

Bachelor's Programme in Chemical Engineering

Computational modelling of folded protein–protein interactions

Meri Mänty

Copyright ©2025 Meri Mänty

Author	Meri Mänty	
Title of thesis	Computational modelling of folded protein–protein interactions	
Programme	Bachelor’s Programme in Chemical Engineering	
Major	CHEM3049 Chemistry and Materials Science	
Thesis supervisor	D.Sc. (Tech.) Minna Nieminen	
Thesis advisor	M.Sc. (Tech.) Adam Harmat	
Date	Number of pages	Language
16.05.2025	22	English

Abstract

This thesis is a literature review focused on folded proteins, their interactions, and the state-of-the-art computational methods for predicting folded protein–protein interactions (PPIs). PPIs can be divided into permanent and transient, depending on the lifetime of the protein complexes, which consist of two or more interacting proteins. Studying protein complexes and PPIs is essential for increased understanding of cellular behaviour and drug-discovery, as aberrant interactions have been connected to diseases. PPIs can be uncovered using experimental methods, such as nuclear magnetic resonance, X-ray crystallography, and cryo-electron microscopy. However, experimental methods are time-consuming and expensive compared to computational modelling, which offers a significantly faster and cheaper alternative.

Computational PPI prediction methods have developed rapidly due to the increasing amount of easily available protein sequence-structure data and advancements in machine learning algorithms. PPIs and protein complexes can be modelled using *ab-initio* docking, template-based docking, or machine learning based approaches, such as AlphaFold-Multimer. *Ab-initio* docking uses the structures of the unbound target proteins, whereas template-based docking relies on templates of known protein complexes. AlphaFold-Multimer uses multiple sequence alignments to predict interfacial contacts between the target protein sequences. The quality of PPI predictions can be assessed using different evaluation criteria, such as root-mean-square-deviation (RMSD) and DockQ values, which are also applied in the CASP-CAPRI blind prediction experiments.

Computational methods offer viable alternatives to experimental methods for modelling most folded PPIs. However, further development in the field of computational PPI prediction is still needed. For example, prediction of protein complexes involving antibodies often remains insufficient, even though predicting them could greatly contribute to the development of new medicines and treatments. The future focus of the field is obtaining more data for the machine learning models, where experimental investigations are key.

Keywords folded proteins, protein structures, protein–protein interactions, protein complexes, computational modelling

Tekijä Meri Mänty

Työn nimi Proteiinien välisten vuorovaikutusten tietokonemallinnus

Koulutusohjelma Kemian tekniikan kandidaattiohjelma

Pääaine CHEM3049 Kemia ja materiaalitiede

Vastuuolettaja TkT Minna Nieminen

Työn ohjaaja DI Adam Harmat

Päivämäärä 16.05.2025 **Sivumäärä** 22

Kieli englanti

Tiivistelmä

Tämä kandidaatintutkielma on kirjallisuuskatsaus laskostuneiden proteiinien (engl. *folded proteins*) rakenteista ja vuorovaikutuksista sekä ajankohtaisista tietokonemallinnuksen menetelmistä, joita käytetään näiden vuorovaikutusten ennustamiseen. Proteiinien väliset vuorovaikutukset voidaan jakaa lyhytaikaisiin (engl. *transient*) ja pitkäaikaisiin (engl. *permanent*) riippuen siitä, kuinka pitkäikäisiä näiden vuorovaikutusten muodostamat proteiinikompleksit ovat. Koska useimmat proteiinit vuorovaikuttavat keskenään soluissa, proteiinikomplekseja ja vuorovaikutuksia tutkimalla voidaan paremmin ymmärtää solujen aineenvaihduntaa ja kehittää uusia lääketieteellisiä hoitomenetelmiä. Proteiinien välisiä vuorovaikutuksia voidaan kokeellisesti tutkia esimerkiksi röntgenkristallografialla, ydinmagneettisella resonanssilla tai kryoelektronimikroskopiolla, mutta kokeelliset menetelmät ovat hitaita ja kalliita verrattuna tietokonemallinnukseen, joka tarjoaa huomattavasti nopeampia ja edullisempia vaihtoehtoja.

Proteiinikompleksien ja -vuorovaikutusten mallinnus on kohdannut suuria edistysaskeleita jatkuvasti lisääntyvän proteiinisekvenssi- ja rakennedatan sekä koneoppimisen hyödyntämisen myötä. Useiden kirjallisuuslähteiden ja tutkimusten perusteella ajankohtaiset proteiinikompleksien mallinnusmenetelmät hyödyntävät yksittäisten proteiinien rakenteita (engl. *ab-initio docking*), ennalta tunnettujen proteiinikompleksien rakenteista tehtyjä templaatteja (engl. *template-based docking*), tai koneoppimista, kuten AlphaFold-Multimer -ohjelman tapauksessa. Mallinnusmenetelmien tuloksia arvioidaan erilaisilla arviointikriteereillä, kuten virheen neliöllisillä keskiarvoilla (RMSD) ja DockQ-arvoilla, joita käytetään myös puolueettomasti järjestetyissä CASP-CAPRI -kokeissa.

Huolimatta merkittävistä edistysaskeleista, alalta löytyy edelleen kohteita jatkokehitykselle. Nykyiset mallinnusmenetelmät eivät esimerkiksi kykene hyväksytysti mallintamaan vasta-aineita sisältäviä proteiinikomplekseja, vaikka niiden mallintaminen voisi tarjota merkittäviä mahdollisuuksia uusien lääketieteellisten hoitojen kehityksessä. Alan tulevaisuus painottuu todennäköisesti koneoppimismalleihin, joiden opettamiseen tarvitaan nykyistä enemmän luotettavaa dataa. Proteiinvuorovaikutusten tutkimiseen käytettyjen mallinnusmenetelmien lisäksi myös kokeelliset menetelmät pysyvätkin edelleen ajankohtaisina, sillä kokeellista tietoa tarvitaan mallinnusmenetelmien kehitykseen sekä tulosten arviointiin ja vertailuun.

Avainsanat laskostuneet proteiinit, proteiinien rakenteet, proteiinien väliset vuorovaikutukset, proteiinikompleksit, tietokonemallinnus

Table of contents

1	Introduction	6
2	Proteins.....	7
2.1	Folded proteins	7
2.2	Protein–protein interactions	11
3	Computational methods for predicting protein–protein interactions..	15
3.1	<i>Ab-initio</i> docking	15
3.2	Template-based docking.....	17
3.3	A machine learning-based approach: AlphaFold-Multimer	18
4	CASP-CAPRI experiments and assessment protocol	21
5	Summary	26
	References.....	28

1 Introduction

Proteins are structurally complex molecules that perform essential functions in cells of all living organisms. For example, proteins participate in molecular transport, gene regulation, supporting cell structures, extracellular signaling, and enzymatic reactions [1]. It has been estimated that 80% of proteins interact with other proteins, forming active protein complexes by protein–protein interactions (PPIs) [2]. Therefore, prediction of PPIs leads to increased understanding of cellular processes and may facilitate the development of new pharmacological treatments.

Interactions between proteins depend on their tertiary structures. Protein 3D structures and PPIs can be uncovered using experimental methods, which are typically time-consuming and expensive [3]. Computational PPI prediction methods offer less resource-consuming alternatives. They have developed at a high rate due to continuously increasing data on protein sequences and the corresponding structures [4]. Additionally, recent machine learning-based approaches have provided significant increases in PPI prediction accuracies [5].

This thesis answers to research questions on how folded proteins interact with each other and what are the computational PPI prediction methods used to predict these interactions. First, folded protein structures, PPIs, and their significance in regulating biological processes will be covered. Next, the computational PPI prediction methods, including *ab-initio* docking, template-based docking, and a machine learning-based approach AlphaFold-Multimer will be introduced. Finally, the performance and assessment of these methods will be discussed by presenting evaluation criteria applied in The Critical Assessment of protein Structure Prediction and the Critical Assessment of Predicted Interactions (CASP-CAPRI experiments) [5].

Computational PPI prediction is considered as one of the upcoming breakthroughs in biological research, although further development is still required. For example, current computational prediction methods perform poorly in predicting short-living (transient) or flexible PPIs [4]. As proteins have by far the most complicated structures and functions of known molecules [1], there also exists various types of PPIs that may require different prediction approaches. This thesis provides guidance for selecting the most appropriate computational method for predicting PPIs between the target proteins.

2 Proteins

Proteins are the main building blocks of all cells [1]. They account for most of the cell's dry mass, and they are responsible for various cellular functions due to their ability to adopt different shapes [1, 6]. For example, enzymes are proteins whose catalytic activities depend strongly on their spatial shapes, or "conformations" [7].

Until the end of the 20th century, the understanding of proteins was dominated by the structure-function paradigm, which assumed that an ordered three-dimensional conformation was a prerequisite for a functional protein [1]. However, intensive reassessment of this classical structure-function paradigm has shown that there also exist many intrinsically disordered proteins that have a high degree of conformational flexibility instead of a stable conformation [8]. These disordered regions account for up to 50% of eukaryotic proteomes [9]. Today, the activity of proteins is not indeed perceived as either active or inactive based on ordered and disordered states, but instead, a continuum between these states [10]. This is important to be aware of, but not further discussed in this thesis. Additional information on disordered proteins and their interactions can be found in literature, such as [11-13]. This thesis focuses on folded proteins with ordered structures.

2.1 Folded proteins

Protein's primary structure is assembled from a set of 20 different amino acids, which join into a linear chain that typically consists of 50 to 2000 subsequent amino acid residues [1]. This polypeptide chain is formed by covalent peptide bonds, that result from dehydration reactions between a carboxyl terminus (COOH) and an amino terminus (NH₂) of adjacent amino acids [1]. The backbone of the polypeptide chain is a repeating structure of core atoms C, N, H, and O. Additionally, the backbone includes different side chains, that are unique for each amino acid and are not involved in peptide bond formation [1]. Based on their electrical charge distribution, side chains are categorized into nonpolar, uncharged polar, negatively charged, and positively charged side chains [1]. Side chains determine the chemical properties of amino acids, and eventually entire proteins, since individual polypeptide chains differ only by the sequence and number of side chains while the backbone chemistry remains the same.

Both the amino acid side chains and the polypeptide backbone participate in protein folding, a process in which the functional three-dimensional structure of a protein is achieved [14]. In the case of folded proteins, the polypeptide chain tends to fold into a 3D structure that has the lowest free energy

[15] and is therefore the most energetically favourable. Protein folding involves formation of secondary and tertiary structures, which together account for the 3D conformation. The secondary structure refers to local spatial arrangements of adjacent amino acids, whereas the tertiary structure is the overall three-dimensional arrangement of the polypeptide chain including the recurring secondary structures [16]. Finally, a quaternary structure is achieved if tertiary structures of several polypeptide chains join together, which is the case for most proteins [17]. Figure 1 illustrates the primary, secondary, tertiary and quaternary structures, and how they are related to each other.

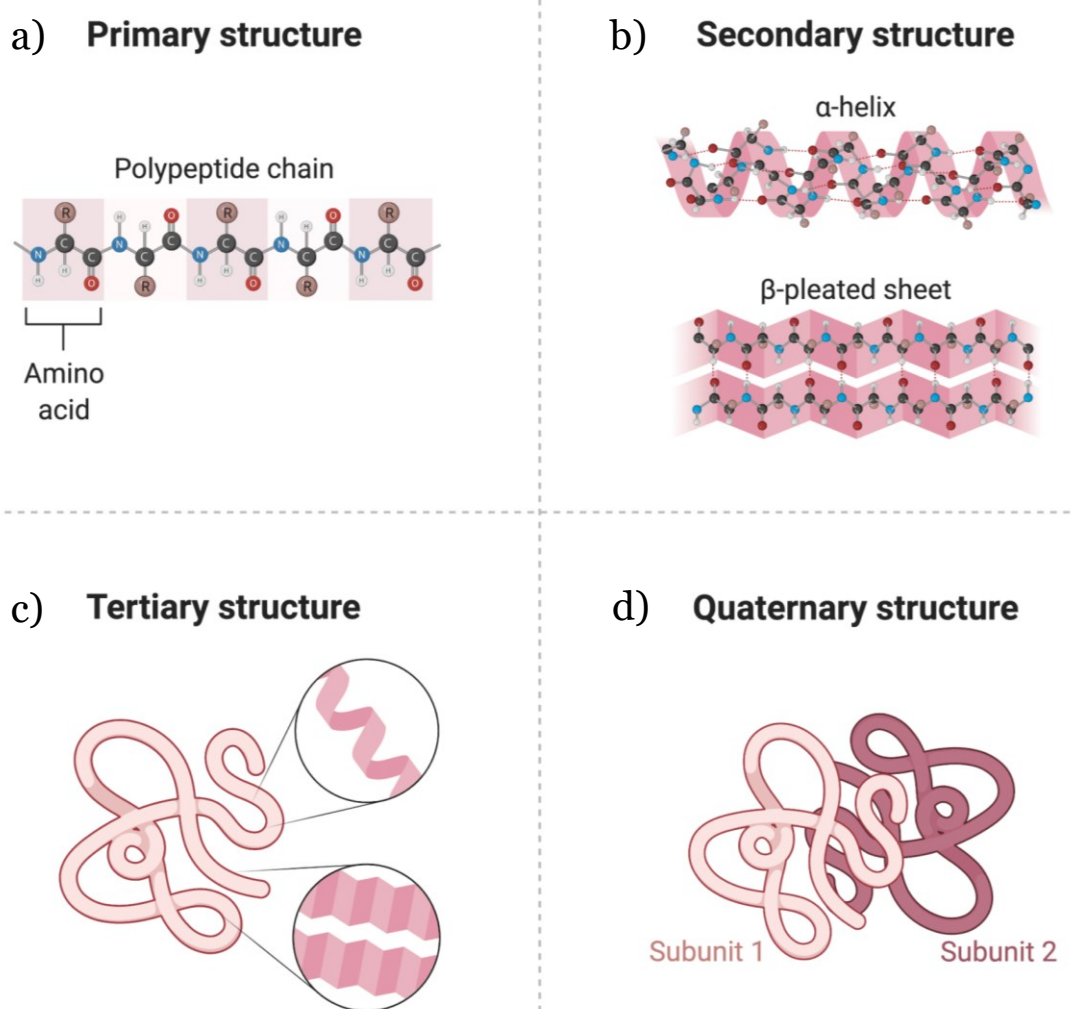


Figure 1. A schematic showing the a) primary, b) secondary, c) tertiary, and d) quaternary structure of a protein. The amino acid side chains are denoted as “R” in the primary structure. [18]

Folding is regulated by noncovalent bonds, including van der Waals attractions, hydrogen bonds, electrostatic interactions, and most importantly, hydrophobic interactions [1, 14]. Noncovalent bonds are much weaker compared to covalent peptide bonds between the backbone atoms. Therefore, the number of weak noncovalent bonds remarkably affects the stability of a fold [1]. The noncovalent bonds that participate in protein folding involve atoms of both the polypeptide backbone and the amino acid side chains.

The protein's secondary structure is achieved when noncovalent bonds, more specifically hydrogen bonds [16], form between nitrogen and oxygen atoms of N-H and C=O groups of the backbone [1]. In general, hydrogen bonds are interactions between an electronegative atom (O, F, N, or S in this context) and a hydrogen atom that is covalently connected to another electronegative atom [19]. Approximately 50% of the structures of all proteins are folded into secondary structures, which include two main recurring motifs, α -helices and β -pleated sheets [20] (see Fig. 1). Secondary structures can occur in all proteins regardless of amino acid sequences, because the hydrogen bonds of secondary structures only involve atoms of the polypeptide backbone, the structure of which remains similar for all polypeptide chains. However, hydrogen bonds that stabilize the protein conformation may also occur between the polypeptide backbone and an amino acid side chain or between two side chains [1], but in this case, they do not participate in the secondary structure formation. Instead, they participate in forming the tertiary structure, where interactions also involve atoms of the amino acid side chains [21].

The tertiary structure of a protein is a compact 3D conformation with low surface-to-volume ratio [21]. The tertiary structure results from noncovalent bonds including hydrophobic interactions, ionic bonds, and hydrogen bonds, but also covalent disulfide bonds [22]. These disulfide linkages remarkably stabilize the protein fold, because a large amount of energy is required to break this covalent bond between sulfur atoms of two cysteine side chains [1]. Whereas cysteine side chains account for these covalent disulfide bonds, other amino acid side chains account for the noncovalent bonds. Uncharged polar side chains can form hydrogen bonds, and charged side chains can form ionic bonds [21], which are electrostatic attractions between two oppositely charged ions [23]. Furthermore, nonpolar side chains account for hydrophobic interactions [1], which are further discussed below. The different interactions and chemical bonds responsible for forming and stabilizing the tertiary structure are compiled in Figure 2.

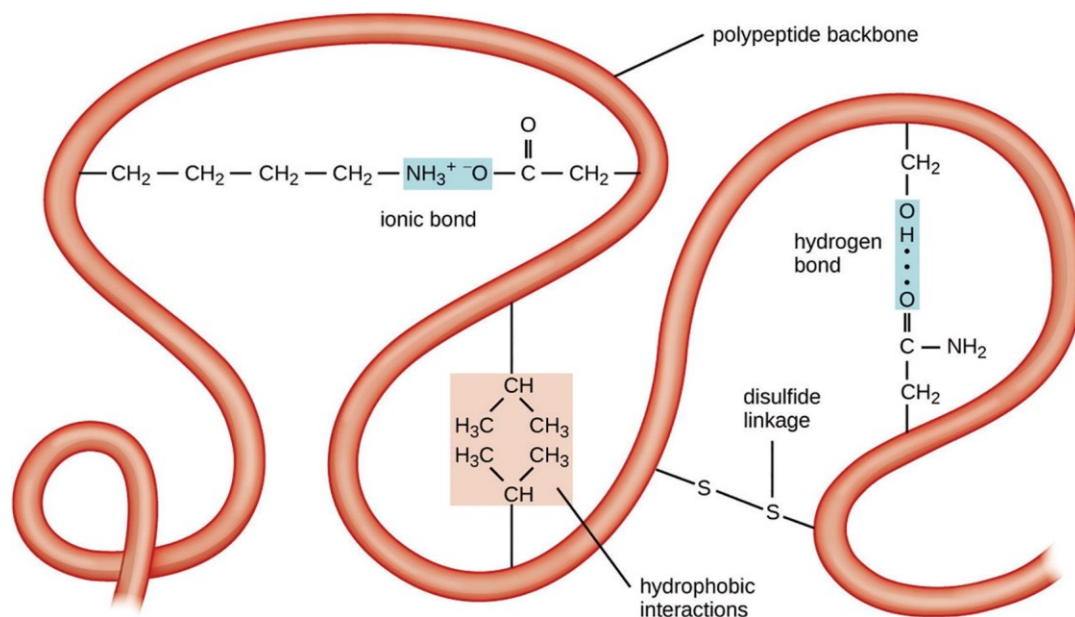


Figure 2. Schematic showing the variety of interactions and chemical bonds that stabilize the tertiary structure of proteins. [21]

Hydrophobic interactions are generally described as the main driving force of protein folding and stability [14]. Hydrophobic effect describes the aggregation process of nonpolar molecules and molecular surfaces in aqueous solutions [24] as a result of hydrophobicity. When hydrophobic residues get closer to each other, they displace the water molecules between them into the bulk solution. The bulk water molecules can move more freely than the interfacial water molecules, which are more ordered at the solute-water interface. Thus, the displacement of water molecules into the bulk leads to increased entropy due to less structured water molecules [24]. This is entropically favourable [25] and contributes aggregation and protein folding. Even though the polypeptide chain becomes more ordered and loses entropy upon folding, for spontaneously folded proteins, this entropy loss is compensated by the increase in solvent (water) entropy and enthalpy gain that resulted from the binding process [26].

Due to the hydrophobic effects, the nonpolar amino acid side chains aggregate and mostly form a protein's internal core which is held together by Van der Waals attractions [1] and hydrogen bonds between the backbone groups [27]. On the other hand, the hydrophilic polar and charged side chains mainly cover the surface of the fold [1]. Therefore, the ratio between polar and nonpolar amino acid side chains greatly impacts the shape of a protein. Figure 3 illustrates the impact of the distribution of hydrophilic and hydrophobic residues to protein folding.

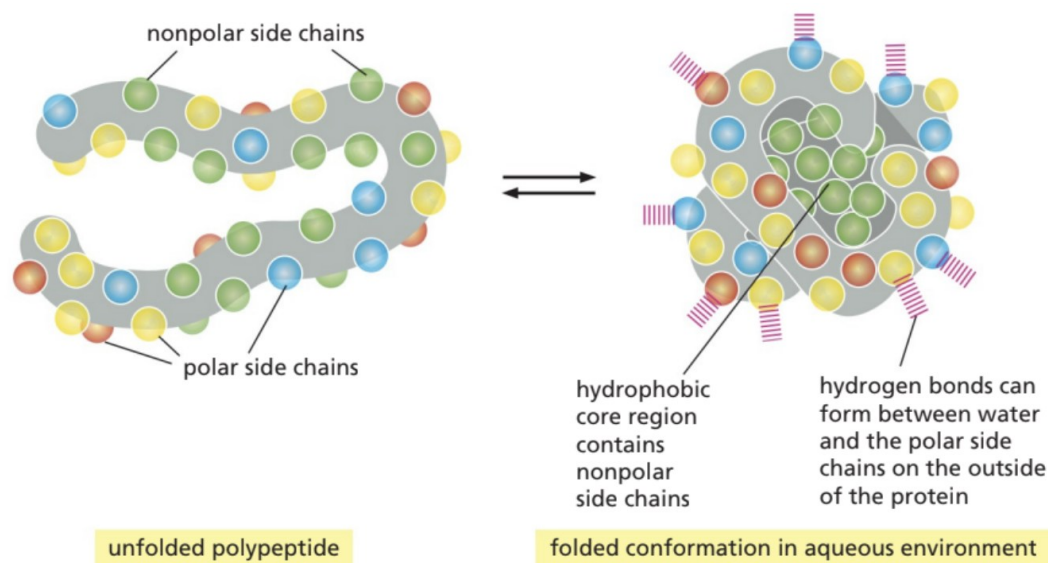


Figure 3. Schematic showing the distribution of the hydrophobic and hydrophilic residues in the sequence and the 3D structure of a folded protein. [1]

In addition to protein folding, noncovalent bonding can further occur between separate tertiary structures, resulting to a quaternary structure. Most cellular proteins form these active multimeric protein complexes [28] of two or more folded proteins, often referred to as “subunits” or “monomers”, with either same or different amino acid sequences [17]. The quaternary structure is critical for the correct function of protein complexes, and it is highly dependent on the chemical structures between the subunit interfaces [17], which participate in protein–protein interactions.

2.2 Protein–protein interactions

About 80% of proteins interact with other proteins by PPIs, forming either binary or higher-order complexes [2, 17]. PPIs play central roles in various biological activities, such as metabolism, enzyme function, muscle contraction, cell signalling, cell cycle, and regulation of other biomolecules [29].

One way to classify PPIs is into permanent and transient PPIs based on the lifetime of the complex [30]. Permanent complexes have strong interactions and are often considered irreversible, whereas transient complexes more readily associate and dissociate, causing an equilibrium [30, 31]. Transient complexes are further divided into weak and strong transient. The weak transient complexes continuously associate and dissociate in solution [30], whereas the strong transient complexes less frequently dissociate or associate when triggered by, for example, a ligand molecule [31]. Ligands are small molecules such as other proteins, nucleic acids or sugars that can bind and

cause conformational changes to a protein [17, 30]. The strength and stability of this protein–ligand interaction can be characterized by binding affinity [32]. More specifically, binding affinity refers to the free energy change associated with the binding process [33]. Experimentally, this is measured as the standard Gibbs free energy of binding (ΔG°) [34].

High binding affinity between the subunit molecules indicates their high tendency to associate into a complex, and on the contrary, low tendency for a complex to dissociate into its subunits. One way of experimentally quantifying the binding affinity of any protein–ligand interaction is by determining the equilibrium dissociation constant (K_D) [35], or its reciprocal equilibrium association constant (K_a) [32]. The K_D can be measured experimentally by, for example, analytical ultracentrifugation [36] or NMR spectroscopy [37]. The K_D and K_a are defined as Equation 1:

$$K_D = \frac{1}{K_a} = \frac{k_{\text{off}}}{k_{\text{on}}} = \frac{[P] \cdot [L]}{[PL]} \quad (1)$$

, where k_{off} is the dissociation rate constant, k_{on} is the association rate constant, $[P]$ is the protein concentration, $[L]$ is the ligand concentration, and $[PL]$ is the concentration of the bound protein–ligand complex [32].

The K_D can be used to calculate the standard Gibbs free energy of binding using Equation 2:

$$\Delta G^\circ = -RT \ln(K_D) \quad (2)$$

, where R is the ideal gas constant, T is temperature, and K_D is the dissociation constant [38].

The lower the K_D value, the slower the dissociation rate, which signifies a stable complex due to greater free energy of binding. For weak transient interactions, K_D is typically in micromolar range, whereas strong transient interactions may have K_D values in nanomolar range. Thus, strong transient complexes have longer lifetimes compared to weak transient complexes, which have lifetimes of seconds [31]. Also, the long-living permanent complexes have their K_D in low nanomolar range [31]. For example, ribosomes belong to permanent complexes [39], and they have a lifespan of 6-7 days in animals according to Mathis *et al.* [40].

Dividing PPIs into distinct groups is challenging, since the boundaries are not explicit, and the stability of a protein complex strongly depends on the physiological conditions and environment [30]. Thus, PPIs may be considered as a continuum between different classes. Figure 4 illustrates the transient-permanent classification of PPIs as a function of continuously increasing binding affinity [31].

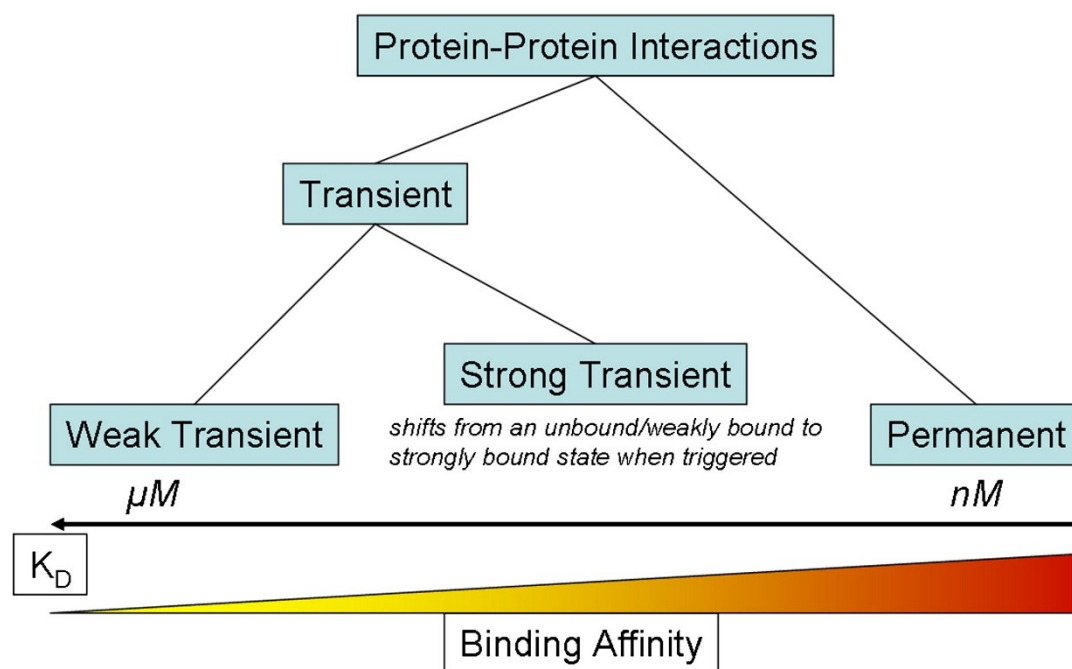


Figure 4. Classification of PPIs into transient and permanent based on their binding affinities. [31]

In addition to binding affinities and lifetimes, transient and permanent complexes differ by their functions and interaction site properties. Transient interaction sites contain more hydrophilic polar residues, whereas permanent interaction sites involve more hydrophobic residues [41]. Additionally, interfaces of transient interactions are typically smaller in size compared to permanent interfaces [41]. Transient PPIs are essential in providing quick responses in various cell signalling chains and biochemical pathways, such as hormone-receptor binding, signal transduction, and correction of misfolded proteins [39]. On the other hand, permanent interactions are required in structures of tighter bonding, such as enzyme-inhibitor, antigen-antibody, and large oligomeric enzyme complex structures [41].

Because most cellular proteins form functional protein complexes with specific conformations, aberrant PPIs can cause diseases such as neurological disorders, infectious diseases, and cancers like leukemia or melanoma [42].

Approximately 645,000 PPIs in the human interactome have been connected to diseases, but by far, drugs have been developed for only a small fraction (2%) of these [43] due to a low number of identified cavities where possible drugs could successfully be targeted to [42]. However, one example of a druggable PPI target in cancer cells is the P53/MDM2 complex involving the oncoprotein MDM2, which deactivates the tumor suppressor protein P53 [42] that can prevent uncontrolled cell division [44]. P53 could be reactivated by manipulating this interaction, which is seen as a potential approach in developing anti-cancer drugs [42]. This example demonstrates why PPI prediction is important and relevant in biological and medicinal research. Both experimental and computational PPI prediction methods offer plenty of possibilities to discover new protein complexes and druggable interaction sites.

Experimental methods used for studying protein complexes either detect direct physical interactions between proteins or determine the overall 3D structure of protein complexes. Experimental methods that detect physical interactions include yeast two-hybrid (Y2H), coimmunoprecipitation (Co-IP), bioluminescence resonance energy transfer (BRET), and affinity purification combined to mass spectrometry (APMS) [4]. On the other hand, experimental methods such as nuclear magnetic resonance (NMR), cryo-electron microscopy (cryo-EM), and X-ray crystallography are used to determine 3D structures of protein complexes [4]. However, computational methods offer faster and cheaper alternatives to the experimental ones. The field of PPI prediction has faced rapid development and breakthroughs after the advent of computational PPI prediction methods.

3 Computational methods for predicting protein–protein interactions

Computational methods model PPIs by using biological data, such as protein sequences or structures [6]. These methods include different molecular docking algorithms, that generally sample multiple binding poses between subunit molecules [45] and rank these using, for example, empirical scoring functions that evaluate the binding affinities among the sampled structures [46]. In other words, docking algorithms involve a sampling component and a scoring component [47]. However, instead of performing sampling and scoring individually, the scoring function can be implemented during and after the sampling, which remarkably improves docking results [48]. Docking algorithms used in PPI prediction can be divided into template-based and *ab-initio* (template-free) docking algorithms depending on if a structure template is used or not [49]. These docking methods are introduced in sections 3.1 and 3.2.

Besides docking algorithms, recently developed computational PPI prediction methods are based on machine learning [6]. In general, machine learning is used to predict the behaviour of unseen test data based on known training data [50]. For PPI prediction, machine learning algorithms are typically trained with data on experimentally determined structures and PPIs of well-studied organisms. For example, the training data input to a machine learning algorithm can be a pair of proteins labelled as interactive or non-interactive [6]. The algorithm tries to identify relevant features of the protein structures in the training data, and the model's performance is benchmarked using the unlabelled test data [6].

Numerous computational methods have been developed for PPI prediction [29]. This thesis introduces *ab-initio* docking, template-based docking, and AlphaFold-Multimer, a novel machine learning-based approach recently developed for PPI prediction. According to multiple literature sources and studies, these computational approaches are frequently applied in PPI prediction. Other computational methods, such as phylogenetic profiling, interlog search methods, or gene co-expression-based methods [29], are outside the scope of this work.

3.1 *Ab-initio* docking

Ab-initio docking predicts structures of protein complexes using known three-dimensional structures of individual unbound proteins [51]. The general steps of *ab-initio* docking algorithms are shown in Figure 5. The step 1 is obtaining the protein 3D structures, which the algorithm uses to perform a

sampling process in step 2. The sampling results to billions of possible protein complexes, where the individual proteins are bound together in different ways [52]. In step 3, the sampled complexes are ranked by scoring, which is a critical step in protein docking [48]. The ranking is typically done with so-called empirical scoring functions [53], which estimate binding affinities for the predicted protein–ligand complex structures [46]. The complex with the highest ranking is considered as the most accurate [54], and therefore it is the final prediction.

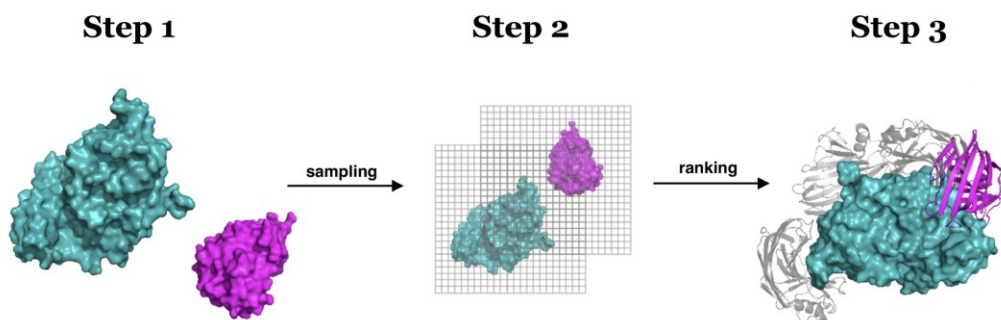


Figure 5. The principle of *ab-initio* docking algorithms. Edited from [52]

Empirical scoring functions include energetic terms, such as van der Waals attractions, hydrophobic effect, hydrogen bonding, and deformation penalty [46, 55]. Deformation penalty refers to the unfavourable entropy change upon binding, because proteins obtain more constrained conformations compared to their free, unbounded states [46].

Improving *ab-initio* methods is challenging without having a large database of complex structures [56, 57], because empirical scoring functions are typically calibrated with experimentally determined protein complex structures and their binding affinities [46], which are used to optimize the weights of energetic terms [55]. One example of an empirical scoring function is Equation 3 developed by Böhm [58]:

$$\Delta G_{bind} = \Delta G_{H-bond} \sum_{H-bond} f(\Delta R, \Delta \alpha) + \Delta G_{ionic} \sum_{ionic} f(\Delta R, \Delta \alpha) + \Delta G_{hydrophobic} \sum_{hydrophobic} |A_{hydrophobic}| + \Delta G_{rotor} N_{rotor} + \Delta G_0 \quad (3)$$

, where the first two terms consider “neutral” and “ionic” hydrogen bonds, the third term considers the hydrophobic effect, the fourth term is related to torsional entropy loss of the ligand, and the fifth term is a regression constant [59]. The calibration of this empirical scoring function was done using 87

protein–ligand complexes and their dissociation constants [59]. More detailed information on the scoring function is provided in references [58-59].

Ab-initio docking yields acceptable predictions for about 40% of the top 10 predictions [52]. Section 4 provides further discussion on how PPI predictions are considered acceptable. Most *ab-initio* algorithms, such as ZDOCK, ROSETTA, and HADDOCK [51, 56, 60], can only predict binary complexes [51]. However, for example, ClusPro can also predict trimers [61]. The prediction of larger complexes by *ab-initio* docking is challenging due to increasing computational costs when the algorithm must calculate a larger energy landscape [56]. To predict larger and more challenging multimeric complexes, template-based docking can be used instead [51].

3.2 Template-based docking

Template-based docking can predict protein complex structures using the amino acid sequences of the unbound target proteins [52]. The general steps of template-based docking algorithms are illustrated in Figure 6. The step 1 is obtaining the target amino acid sequences, which the algorithm uses to search for a known, applicable structure template from the Protein Data Bank [52] in step 2. When choosing the appropriate template, some sequence similarity is required between the target proteins and the proteins in the template [52]. Using the sequence similarity, the target subunit sequences are superimposed and refined onto the template structure [51] in step 3, which results into the final complex structure prediction.

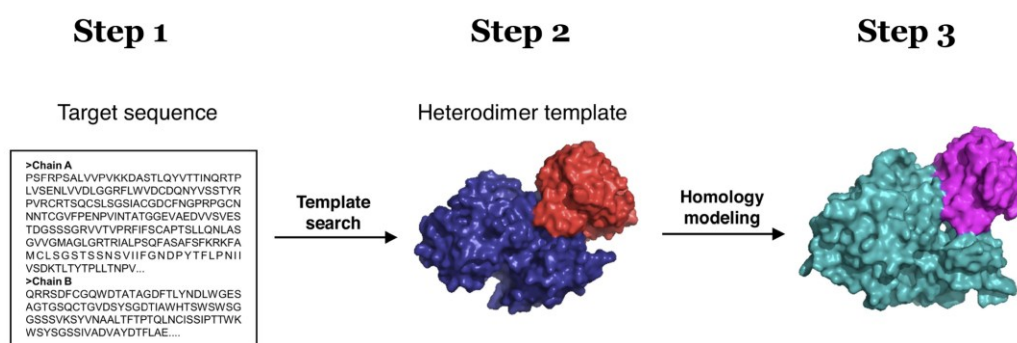


Figure 6. The principle of template-based docking algorithms for an example heterodimer target. Edited from [52]

Template-based docking relies on the evolutionary relationships between the target proteins and the template proteins [49]. Template-based docking assumes that the 3D structures and the binding modes remain similar for

homologous proteins with similar amino acid sequences [49, 60]. Template-based prediction approaches are also referred to as “comparative modelling” or “homology modelling” [56], which are based on the observation that protein pairs most probably interact in the same way if at least 30% of their sequences are identical [62]. On the other hand, if the sequence identity is less than 30%, it is more difficult to say whether the proteins will bind similarly or not [62].

The accuracy and reliability of template-based docking highly depends on the quality and quantity of templates [49]. Good templates are especially available for homo-oligomeric targets [52], because they tend to have conserved binding modes [63]. Homo-oligomers are complexes of two or more identical subunit proteins, and approximately 30-50% of all proteins assemble into these homo-oligomers, which are also referred to as homomers [64]. Indeed, templates are found to model a large majority of known PPIs involving proteins with structural information [51], but only with lenient structural criteria [62]. In general, there exists a lack of available templates that are homologous and reliable enough [51], since good-quality templates are available for less than 20% of all known complexes to date [52].

Negrini *et al.* [62] demonstrated that for three quarters of cases, template-based docking predicts at least one acceptable model in the top 10 models, if a carefully selected good-quality template is used. Additionally, Yan *et al.* [65] demonstrated using a dataset of 54 protein complexes from the Benchmark 4.0 test set by Hwang *et al.* [66] that template-based protein–protein docking had a success rate of 38.9%. The success rate was defined as “the number of cases with at least one correct model divided by the total number of cases in the test set when a specific number of predictions were considered” [65].

3.3 A machine learning-based approach: AlphaFold-Multimer

The field of PPI prediction faced a breakthrough in 2021, when Google DeepMind team released the AlphaFold-Multimer (AF-M) program [67]. AF-M is specifically trained to predict multimeric interfaces between two or more interacting proteins [68]. Currently, AF-M can directly predict protein–protein complexes of up to 20 protein chains [69]. AF-M aligns the inputted target protein sequences into a multiple sequence alignment (MSA) [68]. MSAs can be used to compare and identify similarities between the protein sequences. An example of an MSA is shown in Figure 7.

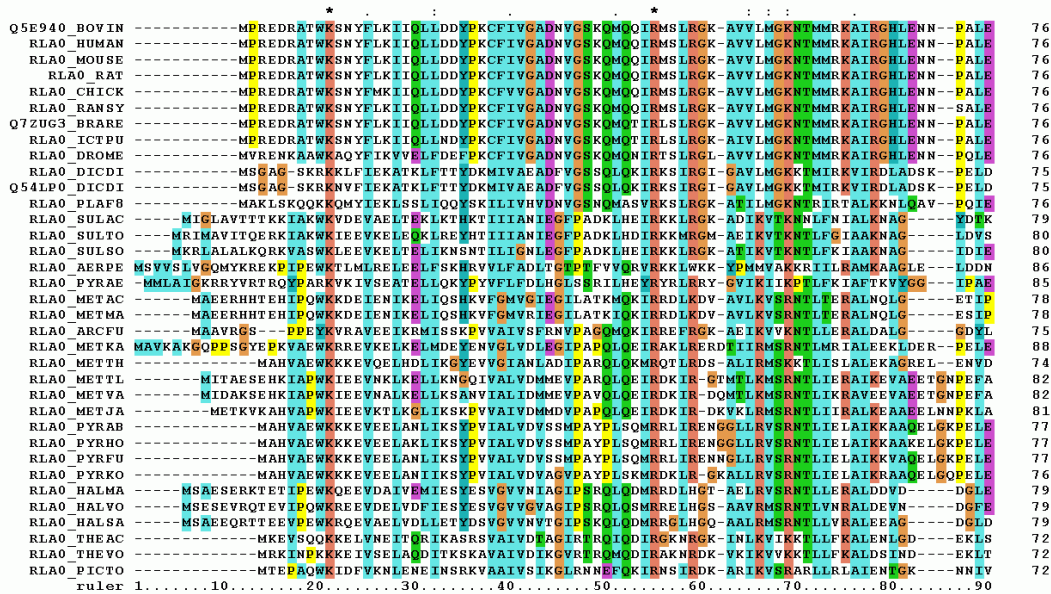


Figure 7. An example of an MSA comprising the first 90 amino acids of the acidic ribosomal protein Po (L10E) from several organisms. The colours indicate similarities between individual amino acid sequences. [70]

The accuracy of AF-M predictions is mostly related to the quality of MSAs [71]. AF-M uses the MSAs of individual protein sequences to compare the correlation of their genetic information [68]. Since AF-M is based on machine learning, it can continuously learn and develop its ability to identify residue conservation and co-evolution directly from the MSAs [68]. However, constructing informative MSAs of multiple protein chains is often challenging for AF-M, because it requires careful sequence pairing in order to uncover information of the interfaces [68].

Along with the predicted interfaces, AlphaFold-Multimer provides two template modelling scores, pTM and ipTM, that estimate model accuracy on a scale from 0 to 100 [72]. The pTM stands for a “predicted” template modelling score, and the ipTM stands for an “interface” pTM [68]. The pTM assesses the average quality of the complex structure prediction, whereas the ipTM more strictly evaluates the accuracy of predicted interfaces only [68, 72]. In practice, the overall confidence of an AF-M model is calculated by combining these pTM- and ipTM-scores into a weighted combination shown in Equation 4:

$$model\ confidence = 0.8 \cdot ipTM + 0.2 \cdot pTM \quad (4)$$

, which takes both inter-chain (ipTM) and intra-chain (pTM) scores into account [68].

Evans *et al.* [68] benchmarked AF-M on two datasets and showed that AF-M yielded best predictions for homomers. AF-M correctly predicted 72% of homomeric interfaces, 36% of which were considered as high-quality predictions. For heteromeric interfaces, the corresponding numbers were 70% and 26%, respectively [68]. This difference in accuracies between homomeric and heteromeric interfaces may result from the tendency of MSAs to more readily encode evolutionary information for homomeric protein–protein interfaces compared to heteromeric interfaces, where the information is more limited [68].

Due to its high accuracy percentages, AF-M may be considered as a breakthrough in the field of PPI prediction. However, as a novel machine learning program, it still requires further improvements to better its performance. For example, Zhu *et al.* [72] showed that AF-M faced some difficulties in protein complex prediction due to AF-M’s limited memory or ability to produce complete MSAs that consider all protein chains. Additionally, Evans *et al.* [68] observed that AF-M could not generally predict the binding of antibodies. Antibodies are immune systems proteins that bind with high affinity to foreign proteins called antigens [73], which participate in various immune functions [74]. Due to their importance, antibody prediction remains as one of the important future work improvements regarding AF-M.

In addition to machine-learning based PPI prediction approaches, such as AlphaFold-Multimer, machine learning can also be applied with beforementioned docking algorithms. It can be used to predict docking scores for millions of ligand molecules and their target proteins by explicitly docking only a smaller subset of molecules [45]. For example, Gentile *et al.* [75] used machine learning to predict docking scores for 1.36 billion molecules against 12 target proteins, using only a few hundred thousand explicitly docked molecules and their docking scores as training data. Due to its speed and scalability, machine learning has the potential to significantly expand our understanding of PPIs and protein complexes.

4 CASP-CAPRI experiments and assessment protocol

The field of PPI prediction is currently developing at a fast rate, and most algorithms and programs are continuously being updated. To improve the field, it is necessary to conduct quality assessment of prediction tools. Typically, evaluations done by the authors of prediction tools tend to be too optimistic due to subjectivity and deep knowledge of the algorithm [76]. To avoid this, objective evaluation criteria and procedures are essential. The state-of-the-art PPI prediction methods are generally evaluated by blind prediction of complex structures, after which the predicted complexes are compared to the known structures. This blind prediction assessment is applied in The Critical Assessment of protein Structure Prediction and the Critical Assessment of Predicted Interactions (CASP-CAPRI experiments) [5].

Community-wide CASP-CAPRI experiments test the performance of protein modelling methods that predict 3D structures of individual proteins (CASP) and protein complexes (CAPRI) by blind prediction [76]. As this thesis focuses on PPIs and protein complexes, only the CAPRI part of the experiment is the focus. The protein complex predictions of CAPRI are compared to unpublished protein complex structures that are experimentally determined by independent researchers and therefore unknown for the CAPRI predictors [77]. The assessment of the CAPRI predictions is done with DockQ scores and the standard CAPRI assessment protocol that includes three quality metrics: f_{nat} , L-RMSD, and i-RMSD_{bb} [5]. These quality metrics are illustrated in Figure 8.

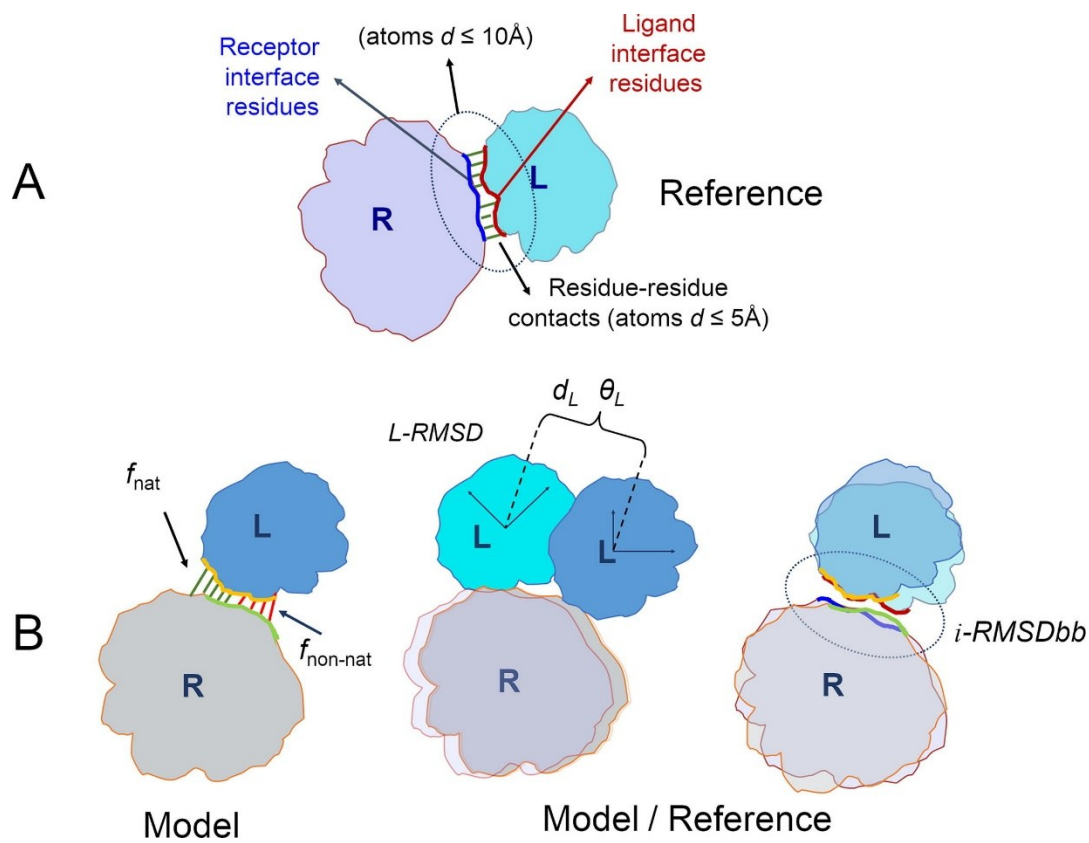


Figure 8. Schematic showing the quality metrics of the CAPRI assessment protocol for protein complex prediction. (A) shows the features of a known reference protein complex involving a receptor, R, and a ligand, L, and (B) shows the quality metrics used for assessing the predictions. [76]

One of the central CAPRI quality metrics are the root-mean-square-deviation (RMSD) values, more explicitly L-RMSD and i-RMSD values. In general, the RMSD values denote the distances between interfacial atoms of the predicted structure and the known reference structure when they are aligned [76]. The L-RMSD values apply to the interface atoms of the ligands only, whereas the i-RMSD values consider interface atoms of both ligand and receptor proteins [76]. Furthermore, i-RMSD values can be divided into i-RMSDbb and i-RMSDsc values. The i-RMSDsc values consider distances of the interfacial side chain atoms, whereas the i-RMSDbb values apply to the interfacial backbone atoms [76].

In addition to the RMSD values, the standard CAPRI assessment protocol uses f_{nat} , which is defined as the fraction of atomic contacts that exist in both the predicted complex structure and in the native reference complex structure [5]. The values of f_{nat} , L-RMSD, and i-RMSDbb are straightforwardly used to qualify CAPRI predictions into categories of acceptable quality,

medium-quality, high-quality, and incorrect [5]. For example, the high-quality CAPRI predictions must have $f_{\text{nat}} \geq 0.5$ and either L-RMSD or i-RMSDbb ≤ 1.0 Å [76]. On the other hand, the medium-quality CAPRI predictions must have $f_{\text{nat}} \geq 0.3$ and either L-RMSD ≤ 5 Å or i-RMSDbb ≤ 2 Å. Furthermore, the acceptable CAPRI predictions must have $f_{\text{nat}} \geq 0.1$ and either L-RMSD ≤ 10 Å or i-RMSDbb ≤ 4 Å [76].

The CAPRI assessment protocol is complemented with DockQ scores [5], which are also used to categorize CAPRI predictions into acceptable, medium-quality, high-quality, and incorrect [76]. The complementary DockQ scores are based on the same RMSD and f_{nat} values that are used in the CAPRI assessment protocol [76]. The DockQ score is a weighted average of the L-RMSD, i-RMSDbb, and f_{nat} shown in Equation 5:

$$\text{DockQ} = \frac{1}{3} \left(f_{\text{nat}} + \frac{1}{1 + \left(\frac{\text{L-RMSD}}{d_1}\right)^2} + \frac{1}{1 + \left(\frac{\text{i-RMSDbb}}{d_2}\right)^2} \right) \quad (5)$$

, where d_1 and d_2 are scaling parameters with respective values of 8.5 Å and 1.5 Å [76].

DockQ scores range from 0 to 1. Predictions with $\text{DockQ} \leq 0.23$ are considered incorrect, whereas the high-quality, medium-quality and acceptable predictions have respective values of $\text{DockQ} \geq 0.80$, $0.49 \leq \text{DockQ} < 0.80$, and $0.23 \leq \text{DockQ} < 0.49$ [76]. For reference, the average DockQ score of the AlphaFold-Multimer program is 0.63 [68], which is within the limits of medium-quality predictions. More explicit information on the assessment done in the CAPRI experiments is provided in references [5, 63].

PPI predictions can be readily assessed using publicly available tools, such as CAPRI-Q [76]. The CAPRI-Q is an enhanced version of the standard CAPRI assessment protocol, as it evaluates PPI predictions by applying the introduced CAPRI assessment protocol, but also other complementary metrics including DockQ scores, TM-scores (see subsection 3.3), and l-DDT tests [76]. The l-DDT test measures the conservation of distances of all heavy atom pairs between the predicted and the reference complex structures [76]. The l-DDT test is complementary to the CAPRI assessment protocol, which uses RMSD values to evaluate only the conservation of interfacial atom distances. More explicit information on the CAPRI-Q tool and its free source code is provided in the reference [76]. The CAPRI-Q can be used to assess any PPI prediction in the same manner as in the assessment of the CASP-CAPRI experiments.

The CASP-CAPRI experiments are concurrently arranged every two years. CASP15-CAPRI experiment in 2022 had a remarkably larger number of targets and predictors compared to previous experiments [5], which reflects increasing interest towards PPI prediction. The CASP15-CAPRI experiment comprised 37 target protein complexes [5], whereas the CASP14-CAPRI experiment involved 12 target complexes [63]. The CASP15-CAPRI showed significant improvements in predicting correct protein complex structures compared to the CASP14-CAPRI experiment due to the development of machine learning-based methods, such as AlphaFold-Multimer, which was used to predict parts of a complex or the whole complex structure [5]. In CASP15-CAPRI, 40% of the predicted final structures were considered as high-quality models, whereas in CASP14-CAPRI, the corresponding percentage was 8% [5]. However, even in CASP15-CAPRI, results were poor for complexes that involved antibodies, nanobodies, or conformational flexibility [5].

Some CAPRI predictor groups have showed that machine learning can be implemented with docking methods to yield better prediction accuracy. For example, Kozakov/Vajda and Seok groups have suggested that inputting machine learning-based protein structure predictions to *ab-initio* docking algorithms could increase the number of predicted interfaces with acceptable or medium accuracy levels [63]. In addition, it has been observed that ranking *ab-initio* docking predictions with a machine learning-based scoring function yielded a 10% increase in the number of near-native predictions [52]. The possibilities of integrating machine learning with other computational methods is worth noticing, but not further discussed in this thesis. The strengths and limitations of the computational PPI prediction methods introduced in this thesis are compiled in Table 1.

Table 1. Summary of the strengths and limitations of the introduced computational methods for predicting folded protein–protein interactions.

	Strengths	Limitations
<i>Ab-initio</i> docking	<ul style="list-style-type: none"> • Can be used to predict completely novel protein complex structures or protein complexes with no available good-quality templates • Yields best predictions especially for heterodimers [52] 	<ul style="list-style-type: none"> • Currently, <i>ab-initio</i> docking is not able to predict complexes that contain more than 3 proteins. • Correct protein 3D structures must be obtained first, which may be time-consuming for newly covered proteins • Calibration of empirical scoring functions requires a large database of known complex structures
Template-based docking	<ul style="list-style-type: none"> • Can be used to predict larger protein complexes compared to <i>ab-initio</i> docking • Performs well if a good-quality homologous template can be found, which is most often the case for homomers 	<ul style="list-style-type: none"> • Limited availability of good-quality templates
AlphaFold-Multimer	<ul style="list-style-type: none"> • Suitable for large-scale PPI prediction and for predicting interfaces of protein complexes involving up to 20 individual protein chains • Achieves nearly equal prediction accuracy for homomers and heteromers (72% vs. 70%) 	<ul style="list-style-type: none"> • Sometimes modelling may fail due to limited memory or out-of-time errors [72] • Low accuracy in predicting antibody complexes • Limitations may occur in predicting complexes that are highly different from the training dataset

5 Summary

This thesis introduced folded protein structures and protein–protein interactions, as well as the state-of-the-art computational methods that are intensively researched and applied in the field of PPI prediction. Experimental PPI prediction methods were also briefly mentioned, but the focus remained on computational prediction of PPIs between ordered, folded proteins. A long-standing limitation of the PPI prediction methods introduced in this thesis is their inability to accurately predict flexible protein complexes with multiple conformational states or disordered regions [5]. While computational methods for predicting both ordered and disordered protein structures are continuously being developed, disorder prediction requires different approaches [78] and further development to predict not only individual protein structures but also their protein complexes [5]. Disordered proteins were not further discussed in this thesis.

To understand the computational modelling of folded PPIs, this thesis began with folded proteins and one way of classifying PPIs into transient and permanent complexes based on binding affinities. Next, the computational PPI prediction methods, including *ab-initio* docking, template-based docking, and a machine learning-based approach AlphaFold-Multimer, were introduced. *Ab-initio* docking relies on experimentally determined 3D protein structures, which may not always be available especially for novel proteins. However, *ab-initio* docking remains useful for dimers and trimers with no suitable structure templates available [52]. If good-quality templates are found, template-based docking can be used to predict larger protein complexes than *ab-initio* docking. However, the quality and the availability of templates need to be improved to retrieve the full potential of template-based docking. Combining template-based and *ab-initio* docking could help overcome the limited availability of good-quality templates [52].

AlphaFold-Multimer can be considered as a breakthrough in the field of PPI prediction, as it has significantly improved prediction accuracies, as demonstrated in, for example, the CASP15-CAPRI experiment. It yields acceptable predictions for most homo- and heteromeric interfaces. However, as is typical for new models, further improvements are still needed. For example, its ability to predict antibody-antigen interactions remains insufficient. Therefore, *ab-initio* and template-based docking currently remain more suitable for modelling complexes involving antibodies [5].

To conclude, the performance of computational PPI prediction methods varies depending on factors such as the number of monomers and whether the complex is homomeric or heteromeric. To improve the accuracy of

computational PPI prediction, it is essential that developers openly report the strengths and limitations of their methods. Moreover, continuous benchmarking and transparent assessment are required for further development of the field of PPI prediction along with its biological and medical applications.

References

1. Alberts, B. et al. *Essential cell biology*. 4th ed. New York, USA: Garland Science, 2007. 863 p. ISBN 978-0-8153-4454-4. P. 121-141.
2. Berggård, T. & Linse, S. & James, P. Methods for the detection and analysis of protein-protein interactions. *Proteomics*. Vol. 7:16. 2007. P. 2833–2842. DOI: 10.1002/pmic.200700131.
3. Li, X. & Han, P. & Wang, G. & Chen, W. & Wang, S., et al. SDNN-PPI: self-attention with deep neural network effect on protein-protein interaction prediction. *BMC Genomics*. Vol. 23. 2022. Art no 474. DOI: 10.1186/s12864-022-08687-2.
4. Durham, J. & Zhang, J. & Humphreys, I. R. & Pei, J. & Cong, Q. Recent advances in predicting and modeling protein-protein interactions. *Trends in Biochemical Sciences*. Vol. 48:6. 2023. P. 527-538. DOI: 10.1016/j.tibs.2023.03.003.
5. Lensink, M.F. & Brysbaert, G. & Raouraoua, N. & Bates, P.A. & Giulini, M. et al. Impact of AlphaFold on structure prediction of protein complexes: The CASP15-CAPRI experiment. *Proteins: Structure, Function, and Bioinformatics*. Vol. 91:12. 2023. P. 1658-1683. DOI: 10.1002/prot.26609.
6. Kewalramani, N. & Emili, A. & Crovella, M. State-of-the-art computational methods to predict protein–protein interactions with high accuracy and coverage. *Proteomics*. Vol. 23:21-22. 2023. Art no 2200292. DOI: 10.1002/pmic.202200292.
7. Ross, T. & Nichols, D. S. ECOLOGY OF BACTERIA AND FUNGI IN FOODS | Influence of Temperature. In: Batt, C. A. & Tortorello, M. L., editors. *Encyclopedia of Food Microbiology*. 2nd ed. Academic Press, 2014. P. 602-609. ISBN: 978-0-12-384733-1. DOI: 10.1016/B978-0-12-384730-0.00087-2.
8. Trivedi, R. & Nagarajaram, H. A. Intrinsically Disordered Proteins: An Overview. *International Journal of Molecular Sciences*. Vol. 23:22. 2022. Art no 14050. DOI: 10.3390/ijms232214050.
9. Dubreuil, B. & Matalon, O. & Levy, E.D. Protein Abundance Biases the Amino Acid Composition of Disordered Regions to Minimize Non-functional Interactions. *Journal of molecular biology*. Vol. 431:24. 2019. P. 4978-4992. DOI: 10.1016/j.jmb.2019.08.008.
10. Kulkarni, P. & Porter, L. & Chou, T. & Chong, S. & Chiti, F. et al. Evolving concepts of the protein universe. *iScience*. Vol. 28:3. 2025. Art no 112012. DOI: 10.1016/j.isci.2025.112012.
11. Chen, R. & Li, X. & Yang, Y. & Song, X. & Wang, C. et al. Prediction of protein-protein interaction sites in intrinsically disordered

- proteins. *Frontiers in Molecular Biosciences*. Vol. 9. 2022. Art no 985022. DOI: 10.3389/fmolb.2022.985022.
12. Kibar, G. & Vingron, M. Prediction of protein–protein interactions using sequences of intrinsically disordered regions. *Proteins: Structure, Function, and Bioinformatics*. Vol. 91:7. 2023. P. 980-990. DOI: 10.1002/prot.26486.
 13. Han, B. & Ren, C. & Wang, W. & Li, J. & Gong, X. Computational Prediction of Protein Intrinsically Disordered Region Related Interactions and Functions. *Genes*. Vol. 14:2. 2023. Art no 432. DOI: 10.3390/genes14020432.
 14. Alaei, L. & Ashengroph, M. & Moosavi-Movahedi, A. A. The concept of protein folding/unfolding and its impacts on human health. *Advances in Protein Chemistry and Structural Biology*. Vol. 126. 2021. P. 227-278. DOI: 10.1016/bs.apcsb.2021.01.007.
 15. Norn, C. & Wicky, B. I. M. & Juergens, D. & Liu, S. & Kim, D. et al. Protein sequence design by conformational landscape optimization. *Proceedings of the National Academy of Sciences*. Vol. 118:11. 2021. Art no 2017228118. DOI: 10.1073/pnas.2017228118.
 16. Smolarczyk, T. & Roterman-Konieczna, I. & Stapor, K. Protein Secondary Structure Prediction: A Review of Progress and Directions. *Current Bioinformatics*. Vol. 15:2. 2020. P. 90-107. DOI: 10.2174/1574893614666191017104639.
 17. Janin, J. & Bahadur, R. P. & Chakrabarti, P. Protein–protein interaction and quaternary structure. *Quarterly Reviews of Biophysics*. Vol. 41:2. 2008. P. 133-180. DOI: 10.1017/S0033583508004708.
 18. Tamang, S. Types of Protein Structure with Diagrams. *Microbe Notes*. [Cited 22 April 2023]. Available at: <https://microbenotes.com/protein-structure/>.
 19. Roy, K. & Kar, S. & Das, R. N. Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment. Academic Press, 2015. Chapter 1, Background of QSAR and Historical Developments, p. 1-46. ISBN 978-0-12-801505-6. DOI: 10.1016/B978-0-12-801505-6.00001-6.
 20. Ji, Y. & Li, Y. The role of secondary structure in protein structure selection. *The European physical journal*. Vol. 32. 2010. P. 103-107. DOI: 10.1140/epje/i2010-10591-5.
 21. Soleymani, F. & Paquet, E. & Viktor, H. & Michalowski, W. & Spinello, D. Protein–protein interaction prediction with deep learning: A comprehensive review. *Computational and structural biotechnology journal*. Vol. 20. 2022. P. 5316-5341.

22. Kang, R. & Chen, R. & Zhang, Q. & Hou, W. & Wu, S. et al. HMGB1 in health and disease. *Molecular Aspects of Medicine*. Vol. 40. 2014. P. 1-116. DOI: 10.1016/j.mam.2014.05.001.
23. Stauffer, E. & Dolan, J. & Newman, R. *Fire Debris Analysis*. Academic Press, 2008. Chapter 3, Review of Basic Organic Chemistry, p. 49-83. ISBN: 9780126639711.
24. Sun, Q. & Fu, Y. & Wang, W. Temperature effects on hydrophobic interactions: Implications for protein unfolding. *Chemical Physics*. Vol. 559. 2022. Art no 111550. DOI: 10.1016/j.chemphys.2022.111550.
25. Dédinaite, A. & Wieland, D.C. & Beldowski, P. & Claesson, P.M. Biolubrication synergy: Hyaluronan – Phospholipid interactions at interfaces. *Advances in colloid and interface science*. Vol. 274. 2019. Art no 102050. DOI: 10.1016/j.cis.2019.102050.
26. Dyakin, V.V. & Uversky, V.N. Arrow of Time, Entropy, and Protein Folding: Holistic View on Biochirality. *International Journal of Molecular Sciences*. Vol. 23:7. 2022. Art no 3687. DOI: 10.3390/ijms23073687.
27. Schell, D. & Tsai, J. & Scholtz, M. & Pace, C.N. Hydrogen bonding increases packing density in the protein interior. *Proteins: Structure, Function, and Bioinformatics*. Vol. 63:2. 2006. P. 278-282. DOI: 10.1002/prot.20826.
28. Zabel, W.J. & Hagner, K.P. & Livesey, B.J. & Marsh, J.A. & Setayeshgar, S. Evolution of protein interfaces in multimers and fibrils. *The Journal of Chemical Physics*. Vol. 150. 2019. Art no 225102. DOI: 10.1063/1.5086042.
29. Poluri, K.M. & Gulati, K. & Sarkar, S. *Protein-Protein Interactions: Principles and Techniques: Volume I*. Gateway East, Singapore: Springer, 2021. 346 p. ISBN 978-981-16-1594-8. Available at: <https://link-springer-com.libproxy.aalto.fi/book/10.1007/978-981-16-1594-8>.
30. Nooren, I.M.A. & Thornton, J.M. Diversity of protein-protein interactions. *The EMBO Journal*. Vol. 22. 2003. P. 3486-3492. DOI: 10.1093/emboj/cdg359.
31. Perkins, J.R. & Diboun, I. & Dessailly, B.H. & Lees, J.G. & Orenco, C. et al., *Transient Protein-Protein Interactions: Structural, Functional, and Network Properties*. *Structure*. Vol. 18:10. 2010. P. 1233-1243. DOI: 10.1016/j.str.2010.08.007.
32. Spassov, D.S. *Binding Affinity Determination in Drug Design: Insights from Lock and Key, Induced Fit, Conformational Selection,*

and Inhibitor Trapping Models. *International Journal of Molecular Sciences*. Vol. 25:13. 2024. Art no 7124. DOI: 10.3390/ijms25137124.

33. Wan, S. & Bhati, A.P. & Zasada, S.J. & Coveney, P.V. Rapid, accurate, precise and reproducible ligand–protein binding free energy prediction. *Interface Focus*. Vol. 10. 2020. Art no 20200007. DOI: 10.1098/rsfs.2020.0007.
34. Gallicchio, E. & Levy, R.M. Recent theoretical and computational advances for modeling protein–ligand binding affinities. *Advances in Protein Chemistry and Structural Biology*. Vol. 85. 2011. p. 27–80. DOI: 10.1016/b978-0-12-386485-7.00002-8.
35. Housden, N.G. & Kleanthous, C. Thermodynamic Dissection of Colicin Interactions. *Methods in Enzymology*. Vol. 488. 2011. P. 123–145. DOI: 10.1016/B978-0-12-381268-1.00006-9.
36. Fedorov, D. & Batys, P. & Hayes, D.B. & Sammalkorpi, M. & Linder, M.B. Analyzing the weak dimerization of a cellulose binding module by analytical ultracentrifugation. *International Journal of Biological Macromolecules*. Vol. 163. 2020. P. 1995–2004. DOI: 10.1016/j.ijbiomac.2020.09.054.
37. Furukawa, A. & Konuma, T. & Yanaka, S. & Sugase, K. Quantitative analysis of protein–ligand interactions by NMR. *Progress in Nuclear Magnetic Resonance Spectroscopy*. Vol. 96. 2016. P. 47–57. DOI: 10.1016/j.pnmrs.2016.02.002.
38. Khan, M.D. & Khan, N. & Sultana, S. & Joshi, R. & Ahmed, S. & Yu, E. et al. Bioelectrochemical conversion of waste to energy using microbial fuel cell technology. *Process Biochemistry*. Vol. 57. 2017. P. 141–158. DOI: 10.1016/j.procbio.2017.04.001.
39. Ozbabacan, S.E. & Engin, H.B. & GURSOY, A. & Keskin, O. Transient protein–protein interactions. *Protein Engineering, Design and Selection*. Vol. 24:9. 2011. P. 635–648. DOI: 10.1093/protein/gzr025
40. Mathis, A.D. & Naylor, B.C. & Carson, R.H. & Evans, E. & Harwell, J. & Knecht, J. et al. Mechanisms of *In Vivo* Ribosome Maintenance Change in Response to Nutrient Signals. *Molecular & cellular proteomics*. Vol. 16:2. 2017. P. 243–254. DOI: 10.1074/mcp.M116.063255.
41. La, D. & Kong, M. & Hoffman, W. & Choi, Y.I. & Kihara, D. Predicting permanent and transient protein–protein interfaces. *Proteins: Structure, Function, and Bioinformatics*. Vol. 81:5. 2013. P. 805–818. DOI: 10.1002/prot.24235.
42. Akbarzadeh, S. & Coşkun, Ö. & Günçer, B. Studying protein–protein interactions: Latest and most popular approaches. *Journal of*

- structural biology. Vol. 216:4. 2024. P. 108118. DOI: 10.1016/j.jsb.2024.108118.
43. Heifetz, A. & Sladek, V. & Townsend-Nicholson, A. & Fedorov, D.G. Characterizing Protein-Protein Interactions with the Fragment Molecular Orbital Method. *Methods in Molecular Biology*. Vol. 2114. 2020. P. 187-205. DOI: 10.1007/978-1-0716-0282-9_13.
 44. Ozaki, T. & Nakagawara, A. Role of p53 in Cell Death and Human Cancers. *Cancers*. Vol. 11:3. 2011. P. 994-1013. DOI: 10.3390/cancers3010994.
 45. Bande, A.Y. & Baday, S. Accelerating Molecular Docking using Machine Learning Methods. *Molecular Informatics*. Vol. 43:6. 2024. Art no e202300167. DOI: 10.1002/minf.202300167.
 46. Wang, R. & Luhua, L. & Wang, S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *Journal of Computer-Aided Molecular Design*. Vol. 16. 2002. P. 11-26. DOI: 10.1023/a:1016357811882.
 47. Zheng, M. & Liu, X. & Xu, Y. & Li, H. & Luo, C. et al. Computational methods for drug design and discovery: focus on China. *Trends in Pharmacological Sciences*. Vol. 34:10. 2013. P. 549-559. DOI: 10.1016/j.tips.2013.08.004.
 48. Vajda, S. & Hall, D.R. & Kozakov, D. Sampling and scoring: A marriage made in heaven. *Proteins: Structure, Function, and Bioinformatics*. Vol. 81:11. 2013. P. 1874-1884. DOI: 10.1002/prot.24343.
 49. Meng, Q. & Guo, F. & Wang, E. & Tang, J. ComDock: A novel approach for protein-protein docking with an efficient fusing strategy. *Computers in biology and medicine*. Vol. 167. 2023. Art no 107660. DOI: 10.1016/j.combiomed.2023.107660.
 50. Wang, Q. & Cai, C. *Machine Learning in Chemical Safety and Health : Fundamentals with Applications*. 1st ed. West Sussex, UK: John Wiley & Sons, 2022. 320 p. ISBN: 978-1119817482.
 51. Rosell, M. & Fernández-Recio, J. Docking approaches for modeling multi-molecular assemblies. *Current Opinion in Structural Biology*. Vol. 64. 2020. P. 59-65. DOI: 10.1016/j.sbi.2020.05.016.
 52. Porter, K.A. & Desta, I. & Kozakov, D. & Vajda, S. What method to use for protein-protein docking? *Current opinion in structural biology*. Vol. 55. 2019. P. 1-7. DOI: 10.1016/j.sbi.2018.12.010.
 53. Wang, J. & Lin, J. Scoring Functions for Prediction of Protein-Ligand Interactions. *Current Pharmaceutical Design*. Vol. 19:12. 2013. P. 2174-2182.

54. Zhang, W. & Meng, Q. & Tang, J. & Guo, F. Exploring effectiveness of *ab-initio* protein–protein docking methods on a novel antibacterial protein complex dataset. *Briefings in Bioinformatics*. Vol. 22:6. 2021. Art no bbab150. DOI: 10.1093/bib/bbab150.
55. Li, J. & Fu, A. & Zhang, L. An Overview of Scoring Functions Used for Protein–Ligand Interactions in Molecular Docking. *Interdisciplinary Sciences: Computational Life Sciences*. Vol. 11. 2019. P. 320-328. DOI: 10.1007/s12539-019-00327-w.
56. Wang, L. & Wen, Z. & Liu, S. & Zhang, L. & Finley, C. et al. Overview of AlphaFold2 and breakthroughs in overcoming its limitations. *Computers in biology and medicine*. Vol. 176. 2024. Art no 108620. DOI: 10.1016/j.compbimed.2024.108620.
57. Vreven, T. & Schweppe, D.K. & Chavez, J.D. & Weisbrod, C.R. & Shibata, S. et al. Integrating Cross-Linking Experiments with *Ab Initio* Protein–Protein Docking. *Journal of molecular biology*. Vol. 430:12. 2018. P. 1814-1828. DOI: 10.1016/j.jmb.2018.04.010.
58. Böhm, H.J. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *Journal of Computer-Aided Molecular Design*. Vol. 8. 1994. P. 243-256. DOI: 10.1007/bf00126743.
59. Wang, R. & Lu, Y. & Wang, S. Comparative Evaluation of 11 Scoring Functions for Molecular Docking. *Journal of Medicinal Chemistry*. Vol. 46:12. 2003. P. 2287-2303. DOI: 10.1021/jm0203783.
60. Vangaveti, S. & Vreven, T. & Zhang, Y. & Weng, Z. Integrating *ab initio* and template-based algorithms for protein–protein complex structure prediction. *Bioinformatics*. Vol. 36:3. 2020. P. 751-757. DOI: 10.1093/bioinformatics/btz623.
61. Kozakov, D. & Hall, D.R. & Xia, B. & Porter, K.A. & Padhorny, D. et al. The ClusPro web server for protein–protein docking. *Nature Protocols*. Vol. 12. 2017. P. 255-278. DOI: 10.1038/nprot.2016.169.
62. Negroni, J. & Mosca, R. & Aloy, P. Assessing the Applicability of Template-Based Protein Docking in the Twilight Zone. *Structure*. Vol- 22:9. 2014. P. 1356-1362. DOI: 10.1016/j.str.2014.07.009.
63. Lensink, M.F. & Brysbaert, G. & Mauri, T. & Nadzirin, N. & Velankar, S. et al. Prediction of protein assemblies, the next frontier: The CASP14-CAPRI experiment. *Proteins: Structure, Function, and Bioinformatics*. Vol. 89:12. 2021. P. 1800-1823. DOI: 10.1002/prot.26222
64. Gaber, A. & Pavšič, M. Modeling and Structure Determination of Homo-Oligomeric Proteins: An Overview of Challenges and Current

- Approaches. *International Journal of Molecular Sciences*. Vol. 22:16. 2021. Art no 9081. DOI: 10.3390/ijms22169081.
65. Yan, Y. & Zhang, D. & Zhou, P. & Li, B. & Huang, S. HDOCK: a web server for protein–protein and protein–DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Research*. Vol. 45:W1. 2017. P. W365-W373. DOI: 10.1093/nar/gkx407.
66. Hwang, H. & Vreven, T. & Janin, J. & Weng, Z. Protein–protein docking benchmark version 4.0. *Proteins: Structure, Function, and Bioinformatics*. Vol. 78:15. 2010. P. 3111-3114.
67. 64. Grassmann, G. & Miotto, M. & Desantis, F. & Rienzo, L.D. & Tartaglia, G.G. et al. Computational Approaches to Predict Protein–Protein Interactions in Crowded Cellular Environments. *Chemical reviews*. Vol. 124. 2024. P. 3932-3977. DOI: 10.1021/acs.chemrev.3c00550.
68. Evans, R. & O'Neill, M. & Pritzel, A. & Antropova, N. & Senior, A. et al. Protein complex prediction with AlphaFold-Multimer. *bioRxiv (Cold Spring Harbor Laboratory)*. Version 2. 2022. DOI: 10.1101/2021.10.04.463034.
69. Kovalevskiy, O. & Mateos-Garcia, J. & Tunyasuvunakool, K. AlphaFold two years on: Validation and impact. *Proceedings of the National Academy of Sciences*. Vol. 121:34. 2024. Art no e2315002121. DOI: 10.1073/pnas.2315002121.
70. Multiple sequence alignment. Wikipedia contributors. Cited 23 Apr 2025 .. Available at: https://en.wikipedia.org/w/index.php?title=Multiple_sequence_alignment&oldid=1245838134.
71. Liu, J. & Guo, Z. & Wu, T. & Roy, R.S. & Quadir, F. et al. Enhancing alphafold-multimer-based protein complex structure prediction with MULTICOM in CASP15. *Communications Biology*. Vol. 6. 2023. Art no 1140. DOI: 10.1038/s42003-023-05525-3.
72. Zhu, W. & Shenoy, A. & Kundrotas, P. & Elofsson, A. Evaluation of AlphaFold-Multimer prediction on multi-chain protein complexes. *Bioinformatics*. Vol. 39:7. 2023. Art no btad424. DOI: 10.1093/bioinformatics/btad424.
73. Carter, M. & Essner, R. & Goldstein, N. & Iyer, M. *Guide to Research Techniques in Neuroscience*. 3rd ed. Amsterdam: Academic Press, 2022. Chapter 14, Intracellular Signaling and Biochemical Assays. P. 309-347. ISBN 978-0-12-818646-6. DOI: 10.1016/B978-0-12-818646-6.00001-4.
74. Daëron, M. The function of antibodies. *Immunological Reviews*. Vol. 328:1. 2024. P. 113-125. DOI: 10.1111/imr.13387.

75. Gentile, F. & Agrawal, V. & Hsing, M. & Ton, A. & Ban, F. et al. Deep Docking: A Deep Learning Platform for Augmentation of Structure Based Drug Discovery. *ACS Central Science*. Vol. 6:6. 2020. P. 939-949. DOI: 10.1021/acscentsci.0c00229.
76. Collins, K.W. & Copeland, M.M. & Brysbaert, G. & Wodak, S.J. & Bonvin, A.M.J.J. CAPRI-Q: The CAPRI resource evaluating the quality of predicted structures of protein complexes. *Journal of molecular biology*. Vol. 436:17. 2024. Art no 168540. DOI: 10.1016/j.jmb.2024.168540.
77. Kryshchuk, A. Schwede, T. & Topf, M. & Fidelis, K. & Moult, J. Critical assessment of methods of protein structure prediction (CASP)—Round XIII. *Proteins: Structure, Function, and Bioinformatics*. Vol. 87:12. 2019. P. 1011-1020. DOI: 10.1002/prot.25823.
78. Kurgan, L. Resources for computational prediction of intrinsic disorder in proteins. *Methods*. Vol. 204. 2022. P. 132-141. DOI: 10.1016/j.ymeth.2022.03.018.