Review

Predictive and Experimental Approaches for Elucidating Protein–Protein Interactions and Quaternary Structures

John Oliver Nealon, Limcy Seby Philomina and Liam James McGuffin * iD

School of Biological Sciences, University of Reading, Reading RG6 6AS, UK;
j.o.nealon@pgr.reading.ac.uk (J.O.N.); l.s.philomina@pgr.reading.ac.uk (L.S.P.)
* Correspondence: l.j.mcguffin@reading.ac.uk; Tel.: +44-(0)118-378-6332

Abstract: The elucidation of protein–protein interactions is vital for determining the function and action of quaternary protein structures. Here, we discuss the difficulty and importance of establishing protein quaternary structure and review in vitro and in silico methods for doing so. Determining the interacting partner proteins of predicted protein structures is very time-consuming when using in vitro methods, this can be somewhat alleviated by use of predictive methods. However, developing reliably accurate predictive tools has proved to be difficult. We review the current state of the art in predictive protein interaction software and discuss the problem of scoring and therefore ranking predictions. Current community-based predictive exercises are discussed in relation to the growth of protein interaction prediction as an area within these exercises. We suggest a fusion of experimental and predictive methods that make use of sparse experimental data to determine higher resolution predicted protein interactions as being necessary to drive forward development.

Keywords: protein; interaction; prediction; homology; docking; quality assessment; experimental

1. Introduction

Accurately predicting how a protein folds, functions and interacts with other molecules, remains one of the most critical problems in bioinformatics. An entirely algorithmic solution to elucidate protein folding would be ideal, and a folding process of some form must exist in nature as postulated in Levinthal's paradox [1]. However, discovering such a solution has eluded the modelling community and as a result, the most effective methods of protein fold prediction use template based approaches. Existing protein structures are compared to the sequence that is being predicted utilizing the extensive library of protein structures and complexes assembled from experimental procedures such as electron microscopy, nuclear magnetic resonance (NMR) spectroscopy, X-ray crystallography and small-angle X-ray scattering (SAXS). X-ray crystallography gives a complete, high-resolution analysis of the three-dimensional structure of proteins, but large, well-ordered single crystals are required. Electron diffraction and electron microscopy (EM) are used mainly where crystals cannot be obtained due to very large complexes, but the drawback of these methods is low-resolution structures. Even though NMR is the leading method for structure determination, after X-ray crystallography, it has limitations in terms of sensitivity, selectivity and target analysis [2,3]. SAXS can be carried out using a wide variety of sample conditions from frozen to the natural solution, although it is lower resolution than NMR and X-ray crystallography [4].

Determining protein structures and their function from sequence data is an essential part of modern molecular biology and leads onto the problem of determining the protein interacting partners and quaternary structural state. A thorough review of scientific literature and progressively reliable computational predictions have resulted in the formation of massive databases of protein interaction

data. These databases have become an essential resource for new predictive methods [5]. The prediction of protein–protein interactions is vital for accurately positioning proteins within signalling pathways and networks. The importance of the development of reliable predictive tools continues to increase with the explosion of genomic and proteomic data [6]. Over the past few decades, the critical assessment of techniques for protein structure prediction (CASP) [7], the critical assessment of prediction of interactions (CAPRI) [8] and the continuous automated model evaluation (CAMEO) [9] prediction experiments have acted as a catalysis spurring the development of new quaternary structure modelling techniques. These experiments take the form of competition, and the results allow researchers to objectively assess the quality of different methods for prediction of protein structures, functions and interactions. In recent years the CASP experiment has included the assessment of quaternary structures, where predictors are encouraged to submit the oligomeric assemblies of the protein targets. Furthermore, the recent addition of the Data Assisted category in CASP aims to evaluate how much the accuracy of 3D models can be improved by the addition of sparse data from NMR, crosslinking, and SAXS experiments.

2. Experimental Methods for Determining Protein–Protein Interactions (PPIs)

Various rapid bioinformatics tools are available for the inherently complex process of PPI prediction, yet experimental methods are still considered more reliable due to the current accuracy limitations of computational methodologies. However, determining the full atom coordinates for a protein complex or quaternary structure, for example using X-ray crystallography, is very costly and time-consuming and the crystallisation process itself can be hit and miss. Alternative experimental techniques for PPI detection are more reliable, but they do not provide detailed atomic information of quaternary structures. Therefore, there is a growing need for techniques that aim to fuse sparse experimental data with predictive data to aid with the modelling of complexes.

As per the statistics from IntAct [10], an open source database system for molecular interaction data, 523,010 interactions (either curated from the literature or direct data depositions) have been submitted until 3 October 2017. The experimental interaction detection methods can be broadly classified into: (i) biochemical; (ii) biophysical; (iii) genetic interference; (iv) imaging technique; (v) phenotype based detection assays; (vi) post-transcriptional interference; and (vii) protein complementation assays. The most common techniques as per IntAct are illustrated in Figure 1. In this section, we focus on advantages and limitations of each method.
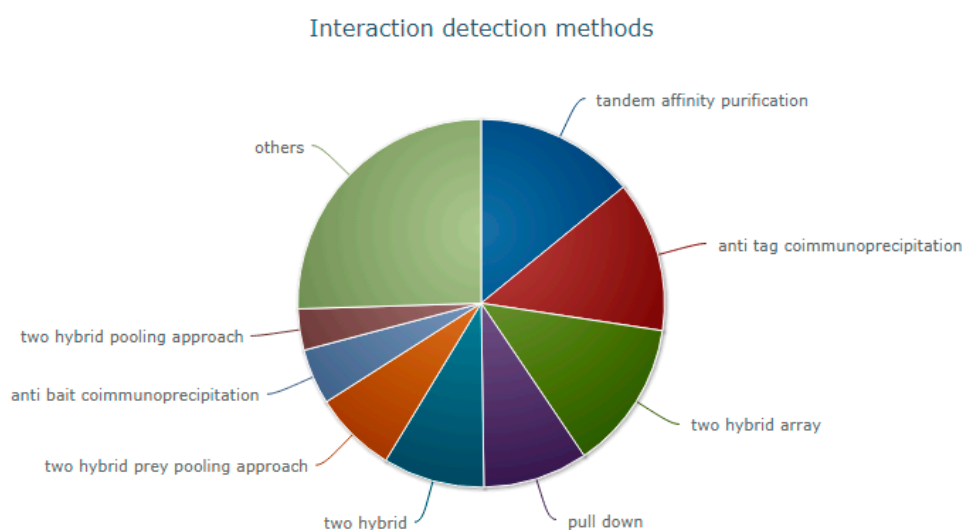

Figure 1. Interaction detection methods.

Tandem affinity purification (TAP) is a biochemical PPI detection method, which accounts for 14.1% of the experimental methods as per IntAct [10]. TAP is a proteomic high throughput technique, which was initially developed in yeast, but can be adapted to a variety of organisms. Wide applicability and simplicity make TAP a useful method for protein purification [11]. This method involves fusion of TAP tag, either N- or C-terminally to the target protein of interest, which is then encoded into the organism, where they form physiological complexes with other proteins. SDS-PAGE and mass spectrometry is then used to examine and identify proteins extracted with the tagged bait [12]. In this highly sensitive and selective method, transient or weak interactions may be lost during the series of purification stages and requirement of a large number of samples is also a disadvantage as it is difficult to purify and identify low abundance binding partners [13]. TAP is used to detect multiple PPIs at one time, but errors can be expected due to the interference of tags in proteins, which can be overcome by subsequent characterization [14].

If antibodies against the target protein are available, Co-immunoprecipitation (Co-IP) is considered to be a simple technique to study PPIs in vivo [15]. In this method, a protein-specific antibody is incubated with a protein mixture, which forms an immune complex with the target protein. Target proteins may interact with other proteins to form a protein complex, which is then immunoprecipitated using immobilized Protein A or Protein G [16]. As protein complexes can be part of large complexes, the co-immunoprecipitating proteins may not interact directly. Due to the lower concentration of antigen, precipitation is not as intensive as other methods [17]. Another disadvantage of this biochemical method is that during the elution of precipitated antigen, it releases antibody which contaminates the antigen [15,16].

Yeast two-hybrid is a high throughput PPI detection method, which accounts for 9.2% of the PPI detection methods submitted in IntAct [10]. In this genetic interference method, which was described initially by Field and Song in 1989, functional transcription factors formed by the interaction between bait and prey protein induces the specific reporter gene expression, allowing the detection of such interactions [18]. Despite being simple to set up, low cost and useful in detecting transient and weak interactions [19], the estimated the high rate of false positives (as high as 50%) suggests any detected interactions must be further verified by other methods [20]. Also, proteins, which are less likely to be present in the nucleus, are excluded, as proteins must be localised to the nucleus for this method. Furthermore, proteins that are not in their natural physiological environment may not fold correctly to interact [21].

The pull-down method, which involves the use of affinity purification, is similar to CoIP in methodology, i.e., uses affinity ligand to capture interacting proteins. This method uses "bait", purified and tagged protein in place of immobilised antibodies in CoIP [22]. Interactions mediated between the immobilized bait protein and a target protein by the negatively charged nucleic acid, which stick to the basic surface on the protein, may generate a false positive result in protein–protein interaction assays [23].

In protein chip technology, expressed purified and screened proteins are printed onto a chip using a microarrayer as discrete spots. A solution containing labelled proteins is incubated with the chip, which is then washed and the position of the labels indicates the interaction between proteins (protein on the chip and protein from the solution) [24]. High signal to noise ratio, high sensitivity and the relatively small quantity of sample requirement makes protein chip technology preferable over other techniques, but the proteins attached to the chip can disrupt protein interactions [25,26]. Even though, current advances in the manufacture of protein chip technology use glass microscope slides and other materials that allow the protein to attach on their surface at high density, the technological challenges in this field still remain [24].

Even though X-ray crystallography is the most popular method of protein structure determination, NMR spectroscopy has played a vital role where protein complexes have disorders and when challenging to obtain a crystal. NMR is handy in weak detection PPIs (K_d > ~100 μM) and hence is a

method preferred to study PPI in excellent detail [27]. As this biophysical technique gives atomic level information in the solution state, this method is of particular interest for PPI analysis by researchers [28].

Each experimental technique has its strengths. However, most of these techniques require expensive and extensive instrumentation and specific knowledge to analyze the results. Hence the field is developing and new and improved methods frequently emerging. An overview of the current experimental methods available is shown in Table 1.

Table 1. Overview of experimental methods for PPI detection and their categories.

Interaction Detection Method	Current Available Techniques	Reference
Biochemical	Affinity technology	[29]
	Aggregation assay	[30]
	Chromatography technology	[29]
	Cosedimentation	[29]
	Cross-linking study	[29]
	Electrophoretic mobility-based method	[31]
	Enzymatic study	[29]
	Footprinting	[29]
	Nucleotide exchange assay	[32]
	Polymerization	[33]
	Probe interaction assay	[29]
Biophysical	Biosensor	[34]
	Circular dichroism	[35]
	Mass spectrometry	[29]
	Differential scanning calorimetry	[36]
	Electron diffraction	[37]
	Electron resonance	[29]
	Enzyme-mediated activation of radical sources	[38]
	Equilibrium dialysis	[39]
	Filter trap assay	[40]
	Fluorescence technology	[30]
	Infrared spectroscopy	[41]
	Intermolecular force	[42]
	Isothermal titration calorimetry	[43]
	Light scattering	[44]
	Luminescence technology	[45]
	Microscale thermophoresis	[46]
	Molecular sieving	[17]
	Neutron diffraction	[47]
	Neutron fibre diffraction	[48]
	Nuclear magnetic resonance	[49]
	Rheology measurement	[50]
	Scintillation proximity assay	[51]
	Small angle neutron scattering	[35]
	Thermal shift binding	[52]
	Ultraviolet- visible spectroscopy	[53]
	X-ray crystallography	[29]
Genetic interference	Chemical RNA modification plus base	[54]
	Random spore analysis	[29]
	Synthetic genetic analysis	[29]
Imaging techniques	Atomic force microscopy	[55]
	Confocal microscopy	[29]
	Electron microscopy	[56]
	Fluorescence microscopy	[29]
	Light microscopy	[29]
	Super-resolution microscopy	[29]
	X-ray tomography	[29]

Table 1. *Cont.*

Interaction Detection Method	Current Available Techniques	Reference
Phenotype-based detection assay	Nuclear translocation assay	[57]
Post-transcriptional interference	Antisense oligonucleotides	[29]
	Antisense RNA	[58]
	RNA interference	[59]
Protein complementation assays	Adenylate cyclase complementation	[60]
	β-galactosidase complementation	[61]
	β lactamase complementation	[62]
	Bimolecular fluorescence complementation	[63]
	Dihydrofolate reductase reconstruction	[64]
	Mammalian protein–protein interaction trap	[65]
	Protein kinase A complementation	[66]
	Reverse ras recruitment system	[67]
	Split luciferase complementation	[68]
	Tox-R dimerization assay	[69]
	Transcriptional complementation assay	[29]

Sparse Experimental Data on Prediction of Protein Interactions and Modelling of Their Complexes

SAXS can provide a wealth of structural information on biomolecules in solution [70]. Advanced usage of SAXS can provide a unique insight into biomolecular behaviour that can only be observed in solution, like transient protein–protein interactions [71]. As SAXS data miss the relevant atomic details, the low-resolution information can be combined with theoretical methods which give the energetic description and atomic details of the interactions [72].

Tang et al. used a method by combining sparse NMR data with evolutionary couplings (ECs), which involves simultaneous analysis of ECs, derived from multiple sequence alignments (MSAs), with NMR chemical shift, nuclear Overhauser effect (NOE) and residual dipolar coupling (RDC) data. The results with this EC-NMR method were said to be more accurate and provided complete structural information compared to the ones obtained with sparse NMR data or ECs alone [73].

In 2007, Latek, Ekonomiuk, and Kolinski combined de novo modelling with limited experimental data and demonstrated that limited experimental data, such as chemical shifts, are sufficient for three-dimensional structure determination if the data are accurate. It was also suggested that weak quality chemical shift data can be improved by the application of additional sparse experimental data [74].

Recently, an integrative modelling approach was used to determine the 3D model of the entire Mediator complex. Robinson et al. combined data from X-ray crystallography, homology modelling, and cryo-electron microscopy with the results from chemical cross-linking and mass spectrometry to produce the model. All Mediator subunits, subunit interfaces and some secondary structural elements were revealed in the elucidated model [75].

Cross-linking is a biochemical PPI detection method, which is suitable for detecting weak interactions. As the reagent used in this technique detects interactions which are not in direct contact, the interactions detected with this method need to be assessed independently. Combined with mass spectrometry, chemical cross-linking can be used to study PPIs. In purified protein complexes they are used to study spatial relationships. Both transient and stable interactions can be studied using the information from the controlled arrays produced using new classes of photo-activated reagents [76,77].

The requirement of only a small amount of sample and the lack of a requirement for crystallisation has made cryo-electron microscopy (cryo-EM) a method preferable over other techniques [78,79]. Statistics from the Electron Microscopy Data Bank, a public repository for electron microscopy density maps of macromolecular complexes and subcellular structures indicates that cryo-EM is becoming a favourite structural biology technique to study challenging biological systems [80]. The high cost of EM equipment, the requirement of expensive maintenance and enormous computational resources

are the current challenges of cryo-EM [78]. However, cryo-EM is suggested as an ideal platform for the integration and modelling of data from other experimental methods for structural biology [81]. Hong-Wei Wang and Jia-Wei Wang explain the technique as a complementary method to X-ray crystallography. They describe how intermediate resolution maps from cryo-EM can assist in X-ray crystallography structure determination of macromolecular complexes [82]. Ning Gao and co-workers used a hybrid method combining structural information, chemical cross-linking proximity mapping followed by mass spectrometry (CX-MS) and cryo-EM for the modelling of five assembly factors in the ITS2 region of the pre-60S ribosome [83]. Residue-level details of PPI of dengue virus coat proteins was predicted using $C\alpha$ atom positions from low-resolution cryo-EM structures [84,85].

Experimental methods have proven invaluable to further our understanding of PPIs. However, each method has its disadvantages. Also, the relatively high cost, time and expertise requirements of in vitro techniques suggest that in silico methods will need to play an increasingly important role for understanding protein–protein interactions at the atomic level.

3. In Silico Methods for Modelling Protein–Protein Complexes

There are two primary categories of methods for modelling protein–protein complexes: (1) homology or template-based modelling and (2) ab initio docking, or template-free modelling. Template-based modelling relies on the existence of appropriate templates in protein databases, and docking approaches can be broadly split into rigid body and flexible solutions. Docking guided by restraints obtained from experimental procedures could be considered a third category and is often referred to as "hybrid modelling". However, the computational pipelines used in any integrative approach would usually comprise either template modelling or docking, or both. For protein structure prediction, ab initio methods have gained importance because the number of quaternary structural models available in the Protein Data Bank (PDB) [86] limits the well-established traditional method of template-based modelling, which is successfully routinely used for tertiary structure prediction. This is true despite the fact that the PDB is growing at an increasing rate due recombinant DNA methods allowing large amounts of target biological macromolecules to be expressed for analysis [2].

The increase in the rate of sequencing of biological macromolecules means that ab initio methods for determining structures and interactions of proteins are becoming more relevant. Protein docking is the computational determination of protein complex structure from individual protein structures. When using docking software, there are two classes of problems: bound docking, where the complex structure is known and the receptor and ligand are pulled apart then reassembled, and unbound docking, where individually determined proteins structures are used. Docking is computationally expensive, and the accuracy can decrease rapidly when the protein chains undergo significant conformational changes upon binding. A brief description of the variety of different methods utilized in homology modelling and docking follows, the methods have been chosen to be representative of some common successful methods as well as some less common approaches.

3.1. Computational Methods for Protein–Protein Interaction Modelling

RosettaDock is a soft body type of protein interaction prediction software and considers backbone or side-chain flexibility, which is advantageous regarding obtaining structures of protein complexes with high-resolution accuracy. The protein docking server predicts the structure of protein complexes given the compositions of the individual components and a relative binding orientation. The server identifies low-energy conformations of a protein–protein interaction near a given starting configuration by optimising rigid-body orientation and side-chain structures [87]. RosettaDock has made a substantial effort in developing flexible-backbone docking approaches in CAPRI rounds 6–12. However, the rigid-body ZDOCK [88] approach was more successful for a number of the targets. ZDOCK and other approaches employing softer potentials are less sensitive to backbone conformational changes.

ZDOCK, which is presently one of the most common rigid-body docking engines, uses a scoring function that includes shape complementarity, electrostatics and a heuristic potential called atomic

contact energy [88–90]. The combination of specific backbone optimization plus a strict potential should be optimal in the high sampling limit where the correct conformation is sure to be encountered, but with insufficient sampling, the results can be inferior to fixed-backbone sampling with a soft potential [91]. A recent development of RosettaDock is the use of SAXS data as a constraint step in selecting predicted complexes by using the SAXS data to reject complexes that violate shape constraints imposed by the SAXS data [92] thereby increasing the accuracy of final predictions.

Rigid body docking as a method produces a large number of docked conformations with favorable surface complementarity usually followed by ranking using energy scoring. The fast Fourier transform (FFT) algorithm uses a correlation approach [93] and systematically explores the space of possible docked conformations using electrostatic interactions or both electrostatic and solvation terms, but the potential is restricted to a correlation function form. FFT approaches are very common and include GRAMM-X which extends the original GRAMM FFT method by employing smoothed potentials, refinement stage, and knowledge-based scoring [94]. The Hex Protein Docking Server (HexServer) is an FFT based protein-docking server that utilizes graphics processing unit (GPU) acceleration. By using multiple GPUs, a typical docking run takes approximately 15 s, which is up to two orders of magnitude faster than conventional FFT based docking approaches using comparable resolution and scoring functions [95] such as GRAMM-X, FRODOCK and MEGADOCK.

MEGADOCK [96] uses a Katchalski–Katzir algorithm [93] and searches probable docking structures in a grid-based 3D space using FFT. MEGADOCK employs a scoring function in which only shape complementarity and electrostatics are considered and thus makes the calculations 8.8 times faster than ZDOCK. The method is set up to perform massive numbers of calculations that are run on parallel computing systems, and along with GPU.proton.DOCK and HexServer demonstrates the increase in search speed when using hardware acceleration.

FRODOCK projects the interaction terms of a potential protein complex into 3D grid-based potentials using spherical harmonics approximations to accelerate the search; this is itself an extension of the FFT alogrithm. The binding energy of the complex as it is formed is approximated as a correlation function composed of van der Waals, electrostatics and desolvation energy terms [97] and this is used for the final ranking in FRODOCK. In FRODOCK 2.0 a new complementary coarse-grained knowledge-based protein-docking potential was introduced [98] and this two-step contact potential was incorporated because of the excellent trade-off between accuracy and low computational cost [99].

M-ZDOCK was developed for predicting the structure of cyclically symmetric multimers based on the construction of an unbound (or partially bound) monomer. Using a grid-based FFT approach purely symmetrical multimers are generated and searched for the highest quality structure rather than creating predicted structures with ZDOCK [88,89] and filtering for adjacent symmetrical structures. Fewer false positives are considered in the search, therefore, reducing the number of overlooked high-quality structures and the amount of computing power required is reduced by searching four instead of six degrees of freedom. Testing known multimer complexes from the PDB, M-ZDOCK was able to find at least one high-quality structure, whereas only two of the four test cases did when using ZDOCK and a filter for symmetry, and the running times are 30–40% faster for M-ZDOCK [100]. M-ZDOCK demonstrates how limiting focus can improve results, which calls for an integrative approach using many methods to cover different eventualities.

The ClusPro [101] docking algorithm evaluates multiple presumed complexes, retaining a pre-set number with encouraging surface complementarities, next a filtering method is applied to this set of structures, selecting those with good electrostatic and DE free energies for further clustering [102]. ClusPro-DC is a recent development of the ClusPro method which discriminates between crystallographic and biological dimers by docking the two subunits to sample the interaction energy landscape exhaustively [101,103]. The ClusPro server also utilizes SAXS data as constraints for selecting possible predicted structures [104] and has demonstrated an improvement in accuracy whilst doing so.

PatchDock [105] and LZerD [106] both utilize non FFT based methods for PPI prediction and could elucidate relevant predicted complexes that FFT based solutions may be missing due to the vagaries of mathematics. The PatchDock algorithm was inspired by object recognition and image segmentation techniques, where two molecules have their surfaces divided into patches based on the surface shape [105]. The surface of the proteins is calculated, a segmentation algorithm for detection of geometric patches (concave, convex and flat surface pieces) is applied and the patches are filtered, so that only patches with residues involved in binding are retained. A surface patch matching procedure applies geometric hashing and pose clustering matching techniques to match the patches previously detected, concave patches are paired with convex and flat patches with any patches. The predicted complexes of the previous stage are examined; all complexes with unacceptable clashes between receptor atoms and ligand atoms are discarded. Finally, the remaining candidates are ranked according to a geometric shape complementarity score [107].

The LZerD [106] software suite provides both pairwise dockings with LZerD and asymmetric multimeric docking with Multi-LZerD. LZerD has demonstrated improved performance since its introduction to the CAPRI experiment due to its continued integration of template based modelling, docking and scoring functions [108,109]. Multimeric docking programs have limited themselves to symmetrical complexes, which makes Multi-LZerD particularly useful as it provides complex asymmetric predictions. This is achieved using pairwise docking predictions from LZerD utilising the 3DZD a rotational invariant mathematical surface representation of proteins [106]. These are then combined using a genetic algorithm and several scoring methods are used, including clashing of atoms determined by atoms being with 3.0 Å of each other in each subunit. Furthermore, a physics-based scoring system that incorporates multiple scoring terms, where repulsive and attractive parts of the term are considered separately; an electrostatics term, which considers repulsive/attractive and short-range/long-range contributions individually; a hydrogen and disulphide bond term; two solvation terms; and a knowledge-based atom contact term [109].

Table 2 summarises the various bioinformatics methods for PPI modelling and how to access them.

Table 2. Overview of bioinformatics methods for modelling PPIs.

Name	Method	URL	Reference
RosettaDock	RosettaDock is a Monte Carlo (MC) based multi-scale docking algorithm.	https://www.rosettacommons.org/	[87]
ZDOCK	FFT used to perform a 3D search of the spatial degrees of freedom between two molecules., utilizes a pairwise statistical potential in the scoring function.	http://zdock.umassmed.edu/	[110]
GRAMM-X	The best surface match between molecules is determined by correlation technique using FFT, uses a smoothed Lennard-Jones potential on a fine grid during the global search FFT stage.	http://vakser.compbio.ku.edu/resources/gramm/	[94]
HexServer	Uses a closed-form 6D spherical polar FFT correlation expression from which arbitrary multi-dimensional multi-property multi-resolution FFT correlations may be generated.	http://hexserver.loria.fr/	[95]
MEGADOCK	MEGADOCK uses a Katchalski-Katzir algorithm and searches probable docking structures in a grid-based 3D space using FFT. MEGADOCK employs a scoring function in which only shape complementarity and electrostatics are considered. The method is set up to perform massive numbers of calculations that are run on parallel computing systems.	http://www.bi.cs.titech.ac.jp/megadock/	[96]
FRODOCK	FRODOCK projects the interaction terms of a potential protein complex into 3D grid-based potentials using spherical harmonics approximations to accelerate the search; this is itself an extension of the FFT alogrithm.	http://frodock.chaconlab.org/	[97]
M-ZDOCK	A grid-based FFT approach generates symmetrical multimers which are searched for the highest quality structure rather than creating predicted structures with ZDOCK and filtering for adjacent symmetrical structures.	http://zdock.umassmed.edu/m-zdock/	[100]
ClusPro	The ClusPro docking algorithm evaluates multiple presumed complexes, retaining a pre-set number with encouraging surface complementarities, next a filtering method is applied to this set of structures, selecting those with good electrostatic and DE free energies for further clustering.	https://cluspro.bu.edu/home.php	[101]
Patchdock	Two molecules have their surfaces divided into patches based on the surface shape. The surface of the proteins is calculated, a segmentation algorithm for detection of geometric patches is applied and the patches are filtered, so that only patches with residues involved in binding are retained. A surface patch matching procedure applies geometric hashing and pose clustering matching techniques to match the patches previously detected.	https://bioinfo3d.cs.tau.ac.il/PatchDock/	[105]
LZerD	Uses the 3DZD a rotational invariant mathematical surface representation of proteins to generate predictions.	http://www.kiharalab.org/proteindocking/lzerd.php	[111]
Multi-LZerD	Uses pairwise docking predictions from LZerD, these are then combined using a genetic algorithm and several scoring methods are used.	http://kiharalab.org/proteindocking/multilzerd.php	[106]

3.2. Quality Assessment of Quaternary Structure Predictions

Quality assessment is vital to structure prediction as it allows us to reduce a potentially considerable number of predicted complexes to a smaller subset for further ranking or examination. Energy based scoring of protein complexes is often carried out by docking programs to filter predictions and takes into account several binding terms: van der Waals energy and shape complementarity; desolvation energy and hydrophobicity; electrostatic interaction energy; translational, rotational and vibrational free energy changes. The assumption made by energy scoring is that proteins will find the lowest energy configuration when forming complexes. However, this is not always true, as some complexes observed by experimental methods have not taken up the lowest possible energy pose. Furthermore, these energy functions provide relative accuracy estimates, with only moderate power in adequately ranking models. Further, when one tries to compare models obtained from different methods, their associated energy scores are often not directly comparable. Therefore, accurate quality estimation methods that are independent of energy scoring are essential for protein structure prediction tools to fulfil their potential as useful techniques for biologists [90].

Consensus clustering can be used to determine the quality of predicted protein complexes. Docking predictions for pairs or groups of protein chains are clustered based on a scoring system that compares the similarity of the predictions using either root mean square deviation (RMSD) of the distance between the atoms of the complex in angstroms or the TM-score [112] of determining the similarity between multimeric protein complexes. Two complexes are considered to be neighbors if they are closer than a threshold angstrom (Å) value and are then placed into a cluster. Once the clusters are created, a prediction is selected out of each cluster as a representative prediction, using shape or physics-based scoring methods, and then the other predictions in the clusters are deleted. Through repeated passes of this process, the total number of predicted poses can be reduced to a more manageable number and processed manually to determine the highest quality predicted pose.

Although many different forms of tertiary model evaluation exist; such as those implemented in ModFOLD6 [113], ProQ3 [114,115] and others [116], recent progression onto quaternary protein modelling means that a reliable generic quaternary structure model quality assessment method has yet to be developed. The ZDOCK method uses an energy-based ranking system that is available separately from the ZDOCK prediction software [90]. ZRANK has also been integrated into MEGADOCK for ranking of predicted structures [96,117]. Similarly, upon evaluation of protein modelling in CASP11 [118], final quaternary model evaluation was recognized as an important but neglected area in need of significant development. Towards the end of the CASP12 experimental cycle, a new category for assessing the quality of the interface in protein–protein interactions was proposed. A preliminary scoring system, ModFOLD IA, for evaluating MultiFOLD models was trialed by the McGuffin group, following the submission of their top submitted complexes for CASP12. The ModFOLD IA method generates interface accuracy scores (IAs) for individual interface residues in MultiFOLD quaternary models. This is achieved first by identifying interface residues (heavy atoms within <5 Å of each other on different protein chains), followed by calculation of the minimum possible distance between residues in each interface pair (Dmin), a comparison of the mean Dmin between each interface pair across alternative models for each protein target (MeanDmin), and finally the generation of an IA score based on the difference between Dmin and MeanDmin per interface residue within each alternative protein model. This simple score aimed to provide a basis for interface residue scoring across generic alternative oligomer models by comparison of the interfaces in individual models to that of the mean interface value for a given target.

4. CASP, CAPRI and CAMEO: Driving the Development of in Silico Quaternary Structure Prediction Methods and Their Fusion with In Vitro Data

The development of methods for the prediction of protein-ligand binding sites and function prediction has been supported in recent years as a direct result of community-wide prediction experiments, such as CASP [7], CAPRI [8] and CAMEO [9].

4.1. CASP

CASP1, the first large-scale experiment to assess protein structure prediction methods, was held in 1994. It consisted of three parts: the collection of targets for prediction from the experimental community, the group of forecasts from the modelling community, and the assessment and discussion of the results. Information was solicited from X-ray crystallographers and NMR spectroscopists on structures about to be solved. Protein–protein interaction prediction was not added to CASP until CASP11 in 2014, however, in the years since, the area has proliferated and was a much more significant part of CASP12 in 2016. CASP has a large number of target sequences for prediction and runs over several months, culminating in a conference to review the experiment.

In Figure 2 computationally predicted and experimentally observed structures for a target protein from the CASP11 preliminary round are shown, highlighting some of the challenges of making the step from tertiary to quaternary structure prediction. The experimentally observed structure was provided by CASP, the target multimer is a dimer of protein Rab11b observed by X-ray spectroscopy and 198 residues in length, and the predicted structures were generated by the McGuffin group using a simple homology modelling procedure. Figure 2A shows the superposition generated by TM-align [112] of the observed and predicted monomers, which have been truncated by removing the disordered region from the predicted structure and the alpha helix from the observed structure. This superposition has a TM-score of 0.93365, which demonstrates the high quality of the initial tertiary structure prediction for this target sequence. The superposition of the complex shown in Figure 2B generated a TM-score of 0.42037 when aligned using MM-align [119]. In Figure 2D the predicted structure shows a disordered region that was predicted to form the interaction site for the dimer. In Figure 2C, the observed structure shows that the disordered regions formed a stable alpha helix upon dimerization. The dimer superposition in Figure 2B has a much lower TM-score than the truncated monomer superposition due to the disordered region becoming ordered and an orientation change about the disordered region, which has not allowed the predicted and observed monomers to align as well as they can be aligned when truncated.
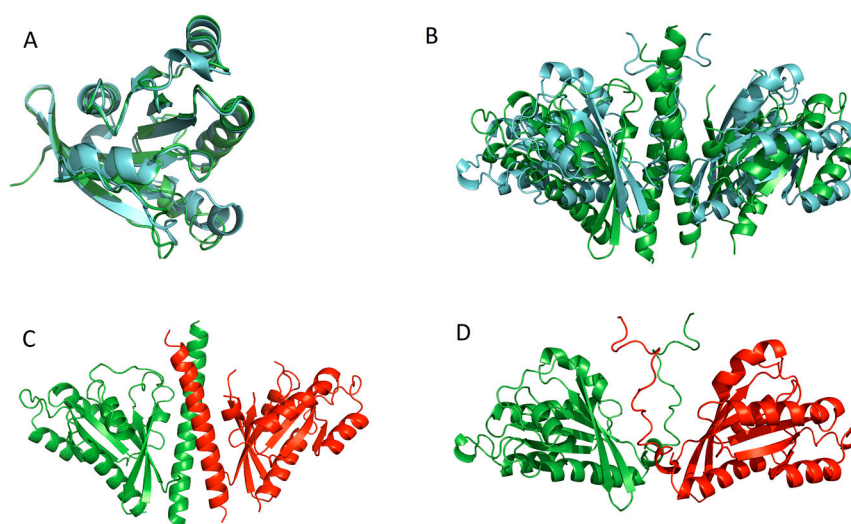

Figure 2. The challenge of making the transition from tertiary to quaternary structure prediction—a homodimer prediction. Even with a high quality starting tertiary structure other factors, such as disorder-order transitions, may come into play. (**A**) Superposition of Truncated Predicted and Truncated Observed Monomers of T0798 target Sequence with the observed monomer colored blue and the predicted monomer colored green; (**B**) Superposition of Predicted and Observed Dimer of T0798 target Sequence with the observed dimer coloured green and the predicted dimer colored blue, aligned with PyMOL; (**C**) Observed Dimer of T0798 Target Sequence with each of the two monomers coloured green and red respectively; (**D**) Predicted Dimer of T0798 Target Sequence with each of the two monomers coloured green and red respectively.

In the CASP 12 experiment the McGuffin group predicted tertiary structures using IntFOLD [120–123] and other server models, then subsequently the highest scoring submitted model was further refined using the ReFOLD [124] procedure. The ReFOLD refinement pipeline consists of three main protocols. The first and second protocol used i3Drefine [125] and NAMD [126] for structure refinement of the starting model. The generated refined models were then assessed and ranked using the new and improved ModFOLD6 [113,127,128]. From this, only top 5 refined models were selected and submitted for the CASP12 experiment, and the top model was used for MultiFOLD analysis. MultiFOLD is an as yet unpublished meta-server method for determining protein quaternary structure that utilises the LZerD, MEGADOCK, FRODOCK, PatchDock and ZDOCK for dimeric complexes (two protein subunits) and M-ZDOCK and Multi-LZerD for multimeric complexes (three or more protein subunits). The predicted quaternary structures were then ranked for submission using a variety of methods including each programs internal ranking procedures, other ranking software and manual intervention using information provided by CASP concerning the origin of the protein. In Figure 3 computationally predicted and experimentally observed structures for a target protein (T0868/T0869 heterodimer) from the CASP 12 preliminary round are shown. The experimentally observed structure, Figure 3A, was provided by CASP and the predicted structure, Figure 3B was generated by the McGuffin group using the MultiFOLD method. A superposition of the observed and predicted structures for the dimer is shown in Figure 3C and has a TM-score of 0.33088 indicating a low similarity between the two structures.
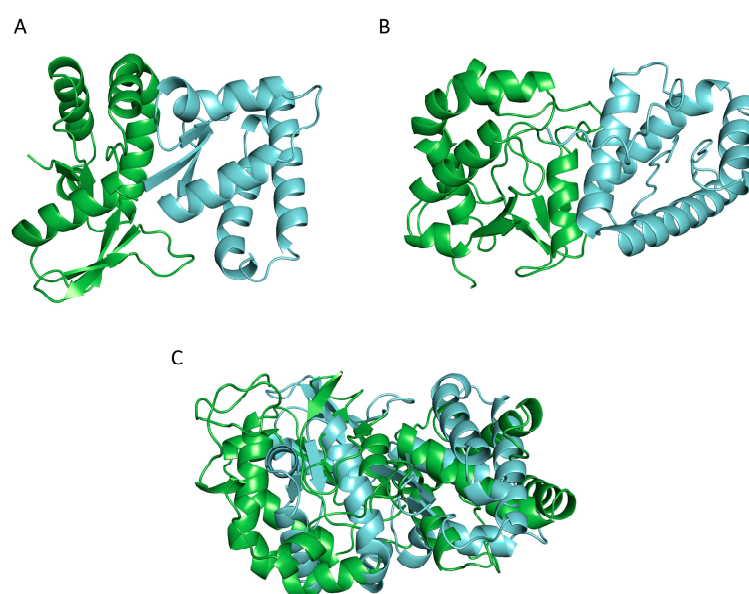

Figure 3. The challenge of making the transition from tertiary to quaternary structure prediction—a hetero dimer prediction. The success of modelling a complex is reliant on the quality of the starting tertiary structure model. (**A**) Observed structure of T0868/T0869 dimer; (**B**) Predicted structure of T0868/T0869 dimer from poor quality initial tertiary structures; (**C**) Superposition of observed and predicted T0868/T0869 dimer with the observed dimer coloured green and the predicted dimer colored blue, aligned with PyMOL.

The ultimate goal of structure prediction is to provide insights into biological functions. However, it is difficult to quantify and benchmark the utility of protein structure prediction for functional inference [129]. The biological function of a protein may have several different meanings; it can include catalyzing chemical reactions, structural support. Most transporting materials across the cell, receiving and sending chemical signals, or responding to stimuli and providing of these functions are realised by interacting with other proteins or small molecules. Therefore, interfaces between proteins, or interfaces between a protein and small molecules are critical to understanding function. Official CASP structural

assessments include global and local metrics that evaluate the atomic level similarity of the structural features of proteins [130–132]. The root mean square deviation (RMSD) was the first metric used in the CASP evaluations, and it is still reported in the automatic evaluation system. The global distance test (GDT) score is useful for the automatic evaluation of predictions as it reflects the absolute and relative accuracy of models for a wide range of target difficulty. In addition to GDT, several other similarity measures are used. Structural quality often tracks with functional quality, but the details of this correlation need to be further explored.

During CASP11, it was seen that the model accuracy was improved when using NMR simulated sparse data. Dramatic improvement was seen in GDT_TS score (Global Distance Test used to quantify prediction performance) for all data assisted models, where GDT_TS was better by >17 GDT_TS units on average compared with the CNS (crystallography and NMR system) results (those models unassisted by data) [133].

4.2. CAMEO

CAMEO continuously applies quality assessment criteria established by the protein structure prediction community. Since the accuracy requirements for different scientific applications vary, there is no "one fits all" score. CAMEO, therefore, offers a variety of scores—assessing different aspects of a prediction (coverage, historical accuracy and completeness) to reflect these requirements [9]. CAMEO operates on a continuous automated basis and is therefore only suitable for server-based predictive methods. The Contact Prediction (CP) section of the CAMEO experiment has been running for 56 weeks with four groups registered as predictors, and 68 target sequences have been tested.

4.3. CAPRI

CAPRI is a blind prediction experiment, and its targets are unpublished crystal or NMR structures and are submitted to CAPRI by their authors in a confidential process to maintain the validity of the experiment. Contributor predictor groups are given the atomic coordinates of two proteins that are known to have relevant biological interactions. The target interactions are modelled with the help of the coordinates and other publicly available data and then submit sets of ten models for assessment. Also, the predictors are invited to upload more massive sets that are communicated to scorer groups who evaluate and rank them and make a separate ten-model submission. After the prediction round is finalised, the CAPRI assessors equate the submissions to the experimental structure and evaluate the models on criteria that depend on the geometry and biological importance of the predicted interactions [8]. CAPRI has four prediction rounds each year with only a few targets per round but is growing with each round.

5. Conclusions

A large number of high throughput experimental techniques and computational programmes are now available for both PPI detection and modelling quaternary structures. Integrating this information together with sparse experimental data, or hybrid modelling, will help biologists to elucidate the complex networks of PPIs, which are intrinsic to every cellular process. A plethora of methods are being explored for in silico quaternary structure prediction and the refinement and quality assessment of such predictions. In docking, methods based on the FFT algorithms are currently prevalent although the field is developing alternative methods. Energy scoring is common to the majority of methods as the main ranking criteria and clustering techniques are also sometimes utilized. However, there is a clear need for the development of generic quality assessment methods for evaluating predicted quaternary structures, if models are to be more seriously adopted by the wider biology community. The combination of predictive and experimental methods and the continued focus of the CASP, CAPRI and CAMEO on quaternary structures, will be essential for driving our continued progress.

Abbreviations

NMR	Nuclear magnetic resonance
SAXS	Small-angle X-ray scattering
PPI	Protein–protein interaction
TAP	Tandem affinity purification
Co-IP	Co-immunoprecipitation
CD	Circular Dichroism
ECs	Evolutionary couplings
GPU	Graphics processing unit
NOE	Nuclear Overhauser effect
RDC	Residual dipolar coupling
Cryo-EM	Cryo-electron microscopy
PDB	Protein Data Bank
RMSD	Root mean square deviation
CASP	Critical Assessment of techniques for Structure Prediction
CAPRI	Critical assessment of prediction of interactions
CAMEO	Continuous automated model evaluation
LDDT	Local distance difference test on all atoms
3DZD	3D Zernike descriptor
FFT	Fast Fourier transform

References

1. Levinthal's Paradox. Available online: http://web.archive.org/web/20110523080407/http://www-miller.ch.cam.ac.uk/levinthal/levinthal.html (accessed on 25 May 2015).
2. RCSB PDB—Holdings Report. Available online: https://www.rcsb.org/pdb/statistics/holdings.do (accessed on 22 November 2017).
3. Emwas, A.-H.M. The strengths and weaknesses of NMR spectroscopy and mass spectrometry with particular focus on metabolomics research. *Methods Mol. Biol.* **2015**, *1277*, 161–193. [CrossRef] [PubMed]
4. Kikhney, A.G.; Svergun, D.I. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett.* **2015**, *589*, 2570–2577. [CrossRef] [PubMed]
5. Szklarczyk, D.; Jensen, L.J. *Protein-Protein Interaction Databases*; Meyerkord, C.L., Fu, H., Eds.; Methods in Molecular Biology; Springer: New York, NY, USA, 2015; ISBN 978-1-4939-2424-0.
6. Ehrenberger, T.; Cantley, L.C.; Yaffe, M.B. *Computational Prediction of Protein-Protein Interaction*; Meyerkord, C.L., Fu, H., Eds.; Springer: New York, NY, USA, 2015; ISBN 978-1-4939-2424-0.
7. Moult, J.; Fidelis, K.; Kryshtafovych, A.; Schwede, T.; Tramontano, A. Critical Assessment of Methods of Protein Structure Prediction (CASP)—Round XII. *Proteins Struct. Funct. Bioinform.* **2017**. [CrossRef] [PubMed]
8. Janin, J. Welcome to CAPRI: A critical assessment of PRedicted interactions. *Proteins Struct. Funct. Genet.* **2002**, *47*, 257–257. [CrossRef]
9. Haas, J.; Roth, S.; Arnold, K.; Kiefer, F.; Schmidt, T.; Bordoli, L.; Schwede, T. The protein model portal—A comprehensive resource for protein structure and model information. *Database* **2013**, *2013*. [CrossRef] [PubMed]
10. IntAct. Available online: https://www.ebi.ac.uk/intact/ (accessed on 4 October 2017).
11. Günzl, A.; Schimanski, B. Tandem Affinity Purification of Proteins. In *Current Protocols in Protein Science*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2001; ISBN 978-0-471-14086-3.

12. Puig, O.; Caspary, F.; Rigaut, G.; Rutz, B.; Bouveret, E.; Bragado-Nilsson, E.; Wilm, M.; Seraphin, B. The tandem affinity purification (TAP) method: A general procedure of protein complex purification. *Methods* **2001**, *24*, 218–229. [CrossRef] [PubMed]

13. Bauch, A.; Superti-Furga, G. Charting protein complexes, signaling pathways, and networks in the immune system. *Immunol. Rev.* **2006**, *210*, 187–207. [CrossRef] [PubMed]

14. Gavin, A.-C.; Bosche, M.; Krause, R.; Grandi, P.; Marzioch, M.; Bauer, A.; Schultz, J.; Rick, J.M.; Michon, A.-M.; Cruciat, C.-M.; et al. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **2002**, *415*, 141–147. [CrossRef] [PubMed]

15. Lee, C. Coimmunoprecipitation Assay. In *Circadian Rhythms: Methods and Protocols*; Rosato, E., Ed.; Humana Press: Totowa, NJ, USA, 2007; pp. 401–406, ISBN 978-1-59745-257-1.

16. Ren, L.; Emery, D.; Kaboord, B.; Chang, E.; Qoronfleh, M.W. Improved immunomatrix methods to detect protein–protein interactions. *J. Biochem. Biophys. Methods* **2003**, *57*, 143–157. [CrossRef]

17. Phizicky, E.M.; Fields, S. Protein–protein interactions: Methods for detection and analysis. *Microbiol. Rev.* **1995**, *59*, 94–123. [PubMed]

18. Fields, S.; Song, O. A novel genetic system to detect protein–protein interactions. *Nature* **1989**, *340*, 245–246. [CrossRef] [PubMed]

19. Estojak, J.; Brent, R.; Golemis, E.A. Correlation of two-hybrid affinity data with in vitro measurements. *Mol. Cell. Biol.* **1995**, *15*, 5820–5829. [CrossRef] [PubMed]

20. Deane, C.M.; Salwinski, L.; Xenarios, I.; Eisenberg, D. Protein interactions: Two methods for assessment of the reliability of high throughput observations. *Mol. Cell. Proteom.* **2002**, *1*, 349–356. [CrossRef]

21. Semple, J.I.; Sanderson, C.M.; Campbell, R.D. The jury is out on "guilt by association" trials. *Brief. Funct. Genomic.* **2002**, *1*, 40–52. [CrossRef]

22. Louche, A.; Salcedo, S.P.; Bigot, S. Protein–protein interactions: Pull-down assays. *Methods Mol. Biol.* **2017**, *1615*, 247–255. [CrossRef] [PubMed]

23. Nguyen, T.N.; Goodrich, J.A. Protein–protein interaction assays: Eliminating false positive interactions. *Nat. Methods* **2006**, *3*, 135–139. [CrossRef] [PubMed]

24. Zhu, H.; Snyder, M. Protein chip technology. *Curr. Opin. Chem. Biol.* **2003**, *7*, 55–63. [CrossRef]

25. Zhu, H.; Klemic, J.F.; Chang, S.; Bertone, P.; Casamayor, A.; Klemic, K.G.; Smith, D.; Gerstein, M.; Reed, M.A.; Snyder, M. Analysis of yeast protein kinases using protein chips. *Nat. Genet.* **2000**, *26*, 283–289. [CrossRef] [PubMed]

26. Chen, Y.; Xu, D. Computational analyses of high-throughput protein–protein interaction data. *Curr. Protein Pept. Sci.* **2003**, *4*, 159–181. [CrossRef] [PubMed]

27. O'Connell, M.R.; Gamsjaeger, R.; Mackay, J.P. The structural analysis of protein–protein interactions by NMR spectroscopy. *Proteomics* **2009**, *9*, 5224–5232. [CrossRef] [PubMed]

28. Gao, G.; Williams, J.G.; Campbell, S.L. Protein–protein Interaction analysis by nuclear magnetic resonance spectroscopy. In *Protein–Protein Interactions: Methods and Applications*; Fu, H., Ed.; Humana Press: Totowa, NJ, USA, 2004; pp. 79–91, ISBN 978-1-59259-762-8.

29. Hermjakob, H.; Montecchi-Palazzi, L.; Bader, G.; Wojcik, J.; Salwinski, L.; Ceol, A.; Moore, S.; Orchard, S.; Sarkans, U.; von Mering, C.; et al. The HUPO PSI's molecular interaction format—A community standard for the representation of protein interaction data. *Nat. Biotechnol.* **2004**, *22*, 177–183. [CrossRef] [PubMed]

30. Narayan, P.; Orte, A.; Clarke, R.W.; Bolognesi, B.; Hook, S.; Ganzinger, K.A.; Meehan, S.; Wilson, M.R.; Dobson, C.M.; Klenerman, D. The extracellular chaperone clusterin sequesters oligomeric forms of the amyloid-β(1–40) peptide. *Nat. Struct. Mol. Biol.* **2011**, *19*, 79–83. [CrossRef] [PubMed]

31. Heegaard, N.H. Affinity in electrophoresis. *Electrophoresis* **2009**, *30*, S229–S239. [CrossRef] [PubMed]

32. Orchard, S.; Hermjakob, H.; Julian, R.K.J.; Runte, K.; Sherman, D.; Wojcik, J.; Zhu, W.; Apweiler, R. Common interchange standards for proteomics data: Public availability of tools and schema. *Proteomics* **2004**, *4*, 490–491. [CrossRef] [PubMed]

33. Manzano, C.; Contreras-Martel, C.; El Mortaji, L.; Izore, T.; Fenel, D.; Vernet, T.; Schoehn, G.; di Guilmi, A.M.; Dessen, A. Sortase-mediated pilus fiber biogenesis in *Streptococcus pneumoniae*. *Structure* **2008**, *16*, 1838–1848. [CrossRef] [PubMed]

34. Rogers, K.R. Principles of affinity-based biosensors. *Mol. Biotechnol.* **2000**, *14*, 109–129. [CrossRef]

35. Wallace, B.A.; Janes, R.W. Synchrotron radiation circular dichroism spectroscopy of proteins: Secondary structure, fold recognition and structural genomics. *Curr. Opin. Chem. Biol.* **2001**, *5*, 567–571. [CrossRef]

36. Jelesarov, I.; Bosshard, H.R. Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition. *J. Mol. Recognit.* **1999**, *12*, 3–18. [CrossRef]

37. Murata, K.; Mitsuoka, K.; Hirai, T.; Walz, T.; Agre, P.; Heymann, J.B.; Engel, A.; Fujiyoshi, Y. Structural determinants of water permeation through aquaporin-1. *Nature* **2000**, *407*, 599–605. [CrossRef] [PubMed]

38. Honke, K.; Kotani, N. The enzyme-mediated activation of radical source reaction: A new approach to identify partners of a given molecule in membrane microdomains. *J. Neurochem.* **2011**, *116*, 690–695. [CrossRef] [PubMed]

39. Van Liempd, S.; Morrison, D.; Sysmans, L.; Nelis, P.; Mortishire-Smith, R. Development and validation of a higher-throughput equilibrium dialysis assay for plasma protein binding. *J. Lab. Autom.* **2011**, *16*, 56–67. [CrossRef] [PubMed]

40. Muchowski, P.J.; Schaffar, G.; Sittler, A.; Wanker, E.E.; Hayer-Hartl, M.K.; Hartl, F.U. Hsp70 and hsp40 chaperones can inhibit self-assembly of polyglutamine proteins into amyloid-like fibrils. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 7841–7846. [CrossRef] [PubMed]

41. Demirdoven, N.; Cheatum, C.M.; Chung, H.S.; Khalil, M.; Knoester, J.; Tokmakoff, A. Two-dimensional infrared spectroscopy of antiparallel β-sheet secondary structure. *J. Am. Chem. Soc.* **2004**, *126*, 7981–7990. [CrossRef] [PubMed]

42. Prakasam, A.K.; Maruthamuthu, V.; Leckband, D.E. Similarities between heterophilic and homophilic cadherin adhesion. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 15434–15439. [CrossRef] [PubMed]

43. Leavitt, S.; Freire, E. Direct measurement of protein binding energetics by isothermal titration calorimetry. *Curr. Opin. Struct. Biol.* **2001**, *11*, 560–566. [CrossRef]

44. Murphy, R.M. Static and dynamic light scattering of biological macromolecules: What can we learn? *Curr. Opin. Biotechnol.* **1997**, *8*, 25–30. [CrossRef]

45. Badr, C.E. Bioluminescence imaging: Basics and practical limitations. *Methods Mol. Biol.* **2014**, *1098*, 1–18. [CrossRef] [PubMed]

46. Duhr, S.; Braun, D. Why molecules move along a temperature gradient. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 19678–19682. [CrossRef] [PubMed]

47. Chatake, T.; Tanaka, I.; Umino, H.; Arai, S.; Niimura, N. The hydration structure of a Z-DNA hexameric duplex determined by a neutron diffraction technique. *Acta Crystallogr. D Biol. Crystallogr.* **2005**, *61*, 1088–1098. [CrossRef] [PubMed]

48. Hanson, B.L. Getting protein solvent structures down cold. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 16393–16394. [CrossRef] [PubMed]

49. Pellecchia, M.; Sem, D.S.; Wuthrich, K. NMR in drug discovery. *Nat. Rev. Drug Discov.* **2002**, *1*, 211–219. [CrossRef] [PubMed]

50. Rammensee, S.; Slotta, U.; Scheibel, T.; Bausch, A.R. Assembly mechanism of recombinant spider silk proteins. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 6590–6595. [CrossRef] [PubMed]

51. Udenfriend, S.; Gerber, L.D.; Brink, L.; Spector, S. Scintillation proximity radioimmunoassay utilizing 125I-labeled ligands. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 8672–8676. [CrossRef] [PubMed]

52. Kranz, J.K.; Clemente, J.C. Binding techniques to study the allosteric energy cycle. *Methods Mol. Biol.* **2012**, *796*, 3–17. [CrossRef] [PubMed]

53. Cotruvo, J.A.J.; Stubbe, J. NrdI, a flavodoxin involved in maintenance of the diferric-tyrosyl radical cofactor in *Escherichia coli* class Ib ribonucleotide reductase. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 14383–14388. [CrossRef] [PubMed]

54. Lowe, T.M.; Eddy, S.R. A computational screen for methylation guide snoRNAs in yeast. *Science* **1999**, *283*, 1168–1171. [CrossRef] [PubMed]

55. Modesti, M.; Ristic, D.; van der Heijden, T.; Dekker, C.; van Mameren, J.; Peterman, E.J.G.; Wuite, G.J.L.; Kanaar, R.; Wyman, C. Fluorescent human RAD51 reveals multiple nucleation sites and filament segments tightly associated along a single DNA molecule. *Structure* **2007**, *15*, 599–609. [CrossRef] [PubMed]

56. Unger, V.M. Electron cryomicroscopy methods. *Curr. Opin. Struct. Biol.* **2001**, *11*, 548–554. [CrossRef]

57. Tanabe, Y.; Fujita, E.; Momoi, T. FOXP2 promotes the nuclear translocation of POT1, but FOXP2(R553H), mutation related to speech-language disorder, partially prevents it. *Biochem. Biophys. Res. Commun.* **2011**, *410*, 593–596. [CrossRef] [PubMed]

58. Denhardt, D.T. Mechanism of action of antisense RNA. Sometime inhibition of transcription, processing, transport, or translation. *Ann. N. Y. Acad. Sci.* **1992**, *660*, 70–76. [CrossRef] [PubMed]

59. Chiu, Y.-L.; Rana, T.M. RNAi in human cells: Basic structural and functional features of small interfering RNA. *Mol. Cell* **2002**, *10*, 549–561. [CrossRef]

60. Karimova, G.; Pidoux, J.; Ullmann, A.; Ladant, D. A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 5752–5756. [CrossRef] [PubMed]

61. Rossi, F.; Charlton, C.A.; Blau, H.M. Monitoring protein–protein interactions in intact eukaryotic cells by β-galactosidase complementation. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 8405–8410. [CrossRef] [PubMed]

62. Galarneau, A.; Primeau, M.; Trudeau, L.-E.; Michnick, S.W. β-lactamase protein fragment complementation assays as in vivo and in vitro sensors of protein protein interactions. *Nat. Biotechnol.* **2002**, *20*, 619–622. [CrossRef] [PubMed]

63. Hu, C.-D.; Chinenov, Y.; Kerppola, T.K. Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation. *Mol. Cell* **2002**, *9*, 789–798. [CrossRef]

64. Remy, I.; Michnick, S.W. Clonal selection and in vivo quantitation of protein interactions with protein-fragment complementation assays. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 5394–5399. [CrossRef] [PubMed]

65. Lemmens, I.; Eyckerman, S.; Zabeau, L.; Catteeuw, D.; Vertenten, E.; Verschueren, K.; Huylebroeck, D.; Vandekerckhove, J.; Tavernier, J. Heteromeric MAPPIT: A novel strategy to study modification-dependent protein–protein interactions in mammalian cells. *Nucleic Acids Res.* **2003**, *31*, e75. [CrossRef] [PubMed]

66. Stefan, E.; Aquin, S.; Berger, N.; Landry, C.R.; Nyfeler, B.; Bouvier, M.; Michnick, S.W. Quantification of dynamic protein complexes using *Renilla luciferase* fragment complementation applied to protein kinase A activities in vivo. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 16916–16921. [CrossRef] [PubMed]

67. Hubsman, M.; Yudkovsky, G.; Aronheim, A. A novel approach for the identification of protein–protein interaction with integral membrane proteins. *Nucleic Acids Res.* **2001**, *29*, E18. [CrossRef] [PubMed]

68. Kato, N.; Jones, J. The split luciferase complementation assay. *Methods Mol. Biol.* **2010**, *655*, 359–376. [CrossRef] [PubMed]

69. Russ, W.P.; Engelman, D.M. TOXCAT: A measure of transmembrane helix association in a biological membrane. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 863–868. [CrossRef] [PubMed]

70. Dyer, K.N.; Hammel, M.; Rambo, R.P.; Tsutakawa, S.E.; Rodic, I.; Classen, S.; Tainer, J.A.; Hura, G.L. High-throughput SAXS for the characterization of biomolecules in solution: A practical approach. *Methods Mol. Biol.* **2014**, *1091*, 245–258. [CrossRef] [PubMed]

71. Jiménez-García, B.; Pons, C.; Svergun, D.I.; Bernadó, P.; Fernández-Recio, J. pyDockSAXS: Protein–protein complex structure by SAXS and computational docking. *Nucleic Acids Res.* **2015**. [CrossRef] [PubMed]

72. Skou, S.; Gillilan, R.E.; Ando, N. Synchrotron-based small-angle X-ray scattering of proteins in solution. *Nat. Protoc.* **2014**, *9*, 1727–1739. [CrossRef] [PubMed]

73. Tang, Y.; Huang, Y.J.; Hopf, T.A.; Sander, C.; Marks, D.S.; Montelione, G.T. Protein structure determination by combining sparse NMR data with evolutionary couplings. *Nat. Meth.* **2015**, *12*, 751–754. [CrossRef] [PubMed]

74. Latek, D.; Ekonomiuk, D.; Kolinski, A. Protein structure prediction: Combining de novo modeling with sparse experimental data. *J. Comput. Chem.* **2007**, *28*, 1668–1676. [CrossRef] [PubMed]

75. Robinson, P.J.; Trnka, M.J.; Pellarin, R.; Greenberg, C.H.; Bushnell, D.A.; Davis, R.; Burlingame, A.L.; Sali, A.; Kornberg, R.D. Molecular architecture of the yeast Mediator complex. *eLife* **2015**, *4*. [CrossRef] [PubMed]

76. Tang, X.; Bruce, J.E. Chemical cross-linking for protein–protein interaction studies. In *Mass Spectrometry of Proteins and Peptides: Methods and Protocols*; Lipton, M.S., Paša-Tolic, L., Eds.; Humana Press: Totowa, NJ, USA, 2009; pp. 283–293, ISBN 978-1-59745-493-3.

77. Kluger, R.; Alagic, A. Chemical cross-linking and protein–protein interactions—A review with illustrative protocols. *Bioorg. Chem.* **2004**, *32*, 451–472. [CrossRef] [PubMed]

78. Nogales, E. The development of cryo-EM into a mainstream structural biology technique. *Nat. Methods* **2016**, *13*, 24–27. [CrossRef] [PubMed]

79. Bai, X.; McMullan, G.; Scheres, S.H. How cryo-EM is revolutionizing structural biology. *Trends Biochem. Sci.* **2015**, *40*, 49–57. [CrossRef] [PubMed]

80. Statistics: EMDataBank. Available online: http://www.emdatabank.org/statistics.html (accessed on 10 October 2017).

81. Skiniotis, G. A snapshot of cryo-EM. *Protein Sci.* **2017**, *26*, 5–7. [CrossRef] [PubMed]

82. Wang, H.-W.; Wang, J.-W. How cryo-electron microscopy and X-ray crystallography complement each other. *Protein Sci.* **2017**, *26*, 32–39. [CrossRef] [PubMed]

83. Wu, S.; Tan, D.; Woolford, J.L.; Dong, M.-Q.; Gao, N. Atomic modeling of the ITS2 ribosome assembly subcomplex from cryo-EM together with mass spectrometry-identified protein–protein crosslinks. *Protein Sci.* **2017**, *26*, 103–112. [CrossRef] [PubMed]

84. Gadkari, R.A.; Srinivasan, N. Prediction of protein–protein interactions in dengue virus coat proteins guided by low resolution cryoEM structures. *BMC Struct. Biol.* **2010**, *10*, 17. [CrossRef] [PubMed]

85. Gadkari, R.A.; Varughese, D.; Srinivasan, N. Recognition of interaction interface residues in low-resolution structures of protein assemblies solely from the positions of Cα atoms. *PLoS ONE* **2009**, *4*. [CrossRef] [PubMed]

86. Bernstein, F.C.; Koetzle, T.F.; Williams, G.J.; Meyer, E.F.; Brice, M.D.; Rodgers, J.R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The protein data bank. *FEBS J.* **1977**, *80*, 319–324.

87. Gray, J.J.; Moughon, S.; Wang, C.; Schueler-Furman, O.; Kuhlman, B.; Rohl, C.A.; Baker, D. Protein–protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J. Mol. Biol.* **2003**, *331*, 281–299. [CrossRef]

88. Chen, R.; Li, L.; Weng, Z. ZDOCK: An initial-stage protein-docking algorithm. *Proteins Struct. Funct. Bioinform.* **2003**, *52*, 80–87. [CrossRef] [PubMed]

89. Pierce, B.G.; Wiehe, K.; Hwang, H.; Kim, B.-H.; Vreven, T.; Weng, Z. ZDOCK server: Interactive docking prediction of protein–protein complexes and symmetric multimers. *Bioinformatics* **2014**, *30*, 1771–1773. [CrossRef] [PubMed]

90. Pierce, B.; Weng, Z. ZRANK: Reranking protein docking predictions with an optimized energy function. *Proteins Struct. Funct. Bioinform.* **2007**, *67*, 1078–1086. [CrossRef] [PubMed]

91. Wang, C.; Schueler-Furman, O.; Andre, I.; London, N.; Fleishman, S.J.; Bradley, P.; Qian, B.; Baker, D. RosettaDock in CAPRI rounds 6–12. *Proteins Struct. Funct. Bioinform.* **2007**, *69*, 758–763. [CrossRef] [PubMed]

92. Sønderby, P.; Rinnan, Å.; Madsen, J.J.; Harris, P.; Bukrinski, J.T.; Peters, G.H.J. Small-angle X-ray scattering data in combination with RosettaDock improves the docking energy landscape. *J. Chem. Inf. Model.* **2017**, *57*, 2463–2475. [CrossRef] [PubMed]

93. Katchalski-Katzir, E.; Shariv, I.; Eisenstein, M.; Friesem, A.A.; Aflalo, C.; Vakser, I.A. Molecular surface recognition: Determination of geometric fit between proteins and their ligands by correlation techniques. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 2195–2199. [CrossRef] [PubMed]

94. Tovchigrechko, A.; Vakser, I.A. GRAMM-X public web server for protein–protein docking. *Nucleic Acids Res.* **2006**, *34*, W310–W314. [CrossRef] [PubMed]

95. Macindoe, G.; Mavridis, L.; Venkatraman, V.; Devignes, M.-D.; Ritchie, D.W. HexServer: An FFT-based protein docking server powered by graphics processors. *Nucleic Acids Res.* **2010**, *38*, W445–W449. [CrossRef] [PubMed]

96. Ohue, M.; Shimoda, T.; Suzuki, S.; Matsuzaki, Y.; Ishida, T.; Akiyama, Y. MEGADOCK 4.0: An ultra–high-performance protein–protein docking software for heterogeneous supercomputers. *Bioinformatics* **2014**, *30*, 3281–3283. [CrossRef] [PubMed]

97. Garzon, J.I.; Lopéz-Blanco, J.R.; Pons, C.; Kovacs, J.; Abagyan, R.; Fernandez-Recio, J.; Chacon, P. FRODOCK: A new approach for fast rotational protein–protein docking. *Bioinformatics* **2009**, *25*, 2544–2551. [CrossRef] [PubMed]

98. Tobi, D. Designing coarse grained-and atom based-potentials for protein–protein docking. *BMC Struct. Biol.* **2010**, *10*, 40. [CrossRef] [PubMed]

99. Moal, I.H.; Torchala, M.; Bates, P.A.; Fernández-Recio, J. The scoring of poses in protein–protein docking: current capabilities and future directions. *BMC Bioinform.* **2013**, *14*, 286. [CrossRef] [PubMed]

100. Pierce, B.; Tong, W.; Weng, Z. M-ZDOCK: A grid-based approach for Cn symmetric multimer docking. *Bioinformatics* **2005**, *21*, 1472–1478. [CrossRef] [PubMed]

101. Kozakov, D.; Hall, D.R.; Xia, B.; Porter, K.A.; Padhorny, D.; Yueh, C.; Beglov, D.; Vajda, S. The ClusPro web server for protein–protein docking. *Nat. Protoc.* **2017**, *12*, 255–278. [CrossRef] [PubMed]

102. Comeau, S.R.; Gatchell, D.W.; Vajda, S.; Camacho, C.J. ClusPro: An automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics* **2004**, *20*, 45–50. [CrossRef] [PubMed]

103. Yueh, C.; Hall, D.R.; Xia, B.; Padhorny, D.; Kozakov, D.; Vajda, S. ClusPro-DC: Dimer classification by the CLUSPRO server for protein–protein docking. *J. Mol. Biol.* **2017**, *429*, 372–381. [CrossRef] [PubMed]

104. Xia, B.; Mamonov, A.; Leysen, S.; Allen, K.N.; Strelkov, S.V.; Paschalidis, I.C.; Vajda, S.; Kozakov, D. Accounting for observed small angle X-ray scattering profile in the protein–protein docking server cluspro. *J. Comput. Chem.* **2015**, *36*, 1568–1572. [CrossRef] [PubMed]

105. Duhovny, D.; Nussinov, R.; Wolfson, H.J. Efficient unbound docking of rigid molecules. In *International Workshop on Algorithms in Bioinformatics*; Springer: New York, NY, USA, 2002; pp. 185–200.

106. Esquivel-Rodríguez, J.; Yang, Y.D.; Kihara, D. Multi-LZerD: Multiple protein docking for asymmetric complexes. *Proteins Struct. Funct. Bioinform.* **2012**. [CrossRef] [PubMed]

107. Schneidman-Duhovny, D.; Inbar, Y.; Polak, V.; Shatsky, M.; Halperin, I.; Benyamini, H.; Barzilai, A.; Dror, O.; Haspel, N.; Nussinov, R. Taking geometry to its edge: Fast unbound rigid (and hinge-bent) docking. *Proteins Struct. Funct. Bioinform.* **2003**, *52*, 107–112. [CrossRef] [PubMed]

108. Peterson, L.X.; Kim, H.; Esquivel-Rodriguez, J.; Roy, A.; Han, X.; Shin, W.-H.; Zhang, J.; Terashi, G.; Lee, M.; Kihara, D. Human and server docking prediction for CAPRI round 30–35 using LZerD with combined scoring functions: Scoring LZerD CAPRI Docking Predictions. *Proteins Struct. Funct. Bioinform.* **2017**, *85*, 513–527. [CrossRef] [PubMed]

109. Peterson, L.X.; Shin, W.-H.; Kim, H.; Kihara, D. Improved performance in CAPRI round 37 using LZerD docking and template-based modeling with combined scoring functions. *Proteins Struct. Funct. Bioinform.* **2017**. [CrossRef] [PubMed]

110. Pierce, B.G.; Hourai, Y.; Weng, Z. Accelerating protein docking in ZDOCK using an advanced 3D convolution library. *PLoS ONE* **2011**, *6*, e24657. [CrossRef] [PubMed]

111. Venkatraman, V.; Yang, Y.D.; Sael, L.; Kihara, D. Protein–protein docking using region-based 3D Zernike descriptors. *BMC Bioinform.* **2009**, *10*, 407. [CrossRef] [PubMed]

112. Zhang, Y. TM-align: A protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **2005**, *33*, 2302–2309. [CrossRef] [PubMed]

113. Maghrabi, A.H.A.; McGuffin, L.J. ModFOLD6: An accurate web server for the global and local quality estimation of 3D protein models. *Nucleic Acids Res.* **2017**, *45*, W416–W421. [CrossRef] [PubMed]

114. Uziela, K.; Shu, N.; Wallner, B.; Elofsson, A. ProQ3: Improved model quality assessments using Rosetta energy terms. *Sci. Rep.* **2016**, *6*. [CrossRef] [PubMed]

115. Uziela, K.; Hurtado, D.M.; Shu, N.; Wallner, B.; Elofsson, A. ProQ3D: Improved model quality assessments using deep learning. *Bioinformatics* **2017**, *33*, 1578–1580. [CrossRef] [PubMed]

116. Elofsson, A.; Joo, K.; Keasar, C.; Lee, J.; Maghrabi, A.H.A.; Manavalan, B.; McGuffin, L.J.; Ménendez Hurtado, D.; Mirabello, C.; Pilstål, R.; et al. Methods for estimation of model accuracy in CASP12. *Proteins Struct. Funct. Bioinform.* **2017**. [CrossRef] [PubMed]

117. Shimoda, T.; Ishida, T.; Suzuki, S.; Ohue, M.; Akiyama, Y. MEGADOCK-GPU: Acceleration of—Docking Calculation on GPUs. In *Proceedings of the International Conference on Bioinformatics, Computational Biology and Biomedical Informatics*; BCB'13; ACM: New York, NY, USA, 2013; pp. 883:883–883:889.

118. Lensink, M.F.; Velankar, S.; Kryshtafovych, A.; Huang, S.-Y.; Schneidman-Duhovny, D.; Sali, A.; Segura, J.; Fernandez-Fuentes, N.; Viswanath, S.; Elber, R.; et al. Prediction of homo- and hetero-protein complexes by protein docking and template-based modeling: A CASP-CAPRI experiment. *Proteins Struct. Funct. Bioinform.* **2016**. [CrossRef] [PubMed]

119. Mukherjee, S.; Zhang, Y. MM-align: A quick algorithm for aligning multiple-chain protein complex structures using iterative dynamic programming. *Nucleic Acids Res.* **2009**, *37*, e83. [CrossRef] [PubMed]

120. McGuffin, L.J.; Atkins, J.D.; Salehe, B.R.; Shuid, A.N.; Roche, D.B. IntFOLD: An integrated server for modelling protein structures and functions from amino acid sequences. *Nucleic Acids Res.* **2015**. [CrossRef] [PubMed]

121. McGuffin, L.J.; Roche, D.B. Automated tertiary structure prediction with accurate local model quality assessment using the intfold-ts method. *Proteins Struct. Funct. Bioinform.* **2011**, *79*, 137–146. [CrossRef] [PubMed]

122. Roche, D.B.; Buenavista, M.T.; Tetchner, S.J.; McGuffin, L.J. The IntFOLD server: An integrated web resource for protein fold recognition, 3D model quality assessment, intrinsic disorder prediction, domain prediction and ligand binding site prediction. *Nucleic Acids Res.* **2011**, *39*, W171–W176. [CrossRef] [PubMed]

123. McGuffin, L.J.; Shuid, A.N.; Kempster, R.; Maghrabi, A.H.A.; Nealon, J.O.; Salehe, B.R.; Atkins, J.D.; Roche, D.B. Accurate template-based modeling in CASP12 using the IntFOLD4-TS, ModFOLD6, and ReFOLD methods. *Proteins Struct. Funct. Bioinform.* **2017**. [CrossRef] [PubMed]

124. Shuid, A.N.; Kempster, R.; McGuffin, L.J. ReFOLD: A server for the refinement of 3D protein models guided by accurate quality estimates. *Nucleic Acids Res.* **2017**, *45*, W422–W428. [CrossRef] [PubMed]

125. Bhattacharya, D.; Cheng, J. i3Drefine software for protein 3D structure refinement and its assessment in CASP10. *PLoS ONE* **2013**, *8*, e69648. [CrossRef] [PubMed]

126. Kumar, S.; Huang, C.; Zheng, G.; Bohm, E.; Bhatele, A.; Phillips, J.C.; Yu, H.; Kale, L.V. Scalable molecular dynamics with NAMD on the IBM Blue Gene/L system. *IBM J. Res. Dev.* **2008**, *52*, 177–188. [CrossRef]

127. Roche, D.B.; Buenavista, M.T.; McGuffin, L.J. Assessing the quality of modelled 3D protein structures using the ModFOLD server. *Methods Mol. Biol.* **2014**, *1137*, 83–103. [CrossRef] [PubMed]

128. McGuffin, L.J.; Buenavista, M.T.; Roche, D.B. The ModFOLD4 server for the quality assessment of 3D protein models. *Nucleic Acids Res.* **2013**, *41*, W368–W372. [CrossRef] [PubMed]

129. Huwe, P.J.; Xu, Q.; Shapovalov, M.V.; Modi, V.; Andrake, M.D.; Dunbrack, R.L. Biological function derived from predicted structures in CASP11: CASP11 biological function prediction. *Proteins Struct. Funct. Bioinform.* **2016**, *84*, 370–391. [CrossRef] [PubMed]

130. Kryshtafovych, A.; Monastyrskyy, B.; Fidelis, K. CASP11 statistics and the prediction center evaluation system: Prediction center in CASP11. *Proteins Struct. Funct. Bioinform.* **2016**, *84*, 15–19. [CrossRef] [PubMed]

131. Kryshtafovych, A.; Moult, J.; Baslé, A.; Burgin, A.; Craig, T.K.; Edwards, R.A.; Fass, D.; Hartmann, M.D.; Korycinski, M.; Lewis, R.J.; et al. Some of the most interesting CASP11 targets through the eyes of their authors: CASP11 target highlights. *Proteins Struct. Funct. Bioinform.* **2016**, *84*, 34–50. [CrossRef] [PubMed]

132. Kryshtafovych, A.; Moult, J.; Bales, P.; Bazan, J.F.; Biasini, M.; Burgin, A.; Chen, C.; Cochran, F.V.; Craig, T.K.; Das, R.; et al. Challenging the state of the art in protein structure prediction: Highlights of experimental target structures for the 10th Critical Assessment of Techniques for Protein Structure Prediction Experiment CASP10: CASP10 Target Highlights. *Proteins Struct. Funct. Bioinform.* **2014**, *82*, 26–42. [CrossRef] [PubMed]

133. Moult, J.; Fidelis, K.; Kryshtafovych, A.; Schwede, T.; Tramontano, A. Critical assessment of methods of protein structure prediction: Progress and new directions in round XI. *Proteins* **2016**, *84*, 4–14. [CrossRef] [PubMed]