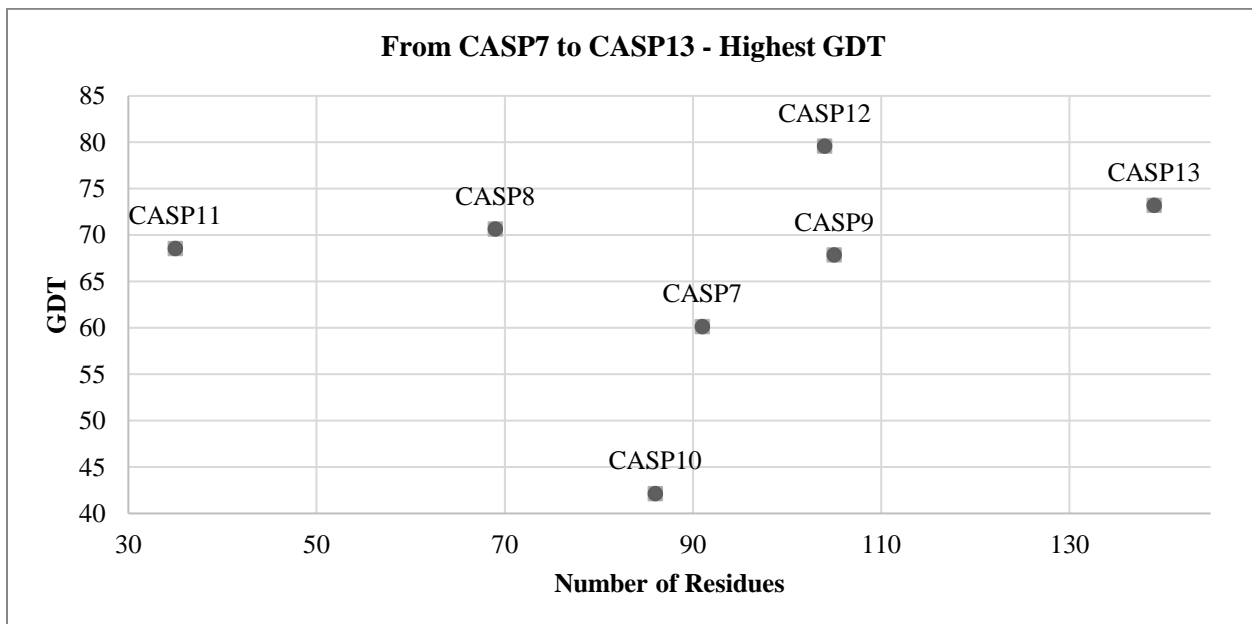


Figure 9: Achievements of Rosetta Server group from CASP7 to CASP13 in terms of the highest GDT score for FM targets.

## 9. Discussion

This mini-review describes the journey followed by the Rosetta fragment-based protein structure prediction. In its first part, it starts by reporting the preliminary studies that underpinned Rosetta's rationales and led to Rosetta's birth. It also details Rosetta's main components such as the fragment picker and fragment assembly. Its second part explores the contributions of Rosetta in CASP since the second round until the latest, i.e. CASP13, by highlighting Rosetta's main challenges and achievements in each round.

From CASP3 to CASP6, Rosetta gained its fame as it dominated the "ab initio" category over all competitors. From CASP7 to CASP10, Rosetta's "Baker-lab" at the University of Washington encountered serious competition with the "Zhang-lab" at the University of Michigan. "I-TASSER" and "Quark" took advantage of the combination of relatively long fragments, when available, resulting from successful threading modelling of some regions of the targets. In CASP11, ROLL and CASP12, Rosetta regained its domination in FM targets category. This may be due particularly to its usage of contact maps predictions as restraints whilst building conformations. In CASP13, usage of inter-residues distances predictions became generalised among top competitors. Moreover, their exploitation of advanced deep learning techniques led them to overtake Rosetta based groups.

Availability of Rosetta as an open source package has considerably participated in the widespread adaptation of some of its components in other molecular structures modelling. Not to add that Baker and his co-workers were amongst the first who made the 'public' contribute through Rosetta@home and Foldit. Arguably, Rosetta has been a pioneer that has paved the way to the emergence of many fragment-assembly techniques that have succeeded. Its early successes in the very early rounds of CASP were truly remarkable in the 'hard' targets category. At the time, Rosetta was able to demonstrate that fragment-based protein structure prediction was an efficient and accurate alternative to 'pure ab initio' approaches, i.e. expensive physics-based techniques to infer new folds.

CASP13 results have dramatically changed the context of the protein structure prediction problem. Whilst "Zhang-group" had been a tight competitor in the previous 6 rounds, A7D, thanks to its usage of deep learning techniques, showed itself as a much 'tougher' rival to Rosetta in the latest round. On the other hand, as there has been no article describing Rosetta's progress regarding contact map prediction since CASP12, it is unclear if Rosetta used an improved version in CASP13. The return of Rosetta to the top of the competition will require extensive improvements for contact map prediction that are likely to rely on deep learning technology. If Rosetta choses that route, the formidable processing power that Rosetta@Home provides could be the key of its success: instead of generating tremendous number of decoys, it could provide the means of training the most complex machine learning based predictors.

### Conflict of interest:

None

### References:

- [1] Lemmon G, Meiler J. Rosetta ligand docking with flexible XML protocols. *Methods Mol Biol* 2012; 819: 143–155.
- [2] Sircar A, Chaudhury S, Kilambi KP, et al. A generalized approach to sampling backbone conformations with RosettaDock for CAPRI rounds 13-19. *Proteins Struct Funct Bioinforma* 2010; 78: 3115–3123.
- [3] Guntas G, Purbeck C, Kuhlman B. Engineering a protein–protein interface using a computationally designed library. *Proc Natl Acad Sci* 2010; 107: 19296–19301.
- [4] Mandell DJ, Coutsiias EA, Kortemme T. Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. *Nat Methods* 2009; 6: 551–552.

- [5] Simons KT, Kooperberg C, Huang E, et al. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *J Mol Biol* 1997; 268: 209–25.
- [6] Leaver-Fay A, Tyka M, Lewis SM, et al. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol* 2011; 487: 545–74.
- [7] Langan RA, Boyken SE, Ng AH, et al. De novo design of bioactive protein switches. *Nature* 2019; 572: 205–210.
- [8] Baker D. What has de novo protein design taught us about protein folding and biophysics? *Protein Sci* 2019; 28: 678–683.
- [9] Marcos E, Chidyausiku TM, McShan AC, et al. De novo design of a non-local  $\beta$ -sheet protein with high stability and accuracy. *Nat Struct Mol Biol* 2018; 25: 1028–1034.
- [10] Lu P, Min D, DiMaio F, et al. Accurate computational design of multipass transmembrane proteins. *Science (80- )* 2018; 359: 1042–1046.
- [11] Silva D, Stewart L, Lam K, et al. Structures and disulfide cross-linking of *de novo* designed therapeutic mini-proteins. *FEBS J* 2018; 285: 1783–1785.
- [12] Chevalier A, Silva DA, Rocklin GJ, et al. Massively parallel de novo protein design for targeted therapeutics. *Nature* 2017; 550: 74–79.
- [13] Huang P-S, Boyken SE, Baker D. The coming of age of de novo protein design. *Nature* 2016; 537: 320–327.
- [14] Simons KT, Bonneau R, Ruczinski I, et al. Ab Initio Protein Structure Prediction of CASP III Targets Using ROSETTA. *Proteins Struct Funct Genet* 1999; 37: 171–176.
- [15] Orengo CA, Bray JE, Hubbard T, et al. Analysis and assessment of ab initio three-dimensional prediction, secondary structure, and contacts prediction. *Proteins Struct Funct Genet* 1999; 37: 149–170.
- [16] Kuhlman B, Bradley P. Advances in protein structure prediction and design. *Nature Reviews Molecular Cell Biology* 2019; 20: 681–697.
- [17] Dill KA, Ozkan SB, Shell MS, et al. The protein folding problem. *Annu Rev Biophys* 2008; 37: 289–316.
- [18] Perez A, Yang Z, Bahar I, et al. FlexE: Using Elastic network models to compare models of protein structure. *J Chem Theory Comput* 2012; 8: 3985–3991.
- [19] Scheraga HA. My 65 years in protein chemistry. *Q Rev Biophys* 2015; 48: 117–177.
- [20] Karplus M. The Levinthal paradox: Yesterday and today. *Fold Des*; 2. Epub ahead of print 1997. DOI: 10.1016/S1359-0278(97)00067-9.
- [21] Dill KA. Dominant forces in protein folding. *Biochemistry* 1990; 29: 7133–7155.
- [22] Lam SD, Das S, Sillitoe I, et al. An overview of comparative modelling and resources dedicated to large-scale modelling of genome sequences. *Acta Crystallogr Sect D Struct Biol* 2017; 73: 628–640.
- [23] Kmiecik S, Gront D, Kolinski M, et al. Coarse-Grained Protein Models and Their Applications. *Chemical Reviews* 2016; 116: 7898–7936.

- [24] KC DB. Recent advances in sequence-based protein structure prediction: Table 1. *Brief Bioinform* 2016; bbw070.
- [25] Deng H, Jia Y, Zhang Y. Protein structure prediction. *Int J Mod Phys B*; 32. Epub ahead of print 20 July 2018. DOI: 10.1142/S021797921840009X.
- [26] Bowie JU, Eisenberg D. An evolutionary approach to folding small alpha-helical proteins that uses sequence information and an empirical guiding fitness function. *Proc Natl Acad Sci U S A* 1994; 91: 4436–40.
- [27] Song Y, Tyka MD, Leaver-fay A, et al. Toward High-Resolution de Novo Structure Prediction for Small Proteins. *Science (80- )* 2005; 309: 1868–1871.
- [28] Subramani A, Wei Y, Floudas CA. ASTRO-FOLD 2.0: An enhanced framework for protein structure prediction. *AIChE J* 2012; 58: 1619–1637.
- [29] Kosciolok T, Jones DT. De novo structure prediction of globular proteins aided by sequence variation-derived contacts. *PLoS One* 2014; 9: e92197.
- [30] Yang J, Yan R, Roy A, et al. The I-TASSER Suite: protein structure and function prediction. *Yang, J, Yan, R, Roy, A, Xu, D, Poisson J, Zhang, Y (2015) I-TASSER Suite protein Struct Funct Predict Nat Meth, 12, 7–8* *Nat Meth* 2015; 12: 7–8.
- [31] Xu D, Zhang Y. Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. *Proteins* 2012; 80: 1715–35.
- [32] Abbass J, Nebel J-C. Reduced Fragment Diversity for Alpha and Alpha-Beta Protein Structure Prediction using Rosetta. *Protein Pept Lett* 2017; 24: 215–222.
- [33] Han KF, Baker D. Global properties of the mapping between local amino acid sequence and local structure in proteins. *Proc Natl Acad Sci U S A* 1996; 93: 5814–8.
- [34] Rohl CA, Strauss CEMEM, Misura KMSMS, et al. Protein structure prediction using Rosetta. *Methods Enzymol* 2004; 383: 66–93.
- [35] Han KF, Baker D. Global properties of the mapping between local amino acid sequence and local structure in proteins. 1996; 93: 5814–5818.
- [36] Han KF, Baker D. Recurring Local Sequence Motifs in Proteins. *J Mol Biol* 1995; 251: 176–187.
- [37] Bystroff C, Simons KT, Han KF, et al. Local sequence-structure correlations in proteins. *Current Opinion in Biotechnology* 1996; 7: 417–421.
- [38] Simons KT, Kooperberg C, Huang E, et al. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *J Mol Biol* 1997; 268: 209–25.
- [39] Simons KT, Ruczinski I, Kooperberg C, et al. Improved recognition of native-like protein structures using a combination of sequence-dependent and sequence-independent features of proteins. *Proteins* 1999; 34: 82–95.
- [40] Bystroff C, Baker D. Blind predictions of local protein structure in CASP2 targets using the I-sites library. *Proteins Struct Funct Genet* 1997; 29: 167–171.
- [41] Leaver-Fay A, O’Meara MJ, Tyka M, et al. Scientific benchmarks for guiding macromolecular energy function improvement. *Methods Enzymol* 2013; 523: 109–143.

- [42] O'Meara MJ, Leaver-Fay A, Tyka MD, et al. Combined covalent-electrostatic model of hydrogen bonding improves structure prediction with Rosetta. *J Chem Theory Comput* 2015; 11: 609–622.
- [43] Alford RF, Leaver-Fay A, Jeliaskov JR, et al. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J Chem Theory Comput* 2017; 13: 3031–3048.
- [44] Bradley P, Malmström L, Qian B, et al. Free modeling with Rosetta in CASP6. *Proteins Struct Funct Genet* 2005; 61: 128–134.
- [45] Chivian D, Kim DE, Malmström L, et al. Prediction of CASP6 structures using automated Robetta protocols. In: *Proteins: Structure, Function and Genetics*. 2005, pp. 157–166.
- [46] Das R, Qian B, Raman S, et al. Structure prediction for CASP7 targets using extensive all-atom refinement with Rosetta@home. In: *Proteins: Structure, Function and Genetics*, pp. 118–128.
- [47] Raman S, Vernon R, Thompson J, et al. Structure prediction for CASP8 with all-atom refinement using Rosetta. *Proteins* 2009; 77 Suppl 9: 89–99.
- [48] Lazaridis T, Karplus M. Effective energy function for proteins in solution. *Proteins Struct Funct Genet* 1999; 35: 133–152.
- [49] Lyskov S, Chou FC, Conchúir SÓ, et al. Serverification of Molecular Modeling Applications: The Rosetta Online Server That Includes Everyone (ROSIE). *PLoS One* 2013; 8: 5–7.
- [50] Dunbrack RL. Rotamer libraries in the 21st century. *Current Opinion in Structural Biology* 2002; 12: 431–440.
- [51] Dunbrack RL, Cohen FE. Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci* 1997; 6: 1661–1681.
- [52] Shapovalov M V., Dunbrack RL. A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions. *Structure* 2011; 19: 844–858.
- [53] Gront D, Kulp DW, Vernon RM, et al. Generalized fragment picking in Rosetta: design, protocols and applications. *PLoS One* 2011; 6: e23294.
- [54] McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. *Bioinformatics* 2000; 16: 404–5.
- [55] Lemam JK, Mueller R, Karakas M, et al. Simultaneous prediction of protein secondary structure and transmembrane spans. *Proteins Struct Funct Bioinforma* 2013; 81: 1127–1140.
- [56] Karplus K. SAM-T08, HMM-based protein structure prediction. *Nucleic Acids Res* 2009; 37: W492–W497.
- [57] Mirabello C, Pollastri G. Porter, PaleAle 4.0: high-accuracy prediction of protein secondary structure and relative solvent accessibility. *Bioinformatics* 2013; 29: 2056–2058.
- [58] Altschul SF, Madden TL, Schäffer AA, et al. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research* 1997; 25: 3389–3402.
- [59] Kirkpatrick S, Gelatt CD, Vecchi MP. Optimization by simulated annealing. *Science* 1983; 220: 671–680.
- [60] Metropolis N, Rosenbluth A, Rosenbluth M, et al. Equation of State Calculations by Fast Computing Machines. *J Chem Phys* 1953; 21: 1087.
- [61] Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server.

*Nucleic Acids Res*; 32.

- [62] Song Y, Dimaio F, Wang RYR, et al. High-resolution comparative modeling with RosettaCM. *Structure* 2013; 21: 1735–1742.
- [63] Wollacott AM, Zanghellini A, Murphy P, et al. Prediction of structures of multidomain proteins from structures of the individual domains. *Protein Sci* 2007; 16: 165–175.
- [64] Soding J. Protein homology detection by HMM-HMM comparison. *Bioinformatics*; 21. Epub ahead of print 2005. DOI: 10.1093/bioinformatics/bti125.
- [65] Yang Y, Faraggi E, Zhao H, et al. Improving protein fold recognition and template-based modeling by employing probabilistic-based matching between predicted one-dimensional structural properties of query and corresponding native properties of templates. *Bioinformatics* 2011; 27: 2076–2082.
- [66] Peng J, Xu J. Raptorx: Exploiting structure information for protein alignment by statistical inference. *Proteins Struct Funct Bioinforma* 2011; 79: 161–171.
- [67] Altschul SF. Basic Local Alignment Search Tool. *J Mol Biol* 1990; 215: 403–410.
- [68] Jaroszewski L, Godzik A, Rychlewski L. Improving the quality of twilight-zone alignments. *Protein Sci* 2000; 9: 1487–1496.
- [69] Rychlewski L, Li W, Jaroszewski L, et al. Comparison of sequence profiles. Strategies for structural predictions using sequence information. *Protein Sci* 2008; 9: 232–241.
- [70] Ginalski K, Rychlewski L. Detection of reliable and unexpected protein fold predictions using 3D-Jury. *Nucleic Acids Res* 2003; 31: 3291–3292.
- [71] Kamisetty H, Ovchinnikov S, Baker D. Assessing the utility of coevolution-based residue-residue contact predictions in a sequence- and structure-rich era. *Proc Natl Acad Sci U S A* 2013; 110: 15674–9.
- [72] Ovchinnikov S, Park H, Varghese N, et al. Protein structure determination using metagenome sequence data. *Science* 2017; 355: 294–298.
- [73] Uziela K, Wallner B. ProQ2: Estimation of model accuracy implemented in Rosetta. *Bioinformatics* 2016; 32: 1411–1413.
- [74] Baker D. Centenary Award and Sir Frederick Gowland Hopkins Memorial Lecture. Protein folding, structure prediction and design. *Biochem Soc Trans* 2014; 42: 225–9.
- [75] Tyka MD, Keedy DA, André I, et al. Alternate states of proteins revealed by detailed energy landscape mapping. *J Mol Biol* 2011; 405: 607–618.
- [76] Cooper S, Khatib F, Treuille A, et al. Predicting protein structures with a multiplayer online game. *Nature* 2010; 466: 756–760.
- [77] Cooper S, Baker D, Popović Z, et al. The challenge of designing scientific discovery games. In: *Proceedings of the Fifth International Conference on the Foundations of Digital Games - FDG '10*. New York, New York, USA: ACM Press, pp. 40–47.
- [78] Eiben CB, Siegel JB, Bale JB, et al. Increased Diels-Alderase activity through backbone remodeling guided by Foldit players. *Nat Biotech* 2012; 30: 190–192.
- [79] Gilski M, Kazmierczyk M, Krzywda S, et al. High-resolution structure of a retroviral protease



- folded as a monomer. *Acta Crystallogr Sect D* 2011; 67: 907–914.
- [80] Khatib F, Cooper S, Tyka MD, et al. From the Cover: Algorithm discovery by protein folding game players. *Proc Natl Acad Sci* 2011; 108: 18949–18953.
- [81] Kleffner R, Flatten J, Leaver-Fay A, et al. Foldit Standalone: a video game-derived protein structure manipulation interface using Rosetta. *Bioinformatics* 2017; 33: 2765–2767.
- [82] Andersen CAF, Palmer AG, Brunak S, et al. Continuum secondary structure captures protein flexibility. *Structure* 2002; 10: 175–184.
- [83] Kinch LN, Shi S, Cheng H, et al. CASP9 target classification. *Proteins Struct Funct Bioinforma* 2011; 79: 21–36.
- [84] Kinch LN, Li W, Schaeffer RD, et al. CASP 11 Target Classification. *Proteins Struct Funct Bioinforma* 2016; 84: 20–33.
- [85] Tress ML, Ezkurdia I, Richardson JS. Target domain definition and classification in CASP8. *Proteins: Structure, Function and Bioinformatics* 2009; 77: 10–17.
- [86] Clarke ND, Ezkurdia L, Kopp J, et al. Domain definition and target classification for CASP7. *Proteins Struct Funct Genet* 2007; 69: 10–18.
- [87] Kinch LN, Kryshchuk A, Monastyrskyy B, et al. CASP13 target classification into tertiary structure prediction categories. *Proteins Struct Funct Bioinforma* 2019; 87: 1021–1036.
- [88] Zemla A. LGA: a method for finding 3D similarities in protein structures. *Nucleic Acids Res* 2003; 31: 3370–3374.
- [89] Aloy P, Stark A, Hadley C, et al. Predictions Without Templates: New Folds, Secondary Structure, and Contacts in CASP5. *Proteins Struct Funct Genet* 2003; 53: 436–456.
- [90] Schrödinger, LLC. *The {PyMOL} Molecular Graphics System, Version~1.8*. November 2015.
- [91] Cozzetto D, Kryshchuk A, Ceriani M, et al. Assessment of predictions in the model quality assessment category. In: *Proteins: Structure, Function and Genetics*. 2007, pp. 175–183.
- [92] Moulton J, Fidelis K, Kryshchuk A, et al. Critical assessment of methods of protein structure prediction: Progress and new directions in round XI. *Proteins Struct Funct Bioinforma* 2016; 84: 4–14.
- [93] Lesk AM. CASP2: Report on ab initio predictions. *Proteins Struct Funct Genet* 1997; 29: 151–166.
- [94] Lesk AM, Conte L Lo, Hubbard TJP. Assessment of novel fold targets in CASP4: Predictions of three-dimensional structures, secondary structures, and interresidue contacts. *Proteins Struct Funct Genet* 2001; 45: 98–118.
- [95] Bonneau R, Tsai J, Ruczinski I, et al. Rosetta in CASP4: Progress in ab initio protein structure prediction. *Proteins Struct Funct Genet* 2001; 45: 119–126.
- [96] Bradley P, Chivian D, Meiler J, et al. Rosetta predictions in CASP5: Successes, failures, and prospects for complete automation. *Proteins Struct Funct Genet* 2003; 53: 457–468.
- [97] Vincent JJ, Tai C-H, Sathyanarayana BK, et al. Assessment of CASP6 predictions for new and nearly new fold targets. *Proteins* 2005; 61 Suppl 7: 67–83.
- [98] Jauch R, Yeo HC, Kolatkar PR, et al. Assessment of CASP7 structure predictions for template free

- targets. *Proteins Struct Funct Bioinforma* 2007; 69: 57–67.
- [99] Ben-David M, Noivirt-Brik O, Paz A, et al. Assessment of CASP8 structure predictions for template free targets. *Proteins Struct Funct Bioinforma* 2009; 77: 50–65.
- [100] Zhang J, Wang Q, Barz B, et al. MUFOLD: A new solution for protein 3D structure prediction. *Proteins* 2010; 78: 1137–52.
- [101] Kinch LN, Yong Shi S, Cong Q, et al. CASP9 assessment of free modeling target predictions. *Proteins Struct Funct Bioinforma* 2011; 79: 59–73.
- [102] Tai CH, Bai H, Taylor TJ, et al. Assessment of template-free modeling in CASP10 and ROLL. *Proteins Struct Funct Bioinforma* 2014; 82: 57–83.
- [103] Kinch LN, Li W, Monastyrskyy B, et al. Evaluation of free modeling targets in CASP11 and ROLL. *Proteins Struct Funct Bioinforma* 2016; 84: 51–66.
- [104] Mabrouk M, Putz I, Werner T, et al. RBO Aleph: leveraging novel information sources for protein structure prediction. *Nucleic Acids Res* 2015; 43: W343–W348.
- [105] Zhang Y. Interplay of I-TASSER and QUARK for template-based and ab initio protein structure prediction in CASP10. *Proteins* 2014; 82 Suppl 2: 175–87.
- [106] Ovchinnikov S, Kim DE, Wang RYR, et al. Improved de novo structure prediction in CASP11 by incorporating coevolution information into Rosetta. *Proteins Struct Funct Bioinforma* 2016; 1–9.
- [107] Joo K, Joung IS, Cheng Q, et al. Contact-assisted protein structure modeling by global optimization in CASP11. *Proteins Struct Funct Bioinforma* 2016; 84: 189–199.
- [108] S O, H K, D B. Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. *Elife* 2014; 3: e02030.
- [109] Park I-H, Gangupomu V, Wagner J, et al. Structure Refinement of Protein Low Resolution Models Using the GNEIMO Constrained Dynamics Method. *J Phys Chem B* 2012; 116: 2365–2375.
- [110] Li SC, Bu D, Xu J, et al. Fragment-HMM: a new approach to protein structure prediction. *Protein Sci* 2008; 17: 1925–34.
- [111] Khoury GA, Smadbeck J, Kieslich CA, et al. Princeton\_TIGRESS 2.0: High refinement consistency and net gains through support vector machines and molecular dynamics in double-blind predictions during the CASP11 experiment. *Proteins Struct Funct Bioinforma* 2017; 85: 1078–1098.
- [112] Kandathil SM, Handl J, Lovell SC. Toward a detailed understanding of search trajectories in fragment assembly approaches to protein structure prediction. *Proteins Struct Funct Bioinforma* 2016; 84: 411–426.
- [113] Karakaş M, Woetzel N, Staritzbichler R, et al. BCL::Fold - De Novo Prediction of Complex and Large Protein Topologies by Assembly of Secondary Structure Elements. *PLoS One*; 7. Epub ahead of print 16 November 2012. DOI: 10.1371/journal.pone.0049240.
- [114] Li J, Deng X, Eickholt J, et al. Designing and benchmarking the MULTICOM protein structure prediction system. *BMC Struct Biol*; 13. Epub ahead of print 2013. DOI: 10.1186/1472-6807-13-2.
- [115] Shrestha R, Zhang KYJ. Improving fragment quality for de novo structure prediction. *Proteins Struct Funct Bioinforma* 2014; 82: 2240–2252.

- [116] Li Y, Liu H, Rata I, et al. Building a Knowledge-Based Statistical Potential by Capturing High-Order Inter-residue Interactions and its Applications in Protein Secondary Structure Assessment. *J Chem Inf Model* 2013; 53: 500–508.
- [117] Simoncini D, Zhang KYJ. Efficient Sampling in Fragment-Based Protein Structure Prediction Using an Estimation of Distribution Algorithm. *PLoS One* 2013; 8: 1–10.
- [118] Abbass J, Nebel J-C. Customised fragments libraries for protein structure prediction based on structural class annotations. *BMC Bioinformatics* 2015; 16: 136.
- [119] Abbass J, Nebel J-C. SCOP-Aided Fragment Assembly Protein Structure Prediction. In: *2019 Fourth International Conference on Advances in Computational Tools for Engineering Applications (ACTEA)*. 2019, pp. 1–5.
- [120] Moutl J, Fidelis K, Kryshtafovych A, et al. Critical assessment of methods of protein structure prediction (CASP)—Round XII. *Proteins Struct Funct Bioinforma* 2018; 86: 7–15.
- [121] Ovchinnikov S, Park H, Kim DE, et al. Protein structure prediction using Rosetta in CASP12. *Proteins Struct Funct Bioinforma* 2018; 86: 113–121.
- [122] Trevizani R, Custódio FL, Dos Santos KB, et al. Critical Features of Fragment Libraries for Protein Structure Prediction. *PLoS One* 2017; 12: e0170131.
- [123] Kryshtafovych A, Schwede T, Topf M, et al. Critical assessment of methods of protein structure prediction (CASP)-Round XIII. *Proteins Struct Funct Bioinforma*. Epub ahead of print October 2019. DOI: 10.1002/prot.25823.
- [124] Kandathil SM, Greener JG, Jones DT. Recent developments in deep learning applied to protein structure prediction. *Proteins Struct Funct Bioinforma* 2019; 87: 1179–1189.
- [125] Senior AW, Evans R, Jumper J, et al. Protein structure prediction using multiple deep neural networks in the 13th Critical Assessment of Protein Structure Prediction (CASP13). *Proteins* 2019; 87: 1141–1148.