Review

Modeling and Structure Determination of Homo-Oligomeric Proteins: An Overview of Challenges and Current Approaches

Aljaž Gaber *,† [ID] and Miha Pavšič *,† [ID]

Department of Chemistry and Biochemistry, Faculty of Chemistry and Chemical Technology,
University of Ljubljana, SI-1000 Ljubljana, Slovenia
* Correspondence: aljaz.gaber@fkkt.uni-lj.si (A.G.); miha.pavsic@fkkt.uni-lj.si (M.P.)
† These authors contributed equally to this work.

Abstract: Protein homo-oligomerization is a very common phenomenon, and approximately half of proteins form homo-oligomeric assemblies composed of identical subunits. The vast majority of such assemblies possess internal symmetry which can be either exploited to help or poses challenges during structure determination. Moreover, aspects of symmetry are critical in the modeling of protein homo-oligomers either by docking or by homology-based approaches. Here, we first provide a brief overview of the nature of protein homo-oligomerization. Next, we describe how the symmetry of homo-oligomers is addressed by crystallographic and non-crystallographic symmetry operations, and how biologically relevant intermolecular interactions can be deciphered from the ordered array of molecules within protein crystals. Additionally, we describe the most important aspects of protein homo-oligomerization in structure determination by NMR. Finally, we give an overview of approaches aimed at modeling homo-oligomers using computational methods that specifically address their internal symmetry and allow the incorporation of other experimental data as spatial restraints to achieve higher model reliability.

Keywords: structure determination; modeling; homo-oligomers

1. Introduction

Many proteins have a natural tendency to self-associate into homo-oligomeric protein complexes, also termed homomers, which are composed of two or more identical subunits. According to the estimation, 30–50% of all proteins oligomerize [1]. In addition, analysis of protein crystal structures demonstrated that roughly 45% of eukaryotic proteins and 60% of prokaryotic proteins that are deposited as single polypeptide chains also exist in a form of homo-oligomeric complex [2]. Adding the specifics of structural characterization of homo-oligomers, for example, symmetry and challenges associated with distinguishing between intra- and inter-molecular inter-residue contacts, the topic is of great importance to any structural biologist. We aim to provide an overview of the topic, first by giving background information on the nature of homo-oligomerization and continuing by describing experimental/computational approaches for homo-oligomer structure determination and modeling.

1.1. Protein Homo-Oligomerization as an Efficient Design Principle

Homo-oligomerization is believed to be nature's solution to form large proteins by avoiding efficiency problems with the synthesis of long polypeptide chains. In comparison to small proteins, larger ones are more favorable due to their higher stability and smaller solvent-exposed surface percentage. Moreover, building larger protein complexes from smaller subunits has several benefits. Such complexes are less prone to translational errors, as only the defected subunits can be discarded and replaced in contrast to the whole large single polypeptide-chain protein. Next, coding efficiency is higher because less information needs to be stored to build a large protein. Furthermore, assembly of (homo)-oligomeric

proteins can be triggered and/or fine-tuned and thus provides an additional layer of regulation, which is crucial in dynamic processes such as actin filament assembly [3] and microtubule growth [4]. Other examples are protein activation as a consequence of dimerization, as in the case of caspase-9 activity [5] and signaling via epidermal growth factor receptor (EGFR) [6]. Additionally, an opposite effect can be achieved—for example, dimerization inhibits the activity of receptor-like protein tyrosine phosphatase-a [7]. Even more prominent examples are where the active site is formed at the interface between subunits, as in the case of HIV-1 dimerization protease. [8]. Besides, oligomerization also enables homotropic allosteric interactions between subunits, for example, in membrane protein αβ TCR [9] and L-Lactate dehydrogenase [10]. Such allosteric regulation was found to be the most common in oligomers with dihedral symmetries, especially in metabolic enzymes [11]. Yet another example are the death domains of several proteins involved in cell death and immune cell signaling where dimerization often leads to protein activation. Here, dimerization is if often mediated via domain swapping where two subunits exchange their parts to form an intertwined dimer [12]. The same principle can also apply to higher-order homo-oligomers, an example is the barnase domain-swapped trimer [13], and is also frequently associated with formation of protein aggregates/deposits [14].

As these advantages are almost intuitive, it is often assumed they should also provide a clear evolutionary benefit. However, Lynch suggested that homo-oligomers could have arisen from stochastic, non-adaptive processes [15,16] and that the benefits of homo-oligomerization are not all-pervasive, but rather dependent on the context and the properties of the individual protein [17]. These and other possible reasons as to why homo-oligomerization is such a frequently encountered property were also extensively discussed elsewhere [1,18–21].

1.2. Most Protein Homo-Oligomers Are Symmetric

Symmetry is an inherent property of almost all homo-oligomers characterized up to date. Although homo-oligomers often exhibit at least some degree of local asymmetry [18], this asymmetry is limited to small differences in the backbone position, differences in sidechain orientations or limited to a certain part of the protein, while the complex as a whole is still symmetric (Figure 1A). Local asymmetry may provide an insight into the mechanism of complex formation [22]. For example, in the case of domain swapped oligomers, local asymmetry may reveal the location of the hinge regions that connect the swapped portions of the subunits [23], as in the case of pancreatic ribonuclease where N-terminal regions are exchanged (Figure 1A). For a comprehensive review on structural asymmetry in homo-dimers, the reader is referred to the paper written by Swapna, Srikeerthana and Srinivasan [24].

On the other hand, global asymmetry (Figure 1B) is rare. In the dataset of annotated biological assemblies (QSbio) [25], less than 5% of homo-oligomeric structures do not have symmetry (Figure 2A, Supplementary Table S1). The most observed symmetry types are cyclic symmetries (C_n in Schönflies notation) with a single axis of rotation, and dihedral symmetries (D_n) with at least one additional axis of rotation, perpendicular to the first one (Figure 2B). Cubic symmetry with the tetrahedral, octahedral and icosahedral arrangement is much less common; together, they account for roughly 1% of all structures. Icosahedral symmetry is often observed in viral envelopes, but there, the envelope is usually composed of several different polypeptide chains [26]. Interestingly, symmetries with an odd number of subunits are less common than those with an even number of subunits. This can be explained by the nature of interactions between subunits, which can be isologous or heterologous. Isologous interactions take place between identical surfaces and amino-acid residues on the interacting subunits, while, in heterologous interactions, different regions on juxtaposed subunits are involved (Figure 2C). Several studies have shown that isologous interactions are more favorable than heterologous [27–30], thus explaining the higher number of oligomers with an even number of subunits [1]. While C_2 symmetric dimers are by far the most populous oligomeric state, dihedral symmetries prevail among

homo-oligomers with a higher number of subunits. These can also be explained by the advantages of isologous interactions over heterologous, as interactions in cyclic homo-oligomers with more than two subunits are, by definition, heterologous. Symmetry is also related to a finite control of protein assembly by producing a closed set of subunits. On the contrary, aggregation of proteins through the non-finite assembly is related to several pathological conditions [31].
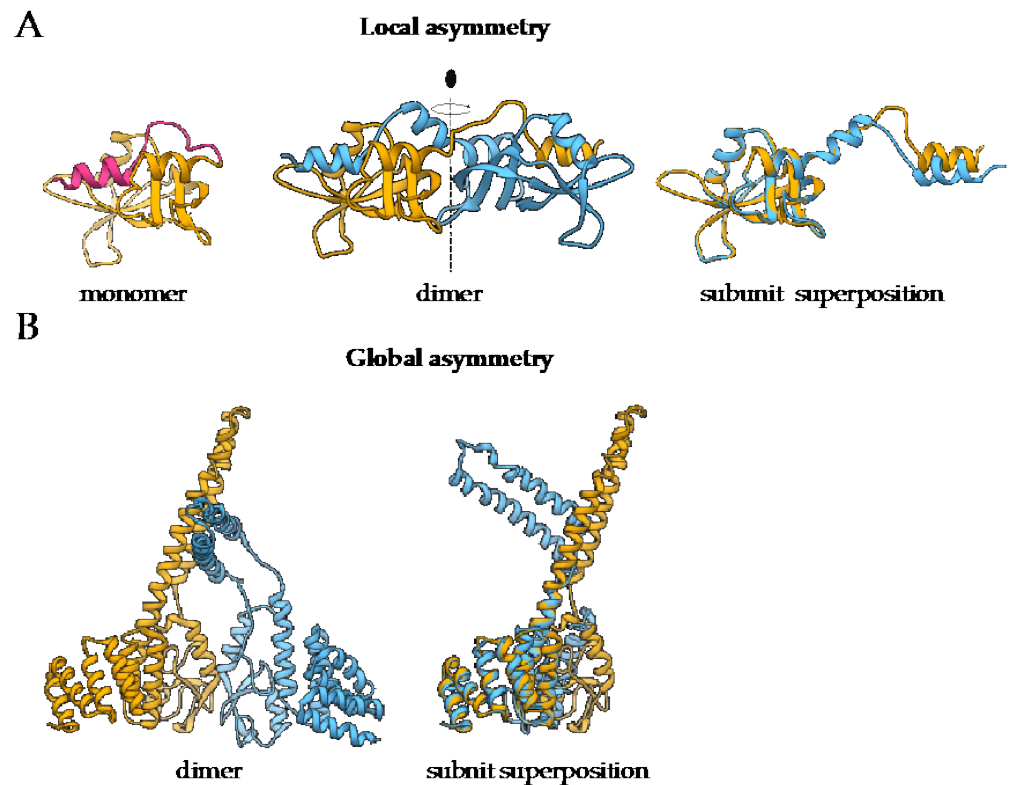

Figure 1. The difference between local and global asymmetry. (**A**) An example of a globally symmetric homo-oligomeric complex with a substantial local asymmetry is the structure of bovine pancreatic ribonuclease N-term-swapped dimer (PDB: 1A2W) [24,32]. During the dimer formation, the N-terminal region of the monomer (PDB: 1A5P, pink) [33] is swapped the juxtaposed subunit. (**B**) Murine CHIP-U-box E3 ubiquitin ligase (PDB: 2C2L) [34] is an example of a globally asymmetric homo-oligomeric complex. Individual subunits are depicted in yellow and blue. For both complexes, the superposition of polypeptide chains is also presented to demonstrate the extent of structural differences between the subunits.
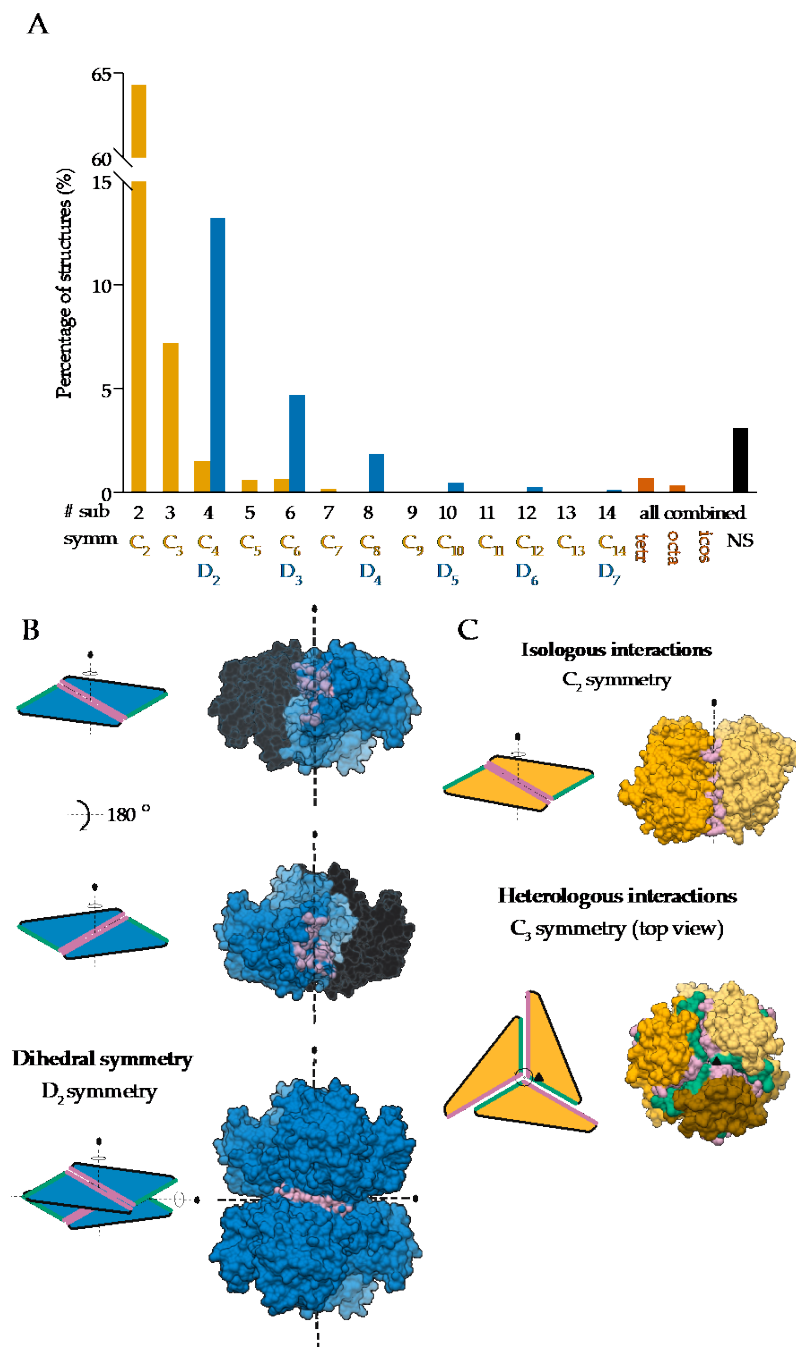
Figure 2. Symmetries observed in the determined structures of homo-oligomeric protein complexes: (**A**) Relative distribution of symmetry types in QSbio with a 90% sequence similarity cutoff. Complexes were classified by the number of subunits (# sub) and the symmetry type (symm). Data for tetrahedral (tetr), octahedral (octa), icosahedral (icos) and non-symmetric (NS) complexes are combined for all number of subunits. (**B**) Assembly of a dihedral symmetry from parts of protein complexes with cyclic symmetry, glucose-6-phosphate 1-dehydrogenase (PDB: 6D23) [35]. In the representation of protein surfaces of the C_2 symmetric subcomplexes, one of the subunits is transparent to enable visualization of the interaction surface. (**C**) Representation of the isologous and heterologous interactions in complexes with odd and even number of subunits in the case of orange carotenoid-binding protein (PDB: 5UI2) [36] and corrinoid adenosyltransferase (PDB: 2R6T) [37], respectively. Individual subunits are depicted with different shades of yellow. Distinct interaction surfaces are depicted with green and pink, respectively. Two- and three-fold axes are denoted by corresponding symbols.

2. High-Resolution Structure Determination of Homo-Oligomers

Experimental structure determination is still the principal and most reliable method of high-resolution structural characterization of proteins. The structural models are deposited in the publicly available Protein Data Bank where the oligomeric state is annotated in terms of stoichiometry and symmetry, which can be either global (operating between complete chains or their assemblies), local (limited to a part of the molecule) or helical [38]. Symmetry is an important factor in structure determination and interpretation process; however, how it is addressed differs between the different experimental approaches.

Currently, the main methods for high-resolution structure determination of proteins are X-ray crystallography and nuclear magnetic resonance (NMR), with cryo-electron microscopy (cryo-EM) joining the group, fueled by technological advances in recent years enabling sub-1.5 Å resolution [39]. Of these methods, X-ray crystallography is the one most inherently linked to symmetry operations because symmetry underlies both diffraction data collection and processing [40] as well as nearly all calculations needed to arrive at the final model of the structure [41]. Moreover, cryo-EM utilizes symmetry during image processing and averaging to achieve a higher signal-to-noise ratio [42]. Since symmetry is also a fundamental property of the vast majority of protein homo-oligomers (as already discussed above), the consideration of symmetry aspects during structure determination and interpretation is critical. Here, we provide an overview of the homo-oligomerization-associated aspects of the listed methods for protein structure determination with a focus on how symmetry can either help or pose significant challenges.

2.1. X-ray Crystallography

2.1.1. Characteristics of Protein Crystals

A critical requirement for X-ray structure determination is a crystal of the molecule/complex of interest. A crystal represents an ordered array of molecules, and the smallest unit from which the complete crystal can be re-created by application of translation and rotation is the asymmetric unit. Therefore, the crystal structure of a protein is determined and reported as the structure(s) of the protein molecule(s) and their ligands within one copy of the asymmetric unit. Next, symmetry operations of the crystal space group, determined already during initial diffraction data processing and later confirmed during phasing and structure refinement process, can be applied to generate neighboring asymmetric units, the unit cell and, by translation, the complete crystal. Application of these operations produces copies of protein molecules, which form a network of inter-molecular contacts stabilizing the crystal [43]. Typically, biological interfaces bury more than 400 Å^2 per subunit (800 Å^2 in total), while crystal contacts bury, on average, less than 400 Å^2 per subunit [1]. It can happen, however, that crystal contacts bury 400–1000 Å^2, and in some instances even more. Here, the main question with which a crystallographer is faced during structure analysis and interpretation is: Which, if any, of these inter-molecular contacts are stable also in solution and hence could be of biological relevance? Below, we provide an overview of the approaches aimed to address this question, first by considering the most important aspects of protein crystallization relevant for homo-oligomers.

Since, in protein crystals, solvent typically occupies approximately half of the volume of the crystal [44], protein molecules generally retain their solution-like structure and activity. However, there are examples where ordered packing into the crystal influenced protein conformation, mainly by preferential stabilization of one of the possible conformations. For example, a comparison of the structures of the same protein molecules determined by different groups or even structures with different local environments within the same crystal demonstrated the effect of packing on side-chain and backbone conformations as well as on hinge-like motions of protein molecules [45]. Furthermore, flash-freezing, a method commonly employed to stabilize crystals before diffraction data collection, can alter intra- and even more inter-molecular contacts within the crystal, particularly by cooling-introduced stabilization of long, polar side-chains of residues, which are then engaged in an extensive network of hydrogen bonds [46]. Even more, preferential packing of one equilibrium

oligomeric state vs. the other, for example, monomer vs. dimer, can complicate structure interpretation [47]. On the other hand, crowding effects exerted due to high protein concentration in the crystallization drop (typically 10 mg/mL or higher) and the presence of other reagent enhancing crowding, such as polyethylene glycol, can make low-affinity complexes more stable and thus make them more likely to be structurally characterized [48]. In addition, since symmetric proteins/protein assemblies tend to crystallize more readily, efforts were undertaken to trigger the formation of symmetric homo-oligomeric assemblies to broaden the crystallization bottle-neck [49,50]. Therefore, caution must be taken during the analysis of inter-molecular contacts and possible homo-oligomeric protein assemblies.

2.1.2. Symmetry Operations

Within the crystal, two types of symmetry operations are possible—crystallographic and non-crystallographic symmetry (NCS) operations. The crystallographic symmetry operations relate to neighboring asymmetric units, while the NCS operations work on molecules within the asymmetric unit. An extensive overview of crystal symmetry is given in the International Tables for Crystallography [51], particularly in section F, dedicated to biological macromolecules [52], with a more easily comprehensive compendium by Dauter and Jaskolski [53]. Due to packing and chirality-maintenance requirements, the crystallographic symmetry operations in protein crystal are limited to translations, rotations around 2-, 3-, 4- and 6-fold axes (rotations for 180°, 120°, 90° and 60°), plus their combinations in the form of screw axes. However, the NCS operations may also include rotations around other axes, for example, 5- and 7-fold axes, and operations not producing a closed set of copies—all of these have implications in the content of the asymmetric unit.

In case that the rotation axis/axes of symmetric homo-oligomeric complexes coincide with the crystallographic rotation axis/axes, the asymmetric unit will contain only part of the homo-oligomer, and the whole assembly can be re-created by application of symmetry operations of the crystal space group producing identical copies of the initial part. Indeed, analysis of a subset of structures in the Protein Data Bank showed that trimers, tetramers and hexamers preferentially crystallized in systems where the homo-oligomer symmetry was incorporated into crystal symmetry—in this case, the asymmetric unit does not contain the full oligomer [54]. However, the symmetry axis/axes of a homo-oligomer may correspond to NCS axis/axes. In this case, the asymmetric unit will contain one or more of the complete assemblies. Symmetric pentamers and heptamers, as an example, contain a rotational axis incompatible with crystal packing. The asymmetric unit will, in these cases, always contain one or more of such complete assemblies. There are also mixed cases where the complete homo-oligomeric assembly can be reconstituted by the application of crystal symmetry operations on partial assemblies containing NCS. The critical difference between crystallographic and non-crystallographic symmetry operations is that crystallographic symmetry operations always produce identical copies. However, the NCS operations may be improper and, as such, relate copies of molecules that are not perfect copies of each other. These copies differ in the conformation of one or more regions, for example, due to different inter-molecular contacts. One of the many examples is the crystal structure of human thyroid hormone receptor mutant with a dimer within the asymmetric unit. The subunits are related by a non-crystallographic 2-fold rotation axis, but they slightly differ with a root mean square deviation of 0.23 Å over C_α atoms [55].

2.1.3. Approaches to Distinguish between Crystal-Only and Biologically Relevant Interaction Interfaces

To determine which molecular assembly, either containing purely crystallographic symmetry, non-crystallographic symmetry or the combination of both, is biologically relevant, an inspection of all different inter-molecular contacts within the crystal is necessary. This is equally true for all protein complexes—for both hetero- as well as for symmetric and asymmetric homo-oligomers.

Various computational tools are available for this purpose and have already been extensively reviewed, for example, by Capitani and coworkers [56] and recently by Elez and coworkers [57]. Therefore, we provide here just a short overview of the approaches that can be classified into three main groups: (1) energy/thermodynamics-based methods; (2) empirical and comparative approaches based on evolutionary, B-factor, pair-atom distance and composite analysis; and (3) approaches incorporating machine learning. Due to different approaches, each tool uses a distinct set of parameters, which are generally distilled into a score signifying the relevance of the interaction. However, these scores are not directly comparable between different tools.

Of the thermodynamics-based methods, the most commonly used tool is PISA (Protein Interfaces, Surfaces, and Assemblies) [58,59]. PISA looks, by making copies of asymmetric unit contents through the application of crystallographic symmetry operations, at all possible inter-molecular interfaces both within and between asymmetric units. The interfaces are analyzed and described in terms of interface surface area, solvation free energy gain upon interface formation ($\Delta^i G$) and associated estimation of interface specificity (P-value) as the measure of the probability of obtaining lower $\Delta^i G$ from randomly picked atoms. Interface surface area is calculated as the difference between accessible surface area (ASA) of separated and complexed protein molecules, divided by the number of molecules in the proposed complex. ASA is commonly calculated by rolling a probe of the radius of 1.4 Å (corresponding to a water molecule) around the protein atoms with slightly increased radii to account for hydrogen atoms [60]. Additionally, the number of hydrogen bonds, disulfide bonds and salt bridges are reported. Interfaces with more negative $\Delta^i G$ (corresponding to hydrophobic interfaces), with higher P-value and with larger interface area are considered as mediating a stable oligomeric assembly. The significance of the observed interaction is reported as complexation significance score (CSS), which is the maximal fraction of the analyzed interface in terms of free energy of binding—higher value (up to 1) corresponds to higher significance. Another energy-based approach is that of ClusPro-DC [61], which underlies the ClusPro [62] docking algorithm mentioned below. The subunits of the proposed oligomer are taken apart and subjected to docking, and, if the docked poses form a close cluster resembling the initial oligomer, the interface is considered as biologically relevant. Contrary to PISA, this approach is limited to homo-dimers.

Empirical approaches work by analyzing specific features of each of the distinct inter-molecular interfaces within the crystal and classifying it as stable/biologically relevant in light of knowledge on already thoroughly analyzed interfaces. For example, EPPIC (evolutionary protein-protein interface classifier) relies on evolutionary analysis considering surface entropy of homologous sequences and few other parameters. However, for clear distinction, a high sequence similarity between query and homologs is needed [63]. The authors of ClusPro-DC tested three tools (PISA, EPPIC and ClusPro-DC) on the same dataset, and the accuracies were 59.6%, 78% and 74.5%, respectively, where accuracy is defined as the percentage of correct classification (true positive and true negatives) over the sum of correct and incorrect classifications [63]. Another approach is CFPScore (combinatorial four-parameter score), which incorporates estimates of binding free energy in inter-protein interactions, interface area, shape complementarity and packing density, reaching a reported prediction accuracy of 96.6% [64]. Other tools/approaches in this group are based on B-factor describing atomic vibrational motions [65], or interface area and complementarity, as in PreBI [66] and COMP [67].

The group of approaches incorporating machine learning (ML) is much broader. One of the newer tools is PIACO (protein interface analysis using covarying signals), which is based on covariance calculated from multiple sequence alignment and few other parameters such as amino acid composition and pair frequency [68]. The prediction accuracy was over 90% [67]. Another tool is PRODIGY-CRYSTAL (PROtein binDIng enerGY prediction) where the classification is based on inter-residue contacts and interaction energies and employs random forest (RF) ML [69,70]. Compared with EPPIC, which reached 88% accuracy, PRODIGY-CRYSTAL scored 92% prediction accuracy on the same

test dataset [71]. Next, RPAIAnalyst (residue pairs across interface) integrates the co-evolutionary aspect of residue pairs at the interface with other properties, such as secondary structure, B-factor and hydrophobicity/polarity, and again uses RF ML approach [72]. The tool reached 84.6% prediction accuracy on the same dataset as used for ClusPro-DC (above). These are just a few tools employing ML, and a more comprehensive list is available elsewhere [57].

To finally confirm the in-solution and also biological relevance of the proposed homo-oligomeric assembly, further experiments are needed. For example, insight into stoichiometry in the solution can be provided by size exclusion chromatography, static and dynamic laser light scattering experiments, analytical ultracentrifugation, mass spectrometry under native conditions and other methods. Further insight, also in terms of overall structural features, possibly of help in distinguishing between non-relevant and relevant assemblies, can be provided by small-angle X-ray scattering (SAXS). Examples are the crystal structures of human aldehyde dehydrogenase 7A1 [73] and of fungal UDP-galactopyranose mutase [74] where SAXS has been utilized to analyze various possible assemblies as interpreted from crystal packing contacts.

2.2. Nuclear Magnetic Resonance Spectroscopy

Nuclear magnetic resonance or NMR spectroscopy—shortly, NMR—is based on collecting various types of spectra, mainly 1D and 2D. Neighboring atoms, covalent bond lengths and other distances, dihedral angles and similar structure characteristics can be estimated from these spectra and are then used as restraints to calculate a convergent ensemble of structural models. Typically, around 20 models, representing different possible conformers, all satisfying the initial restraints, are obtained [75]. Contrary to X-ray crystallography, which is well suited for structural characterization of (very) large proteins and their assemblies, structure determination of proteins larger than approx. 35 kDa using NMR is still significantly challenging due to slower tumbling rates and shorter NMR signal relaxation times [76]. During NMR spectra recording proteins are generally in solution, which may more closely resemble their natural environment than the densely packed crystal. However, an equilibrium between monomeric and specific homo-oligomeric species—plus eventual non-specific associations—may pose significant challenges during structure determination due to spectral degeneracy and difficulties associated with distinguishing between intra- and intermolecular interactions [77]. On the other hand, symmetry in homo-oligomers can be useful since it results in simplified spectra because the complexity of spectra from large symmetric homo-oligomers is at the level of those of the monomer/subunit, especially in the case of cyclic symmetry [78].

An advancement in oligomer structure determination by NMR is represented by the approach using residual dipolar couplings, which provide domain orientation restraints [79], together with the nuclear Overhauser effect (NOE) due to dipole–dipole interactions and classical chemical shifts. These can be combined and used in modeling of monomer using CS-Rosetta, and of oligomer using Rosetta symmetric docking algorithm [77]. Another approach is to use a hybrid method, for example, by including overall shape information derived from SAXS data to guide assembly of monomer structure to homo-oligomers [80].

Interestingly, by using NMR, it has been shown that several homo-oligomers that were described as symmetric using X-ray crystallography display a certain degree of symmetry deviation, mainly due to hydrophilic amino acid residues at the subunit interface [22]. The authors of this recent analysis propose that averaging several conformations within the crystal result in a lower rate of observed symmetry deviations in crystal structures [22].

2.3. Cryo-Electron Microscopy

To determine protein structure, a special variant of cryo-EM is employed—the single particle analysis (SPA). Here, a high number of two-dimensional low-resolution images of the single macromolecule or an assembly—hence, single particle—in various orientations is combined to reconstruct its three-dimensional model [39]. Internal symmetry of the object of interest—as in the case of symmetrical oligomers or symmetrical arrangement of units composed of different chains—makes averaging possible and thus greatly contributes to higher signal-to-noise ratio and to a higher resolution of the final model. Historically, the high symmetry of large icosahedral viruses was employed to determine their structure. An early example from 2008 is the cryo-EM structure of the cytoplasmic polyhedrosis virus [81]. Here, averaging due to the icosahedral symmetrical arrangement of asymmetric units, although composed of several different chains, greatly contributed to the final resolution of 3.88 Å. The same detail-enhancing principle was employed during the determination of cryo-EM structures of other symmetrical assemblies, for example, high-resolution structures of oligomeric enzymes with cyclic, dihedral and tetrahedral symmetry [42]. However, similarly to improper NCS in crystal structures, also here subunits of the oligomer may have different conformations. In the case of higher-order oligomers, a high number of combinations with locally structurally different subunits is possible—this poses significant problems in particle classification [42].

In the process of cryo-EM structure determination, symmetry of macromolecular assemblies is generally detected during the initial analysis and classification of single particles. The most commonly used approach is based on a multivariate statistical analysis (MSA), which was initially used to process noisy images of randomly oriented biological macromolecules [82]. An improved approach based on the MSA is able to detect symmetry from the side- and tilted-view oriented particles, also from images of stoichiometrically non-uniform particles [83]. A recently reported novel approach detects symmetry using the charge density map after particle classification and 3D density map calculation without imposed symmetry. The method works by transforming the calculated density map using symmetry operations and then testing if the initial and transformed map coincide [84].

3. Computational Approaches for Modeling of Homo-Oligomers

High-resolution structure determination of protein homo-oligomers using X-ray crystallography or NMR is often time- and resource-consuming, and sometimes a structure of the monomer is available or can be modeled using homology approaches easily. In these cases, a structural model of homo-oligomer can be generated using computational approaches. For higher model reliability and relevance, additional data on the oligomerization interfaces obtained from other experiments can be used as spatial restraints. The success of modeling, especially of the docking approach, is critically dependent on the starting monomer structure—more success can be expected when their conformation closely resembles the one within the oligomer. Therefore, since due to computational complexity large conformational changes are inherently problematic to model, special care must be taken when selecting the input structure. This is even more critical when domain swapping is expected [85].

Here, we provide an overview of approaches and available computational tools, which are of special interest when modeling homo-oligomers (Table 1), first by considering symmetry-aware docking tools for constructing assemblies from monomeric structures (ab initio). Next, we continue with homology-based modeling approaches where the interface involved in oligomerization is translated from structures of homo-oligomers of homologous proteins.

Table 1. A summary of available software, designed for the modeling of symmetric homo-oligomeric protein complexes. When a webserver is available, only the number of subunits and the additional information, used to guide the modeling, that can be inputted into the webserver are summarized.

Software	Symmetry Types	Additional Information That Can Be Used to Guide the Modeling	Website	References
Ab Initio Docking of Protein Complexes with Cyclic Symmetries				
M-ZDOCK	C_{2-24}, user-defined		https://zdock.umassmed.edu/m-zdock/ (webserver) (accessed on 20 August 2021)	[86]
SymmDock	C_{2-100}, user-defined	interacting residues, distance restraints	http://bioinfo3d.cs.tau.ac.il/SymmDock/ (webserver) (accessed on 20 August 2021)	[87,88]
ClusPro	C_2 and C_3, user-defined	interacting and non-interacting residues, distance restraints (can be grouped), SAXS based restraints	https://cluspro.bu.edu/ (webserver) (accessed on 20 August 2021)	[62,89–91]
GRAMM-X	C_{2-8}, user-defined	interacting residues	http://vakser.compbio.ku.edu/resources/gramm/grammx/ (webserver) (accessed on 20 August 2021)	[92–94]
Ab initioDocking of Protein Complexes with Dihedral and Cubic Symmetries				
MOLFIT	cyclic and dihedral, user-defined		http://www.weizmann.ac.il/Chemical_Research_Support/molfit/home.html (accessed on 20 August 2021)	[95–97]
SAM	any, user-defined		http://sam.loria.fr/ (accessed on 20 August 2021)	[98]
HSYMDOCK	cyclic and dihedral, user-defined orpredicted	interacting residues	http://huanglab.phys.hust.edu.cn/hsymdock/ (webserver) (accessed on 20 August 2021)	[99–101]
GalaxyTongDock	cyclic and dihedral, (up to 12 subunits), user-defined	interacting and non-interacting residues	http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=TONGDOCK_INTRO (webserver) (accessed on 20 August 2021)	[102]
HADDOCK	cyclic and dihedral, up to 20 subunits and up to 10 segment pairs for each symmetry, user-defined	a variety of experimental restraints	https://wenmr.science.uu.nl/haddock2.4/ (webserver) (accessed on 20 August 2021)	[103–105]
Homology-Based Modeling of Homo-Oligomers				
Rosetta SymDock	cyclic, dihedral, icosahedral, helical, (only cyclic and dihedral with up to 10 subunits on the webserver), user-defined		https://rosie.graylab.jhu.edu/symmetric_docking/submit (webserver) (accessed on 20 August 2021)	[106,107]

Table 1. *Cont.*

Software	Symmetry Types	Additional Information That Can Be Used to Guide the Modeling	Website	References
Rosetta Fold-and-dock	cyclic, dihedral, icosahedral, helical, user-defined	a variety of experimental restraints	https://www.rosettacommons.org/ (accessed on 20 August 2021)	[108]
Rosetta SymDock2	cyclic, dihedral, icosahedral, helical, user-defined	a variety of experimental restraints	https://www.rosettacommons.org/ (accessed on 20 August 2021)	[109]
SWISS-MODEL	symmetry is inferred from the templates		https://swissmodel.expasy.org/ (webserver) (accessed on 20 August 2021)	[110–112]
GalaxyGemini	symmetry is inferred from the templates		http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=GEMINI (webserver) (accessed on 20 August 2021)	[113]
GalaxyHomomer	symmetry is inferred from the templates (user-defined), C_n can also be modeled		http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=HOMOMER (webserver) (accessed on 20 August 2021)	[114,115]
HDOCK	cyclic and dihedral, user-defined	the binding site, distance restraints, SAXS based restraints	http://hdock.phys.hust.edu.cn/ (webserver) (accessed on 20 August 2021)	[116]
ClusPro TBM	the user defines stoichiometry, not symmetry		https://tbm.cluspro.org/template_based/index.php (webserver) (accessed on 20 August 2021)	[117]
Rosetta CM	cyclic, dihedral, icosahedral, helical, user-defined	a variety of experimental restraints	https://www.rosettacommons.org/ (accessed on 20 August 2021)	[118,119]

3.1. Ab Initio Docking of Protein Complexes with Cyclic Symmetries

Cyclic symmetry is the most frequently encountered symmetry type in protein homo-oligomers, especially in its simplest form—the C_2 symmetry in symmetrical homo-dimers. There are several tools available where the user can impose such symmetry in docking of subunits—some allow modeling of only a small assembly such as dimer or trimer, while others are less restricted in the number of subunits.

M-ZDOCK [86] is an extension to ZDOCK [120], which uses a grid-based fast Fourier transform (FFT) approach to sampling. The online version allows docking of up to 24 subunits. In contrast to ZDOCK, sampling space is reduced to oligomers that are C_n symmetric. Imposing symmetry at the initial sampling instead of filtering the results at the end also leads to improvements in both the accuracy and computational time [86]. Although M-ZDOCK uses the ZDOCK scoring function [121], which does not provide the user with the ability to include experimentally determined restraints, integration of cross-linking mass spectrometry (XL-MS) data with Z-DOCK was recently reported to improve docking results and even provide insight into the symmetry of the analyzed protein complex [122].

A similar approach is also employed in SymmDock [87,88] which incorporates symmetry-based restraints to the PatchDock algorithm for pairwise docking, which is based on shape complementarity [123]. SymmDock can generate homo-oligomers with cyclic symmetry (up to 100 subunits), but an adapted version was also successfully used to generate models with dihedral D_2 symmetry [124]. To further improve modeling efficiency, users can provide external information on the binding site to restrict the sampling and distance restraints for scoring. The same group also developed the MultiFoXS [125] for fitting SAXS

data to multi-state models, which is especially useful in cases when multiple oligomeric states are present during data acquisition.

ClusPro, another software that uses FFT for sampling [62,90,91], also has an option to directly incorporate cyclic symmetry in docking [89]. However, docking is limited to dimers and trimers only. In contrast to M-ZDOCK and SymmDock, which rotate the subunit to generate symmetrical homo-oligomers, ClusPro rotates the coordinate system. Symmetry is enforced by only considering translations that are within 2 Å from the plane, defined by the symmetry axis of rotation [62]. Similar to SymmDock, distance restraints [126] and the binding site can be defined to narrow the sampling space. In addition to defining residues that participate in interactions (attraction), users can also provide those that are known to be located outside the interaction surface (repulsion). Distance restraints can be combined into groups of restraints and sets of groups. This feature is especially useful in the modeling of homo-oligomers, as distance restraints are often ambiguous (intra- and inter-subunit ambiguity) and/or symmetry related, as we have shown in the case of chemical-crosslinking-based restraints [127]. An option to filter final docking results according to their agreement with SAXS data [128,129] is also integrated.

Symmetry modeling of up to eight subunits can also be defined in GRAMM-X [92,93]. Symmetry is enforced by only considering models, provided by the discrete FFT grid search, that are symmetrical within a defined cutoff [94].

3.2. Ab Initio Docking of Protein Complexes with Dihedral and Cubic Symmetries

Although dihedral, tetrahedral, octahedral and icosahedral symmetric complexes can be in principle generated with additional transformations of protein complexes with cyclic symmetry, the software that incorporate this option in their ab initio docking workflow are rare. However, some were designed specifically for this purpose.

MolFit was the first algorithm to employ FFT to calculate the correlation function [96] and was adapted for the generation of D_2 protein complexes [95] by utilizing two approaches named ab/cd and ab/ac. The first applies translation and rotation to C_2 symmetric docking solutions, provided by MolFit, to assemble the tetramer. The second combines two different C_2 docking solutions (ab and ac), each representing one interaction surface between subunits in the tetramer. Later, the algorithm was extended to generate cyclic and dihedral symmetric complexes with a higher number of subunits [97].

Symmetry assembler (SAM) [98] employs spherical polar Fourier representations for sampling to rapidly assemble protein complexes with any closed symmetry, adopted by homo-oligomeric protein complexes. In all cases, (parts of) protein complexes with cyclic symmetry are first generated. Dihedral complexes are then assembled with additional translation and rotation. Similarly, tetrahedral, octahedral and icosahedral complexes are assembled from C_3 trimers.

HSYMDOCK [101] also enables the building of protein complexes with cyclical [100] or dihedral symmetries. An important feature of this software is that it includes an automatic prediction of the symmetry, without the user's input. Recently, [130], the software was expanded to include long-range interactions in the FFT-based search algorithm [131]. Dihedral symmetries are built with an approach similar to that of the ab/cd algorithm of MolFit and SAM, by an additional C_2 symmetric docking of a previously predicted C_n complex.

GalaxyTongDock [102], similar to M-ZDOCK described above, is also based on ZDOCK [120]. While M-ZDOCK is limited to cyclic symmetries, GalaxyTongDock also models dihedral symmetries with up to 12 subunits. Additionally, the user can provide a list of interacting and non-interacting residues to guide the docking.

Protein complexes with dihedral symmetry can also be modeled with HADDOCK [103,105], one of the most popular software for data-driven docking of biological macromolecules. The HADDOCK protocol consists of three stages. First, rigid body energy minimization is performed to identify docking poses, consistent with provided restraints. Second, there is a semi-flexible refinement in torsion angle space of all residues within the certain ra-

dius from the other molecule (5 Å by default). Third, a final explicit solvent refinement is performed. Various types of data can be used to guide the docking [132], including NMR-based restraints (residual dipolar couplings [133] relaxation anisotropy [134], pseudo contact shifts [135], interface predictions [136], and the radius of gyration obtained from experiments such as SAXS [137] and cryo-EM [138,139].

Multimer docking was introduced to HADDOCK in 2010 [104]. Currently, up to 20 subunits can be submitted to the HADDOCK webserver. The user can impose symmetry by defining C_2 pairs, C_3 triplets, C_4 quadruplets and/or C_5 quintuplets of subunits. By combining these options, complexes with dihedral symmetry can also be assembled. Symmetry is imposed at every stage of the HADDOCK protocol by requiring the intermolecular distances between symmetric C_α to be the same.

A protocol for symmetry docking with Rosetta SymDock, another very popular software for molecular docking, was demonstrated to be successful for modeling cyclic, dihedral, helical and icosahedral complexes [106]. It uses a real space Monte-Carlo-plus-minimization protocol, which is composed of two stages: a fast, low-resolution stage followed by atomic-scale optimization. Symmetry is imposed at both stages [106]—once the first subunit is introduced, the others are generated by symmetry transformations. If the homo-oligomer is composed of more than three subunits, only three adjacent ones are used, with energy only calculated for the central one, to improve the performance. In the second stage, any transformation to the sidechain-atoms of the first subunit is also mimicked in the other subunits. This approach was later expanded to the so-called "Fold-and-Dock" protocol [108] starting with extended polypeptide chains and simultaneously folding the subunits while docking them together, which is especially useful for modeling interleaved homo-oligomers. Recently, an updated version of SymDock—SymDock2—was released [109]. SymDock2 improves docking performance by using a more advanced scoring scheme called Motif Dock Score in the first, low-resolution stage of modeling, and including backbone flexibility in the second stage.

The accuracy of modeling predictions can be further improved by including restraints based on experimental data, such as NMR [140], cross-linking [141], SAXS [142] and sequence co-evolution data [143].

3.3. Homology-Based Modeling of Homo-Oligomers

The increasing number of solved 3D structures of protein complexes, deposited to PDB, enabled the development of homology-based algorithms for homo-oligomer structure prediction.

The first homology-based modeling software that was available to users as an automated pipeline for protein structure prediction was SWISS-MODEL [110–112]. Although it was initially limited to predictions of individual proteins, it was later expanded to model oligomeric structures [144]. The performance of oligomeric structure modeling was later improved by identifying protein–protein interactions with a conservation score that calculates the ratio between the interface and surface residue entropy from multiple sequence alignments of homologous proteins. This builds upon the assumption that residues at the interaction surface are more conserved compared with other residues on the protein surface [145]. Additionally, two features were introduced to improve the quality of the final predictions [145]: (1) quaternary structure score (QS-score), which quantifies the similarity between interfaces as a function of shared interfacial contacts, and (2) supervised machine learning approach, support vector machines (SVM), to predict the expected model-target QS-score.

One of the first software, designed specifically for template-based homo-oligomeric modeling, was GalaxyGemini, developed by the Chaok Seok group [113]. GalaxyGemini runs an HHsearch [146] on the homo-oligomer database, with the user inputted subunit structure, to extract templates based on sequence and tertiary/quaternary structure similarity. Homo-oligomeric models are then generated by superimposing the subunit structure

on the subunits of the homo-oligomer template with TM-align [147] and rigid-body energy minimization to remove steric clashes.

The same group later also developed another software homo-oligomer structure prediction—GalaxyHomomer [114,115], which combines template-based modeling, utilizing GalaxyTBM [148], as well as ab initio modeling, if less than 5 templates with sufficient homology are available. The oligomeric state can be specified by the user or inferred from homology-based templates. When less than 5 templates are available, the oligomeric state is predicted by ab initio modeling using GalaxyTongDock [102] algorithm; however, only C_n symmetries are considered. The key advantages of GalaxyHomomer over its predecessor GalaxyGemini are the final terminal region and loop-remodeling, using GalaxyLoop GalaxyLoop [149–151] while considering the symmetry and overall structure relaxation by GalaxyRefineComplex [152].

Symmetric docking of cyclic and dihedral complexes was also introduced to the HDOCK webserver [116]. HDOCK also employs a hybrid approach, combining template-based and ab initio docking. Template-based modeling is used if a suitable template of the complex is available; otherwise, ab initio docking is performed with HDOCK algorithm for multimer docking, while restricting the search space to abide by the symmetry-imposed constraints. In contrast to HSYMDOCK, developed by the same group, HDOCK also enables users to define the binding site and distance restraints or provide SAXS data to improve the docking result.

Recently, ClusPro template-based modeling (TBM) webserver was launched [117]. However, the functionality of ClusPro TBM still has some limitations in comparison to ClusPro. Users are not able to provide any experimental data to guide the docking, and symmetry cannot be defined. Homo-oligomeric models are generated by searching potential templates that agree with the user-defined stoichiometry and then copying the subunit to the matching positions, followed by interface optimization with fixed backbones.

Symmetry constraints can also be used in Rosetta's comparative modeling protocol, Rosetta CM [119]. Target sequences are modeled onto template backbone, followed by fragment-based modeling of gaps and all-atom optimization. Only templates with the target symmetry are considered [153]. Symmetry is also imposed in the second and the third step of the protocol [118].

3.4. Other Computational Approaches, Used for Modeling Homo-Oligomers

Although not designed for handling symmetric complexes, several docking algorithms and servers, not described above, were successfully employed for docking homo-oligomeric complexes in the recent CASP-CAPRI experiments [94,154]: SWARMDOCK [155], PPI3D [156,157] and LzerD [158,159] MDOCKPP [160]. However, it needs to be noted that they were often combined with the algorithms for symmetry docking to impose symmetry.

To tackle homo-oligomer modeling on a proteome scale, the ProtCHOIR tool was recently developed and applied to the *Mycobacterium abscessus* proteome as a proof-of-concept as part of the *Mabellini* project [161].

The development of deep learning methods led to a dramatic improvement in structure prediction of uncomplexed proteins, especially by the AlphaFold2 algorithm [162]. On the other hand, comparable improvement was not observed in the modeling of protein complexes, although the AlphaFold2 authors claim it often provides good predictions for homo-oligomers, even those with intertwined chains. Given the previous successes of the Rosetta protocol, the recent release of the RoseTTAfold [163], which is also available as a webserver https://robetta.bakerlab.org (accessed on 20 August 2021), also holds great promise.

4. Conclusions

Identification of protein homo-oligomers, and other assemblies in general, from crystal structures has a long history and continues to be an important aspect of protein structure determination. In the last decades, computational methods are becoming more and more

relevant due to the high number of experimental template structures for homology modeling and for deciphering the nature of inter-subunit contacts. However, despite significant advancement in several algorithms designed to specifically tackle modeling of homo-oligomers, and advances in protein–protein docking algorithms in general, there is still plenty of room for improvement.

Highly accurate models can be predicted for homodimers, especially when good templates are available. However, predictions are poorer for higher-order oligomers or cases without suitable homology-template of the complex, as can be seen from CASP (critical assessment of protein structure prediction)-CAPRI (critical assessment of predicted interactions) experiments [94,154]. For example, in the last CASP-CAPRI experiment, no good predictions were made for 4 out of 6 difficult targets [154].

Evaluation of protein assembly predictions in CASP13 led to similar conclusions [164]. The authors of the evaluations also pointed out three challenges that need to be addressed to improve docking predictions: (1) combining modeling of subunits with modeling of the complex; (2) separating intra-chain from inter-chain contacts; and (3) improving the evaluation of isologous interfaces between subunits, especially in the case of inter-subunit interactions between the same residues or residues that are very close in the amino acid sequence. Considering recent advances of the protein structure prediction approaches using deep learning, expectations for improved and novel modelling approaches for protein complexes are likewise high.

References

1. Levy, E.D.; Teichmann, S. Structural, evolutionary, and assembly principles of protein oligomerization. *Prog. Mol. Biol. Transl. Sci.* **2013**, *117*, 25–51. [PubMed]
2. Marsh, J.A.; Rees, H.A.; Ahnert, S.E.; Teichmann, S.A. Structural and evolutionary versatility in protein complexes with uneven stoichiometry. *Nat. Commun.* **2015**, *6*, 1–10. [CrossRef]
3. Stossel, T.P. From signal to pseudopod. How cells control cytoplasmic actin assembly. *J. Biol. Chem.* **1989**, *264*, 18261–18264. [CrossRef]
4. Mitchison, T.; Kirschner, M. Dynamic instability of microtubule growth. *Nature* **1984**, *312*, 237–242. [CrossRef]
5. Renatus, M.; Stennicke, H.R.; Scott, F.L.; Liddington, R.C.; Salvesen, G.S. Dimer formation drives the activation of the cell death protease caspase 9. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 14250–14255. [CrossRef]
6. Yu, X.; Sharma, K.D.; Takahashi, T.; Iwamoto, R.; Mekada, E. Ligand-independent dimer formation of epidermal growth factor receptor (EGFR) is a step separable from ligand-induced EGFR signaling. *Mol. Biol. Cell* **2002**, *13*, 2547–2557. [CrossRef]
7. Jiang, G.; den Hertog, J.; Hunter, T. Receptor-like protein tyrosine phosphatase alpha homodimerizes on the cell surface. *Mol. Cell. Biol.* **2000**, *20*, 5917–5929. [CrossRef]

8.   Navia, M.A.; Fitzgerald, P.M.; McKeever, B.M.; Leu, C.T.; Heimbach, J.C.; Herber, W.K.; Sigal, I.S.; Darke, P.L.; Springer, J.P. Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. *Nature* **1989**, *337*, 615–620. [CrossRef] [PubMed]

9.   Schamel, W.W.A.; Alarcon, B.; Höfer, T.; Minguet, S. The Allostery Model of TCR Regulation. *J. Immunol.* **2017**, *198*, 47–52. [CrossRef]

10. Fushinobu, S.; Ohta, T.; Matsuzawa, H. Homotropic Activation via the Subunit Interaction and Allosteric Symmetry Revealed on Analysis of Hybrid Enzymes ofl-Lactate Dehydrogenase *. *J. Biol. Chem.* **1998**, *273*, 2971–2976. [CrossRef]

11. Bergendahl, L.T.; Marsh, J.A. Functional determinants of protein assembly into homomeric complexes. *Sci. Rep.* **2017**, *7*, 4932. [CrossRef]

12. Park, H.H. Domain swapping of death domain superfamily: Alternative strategy for dimerization. *Int. J. Biol. Macromol.* **2019**, *138*, 565–572. [CrossRef]

13. Zegers, I.; Deswarte, J.; Wyns, L. Trimeric domain-swapped barnase. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 818–822. [CrossRef]

14. Bennett, M.J.; Sawaya, M.R.; Eisenberg, D. Deposition diseases and 3D domain swapping. *Structure* **2006**, *14*, 811–824. [CrossRef] [PubMed]

15. Lynch, M. The evolution of multimeric protein assemblages. *Mol. Biol. Evol.* **2012**, *29*, 1353–1366. [CrossRef]

16. Lynch, M. Evolutionary diversification of the multimeric states of proteins. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, E2821–E2828. [CrossRef] [PubMed]

17. Hagner, K.; Setayeshgar, S.; Lynch, M. Stochastic protein multimerization, activity, and fitness. *Phys. Rev. E* **2018**, *98*, 062401. [CrossRef]

18. Goodsell, D.S.; Olson, A.J. Structural symmetry and protein function. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 105–153. [CrossRef]

19. Ali, M.H.; Imperiali, B. Protein oligomerization: How and why. *Bioorg. Med. Chem.* **2005**, *13*, 5013–5020. [CrossRef]

20. Perica, T.; Marsh, J.A.; Sousa, F.L.; Natan, E.; Colwell, L.J.; Ahnert, S.E.; Teichmann, S.A. The emergence of protein complexes: Quaternary structure, dynamics and allostery. Colworth Medal Lecture. *Biochem. Soc. Trans.* **2012**, *40*, 475–491. [CrossRef]

21. Griffin, M.D.W.; Gerrard, J.A. The relationship between oligomeric state and protein function. *Adv. Exp. Med. Biol.* **2012**, *747*, 74–90. [PubMed]

22. Bonjack, M.; Avnir, D. The near-symmetry of protein oligomers: NMR-derived structures. *Sci. Rep.* **2020**, *10*, 8367. [CrossRef]

23. Bonjack-Shterengartz, M.; Avnir, D. The enigma of the near-symmetry of proteins: Domain swapping. *PLoS ONE* **2017**, *12*, e0180030. [CrossRef] [PubMed]

24. Swapna, L.S.; Srikeerthana, K.; Srinivasan, N. Extent of structural asymmetry in homodimeric proteins: Prevalence and relevance. *PLoS ONE* **2012**, *7*, e36688. [CrossRef] [PubMed]

25. Dey, S.; Ritchie, D.W.; Levy, E.D. PDB-wide identification of biological assemblies from conserved quaternary structure geometry. *Nat. Methods* **2017**, *15*, 67–72. [CrossRef] [PubMed]

26. Johnson, J.E.; Olson, A.J. Icosahedral virus structures and the protein data bank. *J. Biol. Chem.* **2021**, *296*, 100554. [CrossRef]

27. Lukatsky, D.B.; Zeldovich, K.B.; Shakhnovich, E.I. Statistically enhanced self-attraction of random patterns. *Phys. Rev. Lett.* **2006**, *97*, 178101. [CrossRef]

28. Lukatsky, D.B.; Shakhnovich, B.E.; Mintseris, J.; Shakhnovich, E.I. Structural similarity enhances interaction propensity of proteins. *J. Mol. Biol.* **2007**, *365*, 1596–1606. [CrossRef]

29. André, I.; Strauss, C.E.M.; Kaplan, D.B.; Bradley, P.; Baker, D. Emergence of symmetry in homooligomeric biological assemblies. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 16148–16152. [CrossRef]

30. Schulz, G.E. The dominance of symmetry in the evolution of homo-oligomeric proteins. *J. Mol. Biol.* **2010**, *395*, 834–843. [CrossRef]

31. Chiti, F.; Dobson, C.M. Amyloid formation by globular proteins under native conditions. *Nat. Chem. Biol.* **2009**, *5*, 15–22. [CrossRef]

32. Liu, Y.; Hart, P.J.; Schlunegger, M.P.; Eisenberg, D. The crystal structure of a 3D domain-swapped dimer of RNase A at a 2.1-A resolution. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 3437–3442. [CrossRef] [PubMed]

33. Pearson, M.A.; Karplus, P.A.; Dodge, R.W.; Laity, J.H.; Scheraga, H.A. Crystal structures of two mutants that have implications for the folding of bovine pancreatic ribonuclease A. *Protein Sci.* **1998**, *7*, 1255–1258. [CrossRef] [PubMed]

34. Zhang, M.; Windheim, M.; Roe, S.M.; Peggie, M.; Cohen, P.; Prodromou, C.; Pearl, L.H. Chaperoned ubiquitylation—crystal structures of the CHIP U box E3 ubiquitin ligase and a CHIP-Ubc13-Uev1a complex. *Mol. Cell* **2005**, *20*, 525–538. [CrossRef]

35. Ortíz, C.; Botti, H.; Buschiazzo, A.; Comini, M.A. Glucose-6-phosphate dehydrogenase from the human pathogen Trypanosoma cruzi evolved unique structural features to support efficient product formation. *J. Mol. Biol.* **2019**, *431*, 2143–2162. [CrossRef]

36. Kerfeld, C.A.; Sawaya, M.R.; Brahmandam, V.; Cascio, D.; Ho, K.K.; Trevithick-Sutton, C.C.; Krogmann, D.W.; Yeates, T.O. The crystal structure of a cyanobacterial water-soluble carotenoid binding protein. *Structure* **2003**, *11*, 55–65. [CrossRef]

37. Mera, P.E.; St Maurice, M.; Rayment, I.; Escalante-Semerena, J.C. Structural and functional analyses of the human-type corrinoid adenosyltransferase (PduO) from Lactobacillus reuteri. *Biochemistry* **2007**, *46*, 13829–13836. [CrossRef]

38. Burley, S.K.; Bhikadiya, C.; Bi, C.; Bittrich, S.; Chen, L.; Crichlow, G.V.; Christie, C.H.; Dalenberg, K.; Di Costanzo, L.; Duarte, J.M.; et al. RCSB Protein Data Bank: Powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Res.* **2021**, *49*, D437–D451. [CrossRef]

39. Yip, K.M.; Fischer, N.; Paknia, E.; Chari, A.; Stark, H. Atomic-resolution protein structure determination by cryo-EM. *Nature* **2020**, *587*, 157–161. [CrossRef]

40. Powell, H.R. X-ray data processing. *Biosci. Rep.* **2017**, *37*, BSR20170227. [CrossRef]

41. Wlodawer, A.; Minor, W.; Dauter, Z.; Jaskolski, M. Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination. *FEBS J.* **2013**, *280*, 5705–5736. [CrossRef]

42. Vonck, J.; Mills, D.J. Advances in high-resolution cryo-EM of oligomeric enzymes. *Curr. Opin. Struct. Biol.* **2017**, *46*, 48–54. [CrossRef]

43. Wlodawer, A.; Minor, W.; Dauter, Z.; Jaskolski, M. Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *FEBS J.* **2008**, *275*, 1–21. [CrossRef]

44. Kantardjieff, K.A.; Rupp, B. Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci.* **2003**, *12*, 1865–1871. [CrossRef]

45. Eyal, E.; Gerzon, S.; Potapov, V.; Edelman, M.; Sobolev, V. The limit of accuracy of protein modeling: Influence of crystal packing on protein structure. *J. Mol. Biol.* **2005**, *351*, 431–442. [CrossRef]

46. Juers, D.H.; Matthews, B.W. Reversible lattice repacking illustrates the temperature dependence of macromolecular interactions. *J. Mol. Biol.* **2001**, *311*, 851–862. [CrossRef]

47. Dafforn, T.R. So how do you know you have a macromolecular complex? *Acta Crystallogr. D Biol. Crystallogr.* **2007**, *63*, 17–25. [CrossRef] [PubMed]

48. Kuznetsova, I.M.; Turoverov, K.K.; Uversky, V.N. What macromolecular crowding can do to a protein. *Int. J. Mol. Sci.* **2014**, *15*, 23090–23140. [CrossRef] [PubMed]

49. Banatao, D.R.; Cascio, D.; Crowley, C.S.; Fleissner, M.R.; Tienson, H.L.; Yeates, T.O. An approach to crystallizing proteins by synthetic symmetrization. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 16230–16235. [CrossRef] [PubMed]

50. Chesterman, C.; Arnold, E. Co-crystallization with diabodies: A case study for the introduction of synthetic symmetry. *Structure* **2021**, *29*, 598–605. [CrossRef] [PubMed]

51. Chantler, C.; Bunker, B.; Boscherini, F. *International Tables for Crystallography, X-ray Absorption Spectroscopy and Related Techniques*; Wiley: Hoboken, NJ, USA; ISBN 9781119433941. in press.

52. International Tables for Crystallography. *International Tables for Crystallography*, 2012.

53. Dauter, Z.; Jaskolski, M. How to read (and understand) Volume A of International Tables for Crystallography: An introduction for nonspecialists. *J. Appl. Crystallogr.* **2010**, *43*, 1150–1171. [CrossRef]

54. Chruszcz, M.; Potrzebowski, W.; Zimmerman, M.D.; Grabowski, M.; Zheng, H.; Lasota, P.; Minor, W. Analysis of solvent content and oligomeric states in protein crystals–does symmetry matter? *Protein Sci.* **2008**, *17*, 623–632. [CrossRef] [PubMed]

55. Jouravel, N.; Sablin, E.; Togashi, M.; Baxter, J.D.; Webb, P.; Fletterick, R.J. Molecular basis for dimer formation of TRbeta variant D355R. *Proteins* **2009**, *75*, 111–117. [CrossRef] [PubMed]

56. Capitani, G.; Duarte, J.M.; Baskaran, K.; Bliven, S.; Somody, J.C. Understanding the fabric of protein crystals: Computational classification of biological interfaces and crystal contacts. *Bioinformatics* **2016**, *32*, 481–489. [CrossRef]

57. Elez, K.; Bonvin, A.M.J.J.; Vangone, A. Biological vs. Crystallographic Protein Interfaces: An Overview of Computational Approaches for Their Classification. *Crystals* **2020**, *10*, 114. [CrossRef]

58. Krissinel, E.; Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **2007**, *372*, 774–797. [CrossRef]

59. Krissinel, E. Crystal contacts as nature's docking solutions. *J. Comput. Chem.* **2010**, *31*, 133–143. [CrossRef]

60. Shrake, A.; Rupley, J.A. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J. Mol. Biol.* **1973**, *79*, 351–371. [CrossRef]

61. 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]

62. 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]

63. Duarte, J.M.; Srebniak, A.; Schärer, M.A.; Capitani, G. Protein interface classification by evolutionary analysis. *BMC Bioinformatics* **2012**, *13*, 334. [CrossRef]

64. Liu, S.; Li, Q.; Lai, L. A combinatorial score to distinguish biological and nonbiological protein-protein interfaces. *Proteins* **2006**, *64*, 68–78. [CrossRef] [PubMed]

65. Liu, Q.; Li, Z.; Li, J. Use B-factor related features for accurate classification between protein binding interfaces and crystal packing contacts. *BMC Bioinform.* **2014**, *15* (Suppl. 16), S3. [CrossRef]

66. Tsuchiya, Y.; Kinoshita, K.; Ito, N.; Nakamura, H. PreBI: Prediction of biological interfaces of proteins in crystals. *Nucleic Acids Res.* **2006**, *34*, W320–W324. [CrossRef] [PubMed]

67. Tsuchiya, Y.; Nakamura, H.; Kinoshita, K. Discrimination between biological interfaces and crystal-packing contacts. *Adv. Appl. Bioinform. Chem.* **2008**, *1*, 99–113. [CrossRef]

68. Fukasawa, Y.; Tomii, K. Accurate Classification of Biological and non-Biological Interfaces in Protein Crystal Structures using Subtle Covariation Signals. *Sci. Rep.* **2019**, *9*, 12603. [CrossRef]

69. Elez, K.; Bonvin, A.M.J.J.; Vangone, A. Distinguishing crystallographic from biological interfaces in protein complexes: Role of intermolecular contacts and energetics for classification. *BMC Bioinform.* **2018**, *19*, 438. [CrossRef]

70. Jiménez-García, B.; Elez, K.; Koukos, P.I.; Bonvin, A.M.; Vangone, A. PRODIGY-crystal: A web-tool for classification of biological interfaces in protein complexes. *Bioinformatics* **2019**, *35*, 4821–4823. [CrossRef]

71. Baskaran, K.; Duarte, J.M.; Biyani, N.; Bliven, S.; Capitani, G. A PDB-wide, evolution-based assessment of protein-protein interfaces. *BMC Struct. Biol.* **2014**, *14*, 22. [CrossRef]

72. Hu, J.; Liu, H.-F.; Sun, J.; Wang, J.; Liu, R. Integrating co-evolutionary signals and other properties of residue pairs to distinguish biological interfaces from crystal contacts. *Protein Sci.* **2018**, *27*, 1723–1735. [CrossRef]

73. Luo, M.; Tanner, J.J. Structural basis of substrate recognition by aldehyde dehydrogenase 7A1. *Biochemistry* **2015**, *54*, 5513–5522. [CrossRef] [PubMed]

74. Dhatwalia, R.; Singh, H.; Oppenheimer, M.; Karr, D.B.; Nix, J.C.; Sobrado, P.; Tanner, J.J. Crystal structures and small-angle x-ray scattering analysis of UDP-galactopyranose mutase from the pathogenic fungus Aspergillus fumigatus. *J. Biol. Chem.* **2012**, *287*, 9041–9051. [CrossRef] [PubMed]

75. Kwan, A.H.; Mobli, M.; Gooley, P.R.; King, G.F.; Mackay, J.P. Macromolecular NMR spectroscopy for the non-spectroscopist. *FEBS J.* **2011**, *278*, 687–703. [CrossRef]

76. Yu, H. Extending the size limit of protein nuclear magnetic resonance. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 332–334. [CrossRef] [PubMed]

77. Sgourakis, N.G.; Lange, O.F.; DiMaio, F.; André, I.; Fitzkee, N.C.; Rossi, P.; Montelione, G.T.; Bax, A.; Baker, D. Determination of the structures of symmetric protein oligomers from NMR chemical shifts and residual dipolar couplings. *J. Am. Chem. Soc.* **2011**, *133*, 6288–6298. [CrossRef] [PubMed]

78. Foster, M.P.; McElroy, C.A.; Amero, C.D. Solution NMR of large molecules and assemblies. *Biochemistry* **2007**, *46*, 331–340. [CrossRef] [PubMed]

79. Chen, K.; Tjandra, N. The use of residual dipolar coupling in studying proteins by NMR. *Top. Curr. Chem.* **2012**, *326*, 47–67.

80. Wang, J.; Zuo, X.; Yu, P.; Byeon, I.-J.L.; Jung, J.; Wang, X.; Dyba, M.; Seifert, S.; Schwieters, C.D.; Qin, J.; et al. Determination of multicomponent protein structures in solution using global orientation and shape restraints. *J. Am. Chem. Soc.* **2009**, *131*, 10507–10515. [CrossRef]

81. Yu, X.; Jin, L.; Zhou, Z.H. 3.88 A structure of cytoplasmic polyhedrosis virus by cryo-electron microscopy. *Nature* **2008**, *453*, 415–419. [CrossRef]

82. van Heel, M. Multivariate statistical classification of noisy images (randomly oriented biological macromolecules). *Ultramicroscopy* **1984**, *13*, 165–183. [CrossRef]

83. Costa, A.; Patwardhan, A. A novel mirror-symmetry analysis approach for the study of macromolecular assemblies imaged by electron microscopy. *J. Mol. Biol.* **2008**, *378*, 273–283. [CrossRef] [PubMed]

84. Reboul, C.F.; Kiesewetter, S.; Elmlund, D.; Elmlund, H. Point-group symmetry detection in three-dimensional charge density of biomolecules. *Bioinformatics* **2020**, *36*, 2237–2243. [CrossRef] [PubMed]

85. Cozza, G.; Moro, S.; Gotte, G. Elucidation of the ribonuclease A aggregation process mediated by 3D domain swapping: A computational approach reveals possible new multimeric structures. *Biopolymers* **2008**, *89*, 26–39. [CrossRef]

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

87. Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H.J. PatchDock and SymmDock: Servers for rigid and symmetric docking. *Nucleic Acids Res.* **2005**, *33*, W363–W367. [CrossRef] [PubMed]

88. Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H.J. Geometry-based flexible and symmetric protein docking. *Proteins* **2005**, *60*, 224–231. [CrossRef] [PubMed]

89. Comeau, S.R.; Camacho, C.J. Predicting oligomeric assemblies: N-mers a primer. *J. Struct. Biol.* **2005**, *150*, 233–244. [CrossRef]

90. Kozakov, D.; Beglov, D.; Bohnuud, T.; Mottarella, S.E.; Xia, B.; Hall, D.R.; Vajda, S. How good is automated protein docking? *Proteins* **2013**, *81*, 2159–2166. [CrossRef]

91. Desta, I.T.; Porter, K.A.; Xia, B.; Kozakov, D.; Vajda, S. Performance and Its Limits in Rigid Body Protein-Protein Docking. *Structure* **2020**, *28*, 1071–1081. [CrossRef]

92. Tovchigrechko, A.; Vakser, I.A. Development and testing of an automated approach to protein docking. *Proteins* **2005**, *60*, 296–301. [CrossRef] [PubMed]

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

94. 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 homoprotein and heteroprotein complexes by protein docking and template-based modeling: A CASP-CAPRI experiment. *Proteins* **2016**, *84* (Suppl. 1), 323–348. [CrossRef]

95. Berchanski, A.; Eisenstein, M. Construction of molecular assemblies via docking: Modeling of tetramers with D2 symmetry. *Proteins* **2003**, *53*, 817–829. [CrossRef] [PubMed]

96. 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]

97. Berchanski, A.; Segal, D.; Eisenstein, M. Modeling oligomers with Cn or Dn symmetry: Application to CAPRI target 10. *Proteins* **2005**, *60*, 202–206. [CrossRef]

98. Ritchie, D.W.; Grudinin, S. Spherical polar Fourier assembly of protein complexes with arbitrary point group symmetry. *J. Appl. Crystallogr.* **2016**, *49*, 158–167. [CrossRef]

99. Huang, S.-Y.; Zou, X. An iterative knowledge-based scoring function for protein-protein recognition. *Proteins Struct. Funct. Bioinform.* **2008**, *72*, 557–579. [CrossRef]

100. Yan, Y.; Wen, Z.; Wang, X.; Huang, S.-Y. Addressing recent docking challenges: A hybrid strategy to integrate template-based and free protein-protein docking. *Proteins* **2017**, *85*, 497–512. [CrossRef]

101. Yan, Y.; Tao, H.; Huang, S.-Y. HSYMDOCK: A docking web server for predicting the structure of protein homo-oligomers with Cn or Dn symmetry. *Nucleic Acids Res.* **2018**, *46*, W423–W431. [CrossRef]

102. Park, T.; Baek, M.; Lee, H.; Seok, C. GalaxyTongDock: Symmetric and asymmetric ab initio protein-protein docking web server with improved energy parameters. *J. Comput. Chem.* **2019**, *40*, 2413–2417. [CrossRef]

103. Dominguez, C.; Boelens, R.; Bonvin, A.M.J.J. HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **2003**, *125*, 1731–1737. [CrossRef] [PubMed]

104. Karaca, E.; Melquiond, A.S.J.; de Vries, S.J.; Kastritis, P.L.; Bonvin, A.M.J.J. Building macromolecular assemblies by information-driven docking: Introducing the HADDOCK multibody docking server. *Mol. Cell. Proteom.* **2010**, *9*, 1784–1794. [CrossRef]

105. van Zundert, G.C.P.; Rodrigues, J.P.G.L.; Trellet, M.; Schmitz, C.; Kastritis, P.L.; Karaca, E.; Melquiond, A.S.J.; van Dijk, M.; de Vries, S.J.; Bonvin, A.M.J. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* **2016**, *428*, 720–725. [CrossRef]

106. André, I.; Bradley, P.; Wang, C.; Baker, D. Prediction of the structure of symmetrical protein assemblies. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 17656–17661. [CrossRef] [PubMed]

107. Lyskov, S.; Chou, F.-C.; Conchúir, S.Ó.; Der, B.S.; Drew, K.; Kuroda, D.; Xu, J.; Weitzner, B.D.; Douglas Renfrew, P.; Sripakdeevong, P.; et al. Serverification of Molecular Modeling Applications: The Rosetta Online Server That Includes Everyone (ROSIE). *PLoS ONE* **2013**, *8*, e63906. [CrossRef] [PubMed]

108. Das, R.; André, I.; Shen, Y.; Wu, Y.; Lemak, A.; Bansal, S.; Arrowsmith, C.H.; Szyperski, T.; Baker, D. Simultaneous prediction of protein folding and docking at high resolution. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 18978–18983. [CrossRef]

109. Roy Burman, S.S.; Yovanno, R.A.; Gray, J.J. Flexible Backbone Assembly and Refinement of Symmetrical Homomeric Complexes. *Structure* **2019**, *27*, 1041–1051.e8. [CrossRef]

110. Guex, N.; Peitsch, M.C. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* **1997**, *18*, 2714–2723. [CrossRef]

111. Schwede, T.; Kopp, J.; Guex, N.; Peitsch, M.C. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res.* **2003**, *31*, 3381–3385. [CrossRef]

112. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F.T.; de Beer, T.A.P.; Rempfer, C.; Bordoli, L.; et al. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46*, W296–W303. [CrossRef]

113. Lee, H.; Park, H.; Ko, J.; Seok, C. GalaxyGemini: A web server for protein homo-oligomer structure prediction based on similarity. *Bioinformatics* **2013**, *29*, 1078–1080. [CrossRef]

114. Baek, M.; Park, T.; Heo, L.; Park, C.; Seok, C. GalaxyHomomer: A web server for protein homo-oligomer structure prediction from a monomer sequence or structure. *Nucleic Acids Res.* **2017**, *45*, W320–W324. [CrossRef]

115. Baek, M.; Park, T.; Heo, L.; Seok, C. Modeling Protein Homo-Oligomer Structures with GalaxyHomomer Web Server. In *Protein Structure Prediction*; Kihara, D., Ed.; Springer US: New York, NY, 2020; pp. 127–137. ISBN 9781071607084.

116. Yan, Y.; Tao, H.; He, J.; Huang, S.-Y. The HDOCK server for integrated protein-protein docking. *Nat. Protoc.* **2020**, *15*, 1829–1852. [CrossRef] [PubMed]

117. Porter, K.A.; Padhorny, D.; Desta, I.; Ignatov, M.; Beglov, D.; Kotelnikov, S.; Sun, Z.; Alekseenko, A.; Anishchenko, I.; Cong, Q.; et al. Template-based modeling by ClusPro in CASP13 and the potential for using co-evolutionary information in docking. *Proteins* **2019**, *87*, 1241–1248. [CrossRef] [PubMed]

118. DiMaio, F.; Leaver-Fay, A.; Bradley, P.; Baker, D.; André, I. Modeling symmetric macromolecular structures in Rosetta3. *PLoS ONE* **2011**, *6*, e20450. [CrossRef] [PubMed]

119. Song, Y.; DiMaio, F.; Wang, R.Y.-R.; Kim, D.; Miles, C.; Brunette, T.; Thompson, J.; Baker, D. High-resolution comparative modeling with RosettaCM. *Structure* **2013**, *21*, 1735–1742. [CrossRef] [PubMed]

120. 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]

121. Mintseris, J.; Pierce, B.; Wiehe, K.; Anderson, R.; Chen, R.; Weng, Z. Integrating statistical pair potentials into protein complex prediction. *Proteins* **2007**, *69*, 511–520. [CrossRef]

122. Vreven, T.; Schweppe, D.K.; Chavez, J.D.; Weisbrod, C.R.; Shibata, S.; Zheng, C.; Bruce, J.E.; Weng, Z. Integrating Cross-Linking Experiments with Ab Initio Protein–Protein Docking. *J. Mol. Biol.* **2018**, *430*, 1814–1828. [CrossRef]

123. Duhovny, D.; Nussinov, R.; Wolfson, H.J. *Efficient Unbound Docking of Rigid Molecules. In Proceedings of the Algorithms in Bioinformatics*; Springer: Berlin/Heidelberg, Germany, 2002; pp. 185–200.

124. Gaber, A.; Kim, S.J.; Kaake, R.M.; Benčina, M.; Krogan, N.; Šali, A.; Pavšič, M.; Lenarčič, B. EpCAM homo-oligomerization is not the basis for its role in cell-cell adhesion. *Sci. Rep.* **2018**, *8*, 13269. [CrossRef] [PubMed]

125. Schneidman-Duhovny, D.; Hammel, M.; Tainer, J.A.; Sali, A. FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. *Nucleic Acids Res.* **2016**, *44*, W424–W429. [CrossRef]

126. Xia, B.; Vajda, S.; Kozakov, D. Accounting for pairwise distance restraints in FFT-based protein–protein docking. *Bioinformatics* **2016**, *32*, 3342–3344. [CrossRef]

127. Gaber, A.; Gunčar, G.; Pavšič, M. Proper evaluation of chemical cross-linking-based spatial restraints improves the precision of modeling homo-oligomeric protein complexes. *BMC Bioinform.* **2019**, *20*, 464. [CrossRef]

128. 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]

129. Ignatov, M.; Kazennov, A.; Kozakov, D. ClusPro FMFT-SAXS: Ultra-fast Filtering Using Small-Angle X-ray Scattering Data in Protein Docking. *J. Mol. Biol.* **2018**, *430*, 2249–2255. [CrossRef]

130. Yan, Y.; Huang, S.-Y. CHDOCK: A hierarchical docking approach for modeling Cn symmetric homo-oligomeric complexes. *Biophys. Rep.* **2019**, *5*, 65–72. [CrossRef]

131. Yan, Y.; Huang, S.-Y. Protein-Protein Docking with Improved Shape Complementarity. In Proceedings of the Intelligent Computing Theories and Application, Wuhan, China, 15–18 August 2018; Springer International Publishing: Berlin/Heidelberg, Germany, 2018; pp. 600–605.

132. van Dijk, A.D.J.; Boelens, R.; Bonvin, A.M.J.J. Data-driven docking for the study of biomolecular complexes. *FEBS J.* **2005**, *272*, 293–312. [CrossRef] [PubMed]

133. van Dijk, A.D.J.; Fushman, D.; Bonvin, A.M.J.J. Various strategies of using residual dipolar couplings in NMR-driven protein docking: Application to Lys48-linked di-ubiquitin and validation against 15N-relaxation data. *Proteins* **2005**, *60*, 367–381. [CrossRef] [PubMed]

134. van Dijk, A.D.J.; Kaptein, R.; Boelens, R.; Bonvin, A.M.J.J. Combining NMR relaxation with chemical shift perturbation data to drive protein-protein docking. *J. Biomol. NMR* **2006**, *34*, 237–244. [CrossRef]

135. Schmitz, C.; Bonvin, A.M.J.J. Protein–protein HADDocking using exclusively pseudocontact shifts. *J. Biomol. NMR* **2011**, *50*, 263–266. [CrossRef] [PubMed]

136. de Vries, S.J.; Bonvin, A.M.J.J. CPORT: A consensus interface predictor and its performance in prediction-driven docking with HADDOCK. *PLoS ONE* **2011**, *6*, e17695. [CrossRef] [PubMed]

137. Karaca, E.; Bonvin, A.M.J.J. On the usefulness of ion-mobility mass spectrometry and SAXS data in scoring docking decoys. *Acta Crystallogr. D Biol. Crystallogr.* **2013**, *69*, 683–694. [CrossRef] [PubMed]

138. van Zundert, G.C.P.; Melquiond, A.S.J.; Bonvin, A.M.J.J. Integrative Modeling of Biomolecular Complexes: HADDOCKing with Cryo-Electron Microscopy Data. *Structure* **2015**, *23*, 949–960. [CrossRef] [PubMed]

139. Trellet, M.; van Zundert, G.; Bonvin, A.M.J.J. Protein–Protein Modeling Using Cryo-EM Restraints. In *Structural Bioinformatics: Methods and Protocols*; Gáspári, Z., Ed.; Springer US: New York, NY, 2020; pp. 145–162. ISBN 9781071602706.

140. Shen, Y.; Lange, O.; Delaglio, F.; Rossi, P.; Aramini, J.M.; Liu, G.; Eletsky, A.; Wu, Y.; Singarapu, K.K.; Lemak, A.; et al. Consistent blind protein structure generation from NMR chemical shift data. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 4685–4690. [CrossRef] [PubMed]

141. Kahraman, A.; Herzog, F.; Leitner, A.; Rosenberger, G.; Aebersold, R.; Malmström, L. Cross-link guided molecular modeling with ROSETTA. *PLoS ONE* **2013**, *8*, e73411. [CrossRef]

142. 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]

143. Ovchinnikov, S.; Kamisetty, H.; Baker, D. Robust and accurate prediction of residue–residue interactions across protein interfaces using evolutionary information. *Elife* **2014**, *3*, e02030. [CrossRef]

144. Biasini, M.; Bienert, S.; Waterhouse, A.; Arnold, K.; Studer, G.; Schmidt, T.; Kiefer, F.; Gallo Cassarino, T.; Bertoni, M.; Bordoli, L.; et al. SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **2014**, *42*, W252–W258. [CrossRef]

145. Bertoni, M.; Kiefer, F.; Biasini, M.; Bordoli, L.; Schwede, T. Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Sci. Rep.* **2017**, *7*, 1–15. [CrossRef]

146. Söding, J. Protein homology detection by HMM-HMM comparison. *Bioinformatics* **2005**, *21*, 951–960. [CrossRef]

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

148. Ko, J.; Park, H.; Seok, C. GalaxyTBM: Template-based modeling by building a reliable core and refining unreliable local regions. *BMC Bioinform.* **2012**, *13*, 198. [CrossRef] [PubMed]

149. Lee, J.; Lee, D.; Park, H.; Coutsias, E.A.; Seok, C. Protein loop modeling by using fragment assembly and analytical loop closure. *Proteins* **2010**, *78*, 3428–3436. [CrossRef] [PubMed]

150. Park, H.; Seok, C. Refinement of unreliable local regions in template-based protein models. *Proteins* **2012**, *80*, 1974–1986. [CrossRef] [PubMed]

151. Park, H.; Lee, G.R.; Heo, L.; Seok, C. Protein loop modeling using a new hybrid energy function and its application to modeling in inaccurate structural environments. *PLoS ONE* **2014**, *9*, e113811.

152. Heo, L.; Lee, H.; Seok, C. GalaxyRefineComplex: Refinement of protein-protein complex model structures driven by interface repacking. *Sci. Rep.* **2016**, *6*, 32153. [CrossRef]

153. Park, H.; Kim, D.E.; Ovchinnikov, S.; Baker, D.; DiMaio, F. Automatic structure prediction of oligomeric assemblies using Robetta in CASP12. *Proteins* **2018**, *86 Suppl 1*, 283–291. [CrossRef]

154. Lensink, M.F.; Brysbaert, G.; Nadzirin, N.; Velankar, S.; Chaleil, R.A.G.; Gerguri, T.; Bates, P.A.; Laine, E.; Carbone, A.; Grudinin, S.; et al. Blind prediction of homo- and hetero-protein complexes: The CASP13-CAPRI experiment. *Proteins* **2019**, *87*, 1200–1221. [CrossRef]

155. Torchala, M.; Moal, I.H.; Chaleil, R.A.G.; Fernandez-Recio, J.; Bates, P.A. SwarmDock: A server for flexible protein–protein docking. *Bioinform.* **2013**, *29*, 807–809. [CrossRef] [PubMed]

156. Dapkunas, J.; Timinskas, A.; Olechnovic, K.; Margelevicius, M.; Diciunas, R.; Venclovas, C. The PPI3D web server for searching, analyzing and modeling protein-protein interactions in the context of 3D structures. *Bioinformatics* **2017**, *33*, 935–937. [CrossRef] [PubMed]

157. Dapkūnas, J.; Venclovas, Č. Template-Based Modeling of Protein Complexes Using the PPI3D Web Server. *Methods Mol. Biol.* **2020**, *2165*, 139–155. [PubMed]

158. Esquivel-Rodriguez, J.; Filos-Gonzalez, V.; Li, B.; Kihara, D. Pairwise and multimeric protein-protein docking using the LZerD program suite. *Methods Mol. Biol.* **2014**, *1137*, 209–234.

159. Christoffer, C.; Chen, S.; Bharadwaj, V.; Aderinwale, T.; Kumar, V.; Hormati, M.; Kihara, D. LZerD webserver for pairwise and multiple protein–protein docking. *Nucleic Acids Res.* **2021**, *49*, W359–W365. [CrossRef] [PubMed]

160. Huang, S.-Y.; Zou, X. MDockPP: A hierarchical approach for protein-protein docking and its application to CAPRI rounds 15–19. *Proteins Struct. Funct. Bioinform.* **2010**, *78*, 3096–3103. [CrossRef]

161. Torres, P.H.M.; Rossi, A.D.; Blundell, T.L. ProtCHOIR: A tool for proteome-scale generation of homo-oligomers. *Brief. Bioinform.* **2021**. [CrossRef]

162. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *15*, 1–7. [CrossRef]

163. Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G.R.; Wang, J.; Cong, Q.; Kinch, L.N.; Schaeffer, R.D.; et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **2021**, *373*, 871–876. [CrossRef]

164. Guzenko, D.; Lafita, A.; Monastyrskyy, B.; Kryshtafovych, A.; Duarte, J.M. Assessment of protein assembly prediction in CASP13. *Proteins* **2019**, *87*, 1190–1199. [CrossRef]