



## Review

## Coarse-grained (hybrid) integrative modeling of biomolecular interactions

Jorge Roel-Touris<sup>1</sup>, Alexandre M.J.J. Bonvin<sup>\*,1</sup>

Bijvoet Centre for Biomolecular Research, Faculty of Science – Chemistry, Utrecht University, Utrecht 3584CH, the Netherlands

## ARTICLE INFO

## Article history:

Received 27 February 2020  
Received in revised form 23 April 2020  
Accepted 6 May 2020  
Available online 15 May 2020

Dedicated to the 8th anniversary of CSBJ journal upon invitation from Dr. Gianni Panagiotou.

## Keywords:

Complexes  
Docking  
Molecular representations  
Force field  
Software

## ABSTRACT

The computational modeling field has vastly evolved over the past decades. The early developments of simplified protein systems represented a stepping stone towards establishing more efficient approaches to sample intricate conformational landscapes. Downscaling the level of resolution of biomolecules to coarser representations allows for studying protein structure, dynamics and interactions that are not accessible by classical atomistic approaches. The combination of different resolutions, namely hybrid modeling, has also been proved as an alternative when mixed levels of details are required. In this review, we provide an overview of coarse-grained/hybrid models focusing on their applicability in the modeling of biomolecular interactions. We give a detailed list of ready-to-use modeling software for studying biomolecular interactions allowing various levels of coarse-graining and provide examples of complexes determined by integrative coarse-grained/hybrid approaches in combination with experimental information.

© 2020 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## Contents

1. Introduction	1182
2. Historical perspective	1183
3. Coarse-grained/Hybrid modeling of biomolecular interactions	1183
3.1. Simplified representations and topologies	1183
3.2. Parametrization of coarse-grained force fields	1184
3.2.1. Classical parametrization strategies	1184
3.2.2. Machine learning-based parametrization	1185
3.3. Combining different levels of resolution	1186
3.4. Sampling and scoring schemes	1186
3.5. Backmapping from coarse-grained to atomistic resolution	1186
4. Application examples of integrative modeling of protein interactions	1187
5. Concluding remarks	1188
Declaration of Competing Interest	1188
Acknowledgments	1188
Appendix A. Supplementary data	1188
References	1188

\* Correspondence author.

E-mail address: [a.m.j.j.bonvin@uu.nl](mailto:a.m.j.j.bonvin@uu.nl) (A.M.J.J. Bonvin).<sup>1</sup> Both authors have devised the content and contributed to the writing.

## 1. Introduction

The chemistry that supports life is extremely sophisticated. Despite advances over the past decades, the scientific community still lacks fundamental knowledge to fully understand the biology behind the cell at atomic level. We know that basic subunit atoms (i.e. carbon, oxygen, hydrogen and nitrogen) can combine and form complex molecules such as lipids, carbohydrates, nucleic acids and proteins. At the same time, these biomolecules associate and create more intricate assemblies that adopt specific three-dimensional (3D) structures, essential for their biological functions. Their interactions mediate a wide range of biological functions such as for example signal transduction, molecular recognition or transport. Indeed, roughly 80% of the proteins might function upon association with other biomolecules [1]. It is therefore of great importance to understand how these macromolecules interact. Next to experimental methods, complementary computational approaches have been developed with the so-called integrative modeling emerging as the most promising strategy [2]. In short, integrative modeling aims at obtaining structural insights into a given system under study that cannot be revealed by a single approach alone. To do so, it combines data from multiple information sources (e.g. nuclear magnetic resonance (NMR) spectroscopy, cryo-electron microscopy (cryo-EM), mass spectrometry (MS), small angle x-ray scattering (SAXS), bioinformatics analysis. . .) [3] into computational approaches to model the assemblies. Integrative modelling has been extensively used to model increasingly larger systems in the recent past [4]. In this sense, we are probably closer than ever to construct a predictive model of an entire cell [5].

Classical atomistic computational modeling of interactions remains inefficient for many molecular assemblies. Larger systems often require longer simulations and their complex conformational landscapes cannot be efficiently and thoroughly sampled by atomistic approaches. The simplification of large systems to coarser representations offers a valuable approach to alleviate those limitations. There is already a huge body of literature on this topic and, in the present work, we do not aspire to give the most comprehensive review covering all possible contributions, but will focus on the modeling of biomolecular interactions. i.e. complexes, involving proteins, peptides and nucleic acids (DNA and RNA). The remaining of the text is organized as follows: We first start with a brief historical overview of the development of coarse-graining. We then describe several representative designs of simplified systems and parametrization strategies and discuss how these can be implemented into the modeling of biomolecular complexes, both for the generation of possible conformations (sampling) and the discrimination between native and non-native models (scoring). Finally, we provide an overview of currently available software that support coarse-grained modeling of biomolecular complexes and highlight several representative applications.

## 2. Historical perspective

The structural characterization of lysozyme in 1967 [6] spurred Arieh Warshel to study enzymatic reaction mechanisms. His developments in this field under the supervision of Martin Karplus, inaugurated the now well-established quantum mechanics/molecular mechanics (QM/MM) methods [7]. In parallel, Michael Levitt, a PhD student at the Medical Research Council at that time, was making significant advances for studying molecular conformations by computational approaches: Together with Shneior Lifson in 1972 at the Weizmann Institute in Israel, Levitt and Warshel started working on a simplified representation of a protein, where spheres would represent amino acids. In fact, this project, later on in 1975, turned out in the very first computer simulation of a pro-

tein system (pancreatic trypsin inhibitor) using a coarse-grained model [8]. These simulations suggested that the protein folding process has a relatively small number of conformations, and challenged the so-called “Levinthal paradox” [9]. In this work, each residue was represented by only two beads: The C $\alpha$  atom and the centroid of the side chain. Non-bonded interactions were assumed to occur only between side chains. By doing so, only torsion angles between 4 consecutive C $\alpha$  atoms were considered, considerably reducing the conformational space (one degree of freedom per residue). For all these premature findings Karplus, Levitt and Warshel were awarded with the Nobel Prize in Chemistry in 2013.

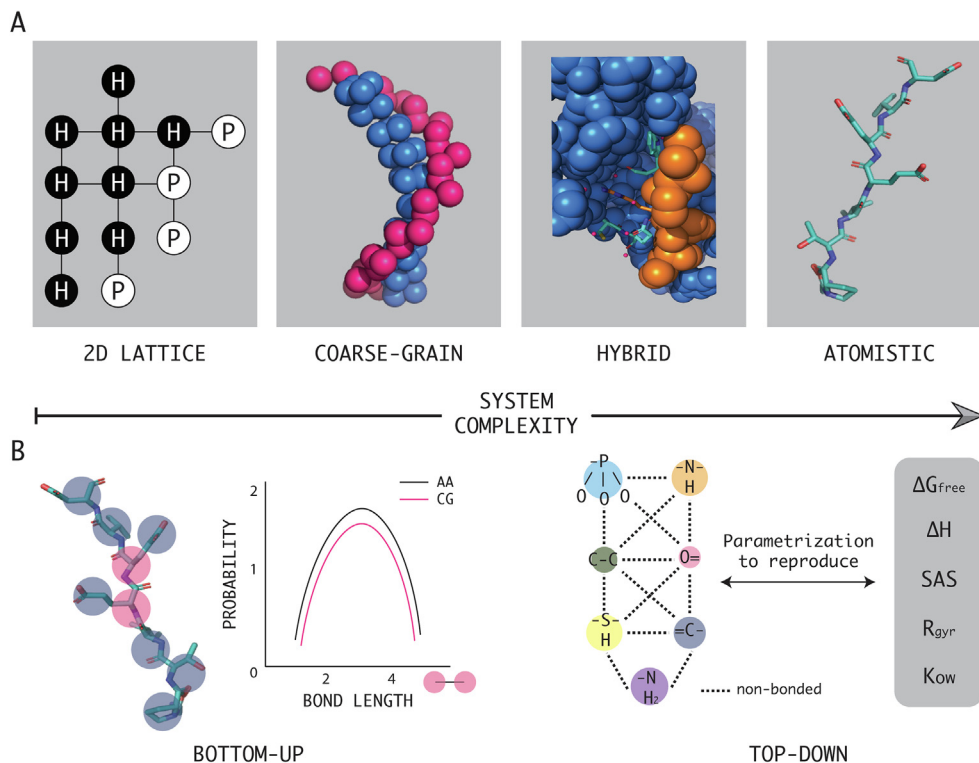
In 1975, Chothia and Janin established the structural basis of the hydrophobic effect as fundamental to the stabilization of protein association [10]. All these pioneering findings were used as a basis for the first computational analysis of a protein–protein complex: In 1978, Wodak and Janin studied the association of BPTI and trypsin using a coarse-grained representation of the system [11]. They used a combination of a simple averaged potential energy function including non-bonded (van der Waals) and residue-solvent interactions. Whilst encouraging, this early model totally neglected electrostatic interactions and was thus unable to describe hydrogen bonds and salt bridges, which, later on in 1984, were suggested to provide the specificity of the association [12]. In spite of the incompleteness of this work, they shed light on the idea that a simplified protein model could be an effective alternative to screen a relatively large number of possible interfaces, which constituted the first coarse-grained docking simulation. Ever since, coarse-grained/hybrid modeling approaches have gained importance in the computational structural biology field [13] and have become central in the study of folding, dynamics and association mechanisms of biomolecules.

## 3. Coarse-grained/Hybrid modeling of biomolecular interactions

In this section, we will focus on macromolecular docking approaches allowing some level of coarse-grained/hybrid representations for the modeling of interactions. These usually include two different steps: The generation of possible complex conformations, referred to as sampling, and the discrimination between biologically and non-biologically relevant models referred to as scoring. The latter might also be an integral part of the sampling process, especially when experimental or predicted information is included to bias the sampling (e.g. restraints-driven sampling). We first describe various strategies to simplify the representation of polypeptides and nucleic acids and discuss existing parametrization strategies and force fields. We then focus on how coarse-grained/hybrid approaches can be applied during the sampling and scoring steps for modeling biomolecular interactions and end with a short discussion of backmapping approaches to restore full atomistic representations.

### 3.1. Simplified representations and topologies

In general, a coarse-grained model aims at decreasing the complexity of a system by grouping several atoms into larger “pseudo-atoms” or “beads”, thereby reducing the number of degrees of freedom. This results both in more efficient computations and a possible smoothening of the energy landscape that might facilitate the identification of relevant states of the system. In the context of proteins, the simplest models introduced are the hydrophobic/polar (HP) models (see Fig. 1). These simplify the representations of a polypeptide chain [14] by considering only two type of beads (H and P), which, to some extent, are an approximation of two types



**Fig. 1.** Examples of various coarse-grained models. (A) The panels from left to right illustrate the increase in the complexity of the system (i.e. decreased coarse-graining): A 2-D lattice representation of a HP model, a coarse-grained (4:1 mapping) of a dsDNA molecule, a hybrid representation of a protein–protein interface (AA/CG) and an atomistic model of a peptide. (B) The two traditional parametrization strategies. *Bottom-up*: Bond-lengths are parametrized by mapping to distributions of reference atomistic simulations. *Top-down*: Models are designed to match specific properties (e.g. thermodynamic quantities) of the system.

of residues: hydrophobic (H) and polar (P) [15]. Albeit very minimalistic, HP representations have proven useful to study larger conformational changes and longer time scales. These models, and their variants, have been extensively studied in the past decade [16–19] and reviewed elsewhere [20]. Another example of a low-resolution model to represent proteins is SICH0 (Side Chain Only) [21]. In the model developed by Kolinski and Skolnick [21], each amino acid is represented as a unique interaction site, located at the center of the side-chain. It is thus computationally very efficient but completely neglects backbone conformations ( $\varphi/\psi$  dihedrals) [22].

In order to overcome the inaccuracies of very simplistic representations, higher resolution models have been developed. PRIMO/PRIMONA, for proteins and nucleic acids, was proposed as a reduced quasi-atomistic resolution model [23]. Feig and co-workers [23] represent polypeptide backbones with three beads ( $C\alpha$ , N and a combined carbonyl site) and side-chains as a combination of up to five different particles. In the case of nucleotides, adenine, cytosine and uracil are represented by four coarse-grained particles, and guanine and thymine by five. The sugar-phosphate backbone of the PRIMONA model consist of eight different CG beads. In contrast, the HiRE-RNA model designed by Pasquali and Derremaux [24] only considers three of the seven backbone torsional angles ( $\alpha$ ,  $\beta$  and  $\gamma$ ); each RNA nucleotide is represented by six (pyrimidine bases) or seven (purine bases) beads, allowing for a reduction of ~70% of the number of particles compared to a fully atomistic structure. Similar to PRIMO, in the SIRAH model [25] the positions of the nitrogen,  $\alpha$  carbon and oxygen from the peptide bonds are kept at pseudo-atomistic resolution, while side chains are treated at a lower degree of detail (from one to five different beads). This model also allows for the study of protein-DNA interactions by molecular dynamics through the use of an explicit/CG solvation scheme [26,27].

Other coarse-grained models have been designed to be easily transferable and applicable to multiple systems. Among those, MARTINI is probably the most popular one. The current “MARTINI dome” includes: lipids [28], proteins [29], polymers [30,31], carbohydrate [32], water [33], glycolipids [34], nucleotides [35,36] and nanoparticles [37]. The systems are represented by four different basic particles – nonpolar (N), polar (P), apolar (C) and charged (Q) – that are further classified based on their degree of polarity and hydrogen bonding properties, giving a total of eighteen unique “building blocks”. The MARTINI force field for proteins, in its latest official release (2.2p), includes off-center charges for polar and charged residues [38]. These represent a good proxy for hydrogen bond and salt bridges formation and thus for molecular recognition. For nucleic acids, much like PRIMONA, the MARTINI model specifically accounts for Watson-Crick base pairing (eight additional beads) to stabilize the DNA double helix structure.

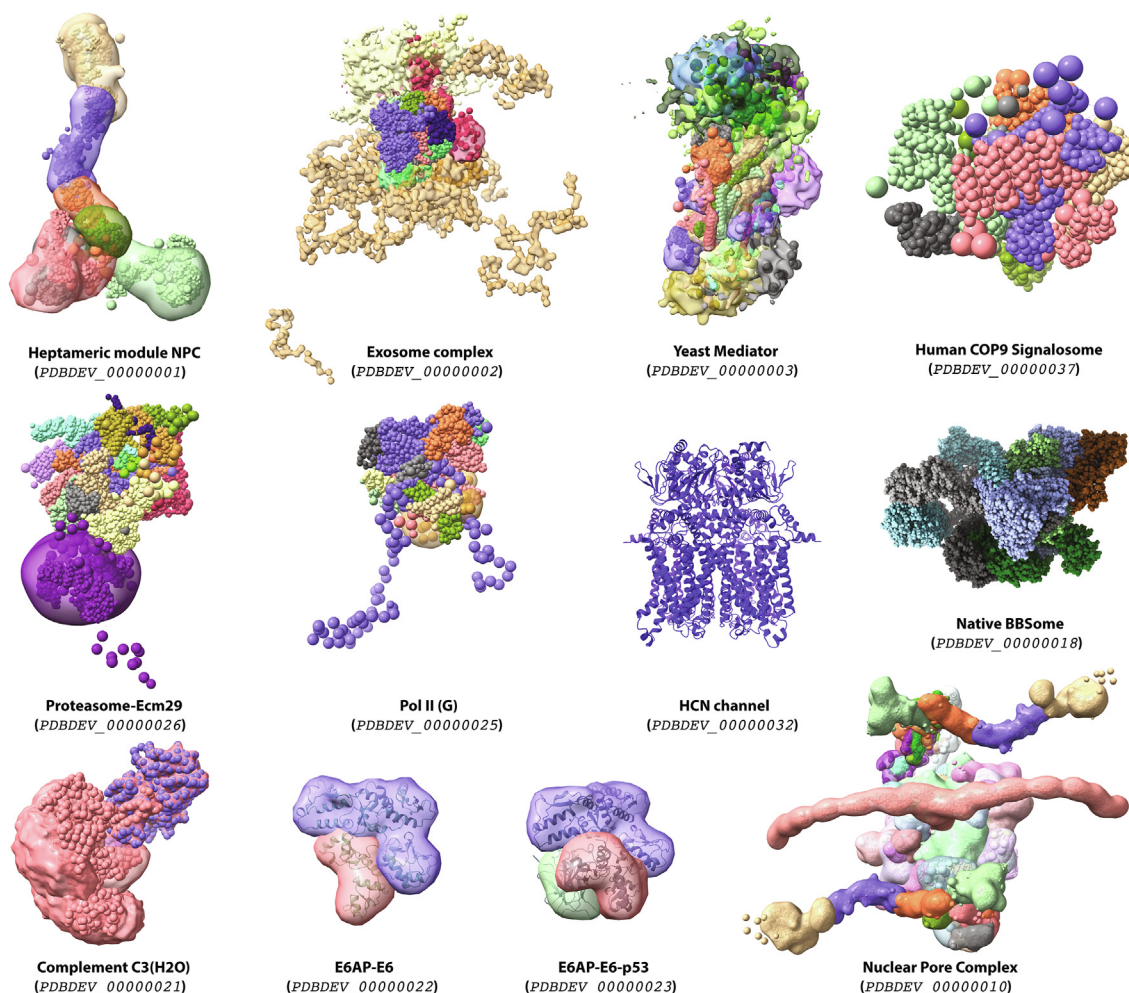
### 3.2. Parametrization of coarse-grained force fields

#### 3.2.1. Classical parametrization strategies

In the context of molecular modeling, the set of parameters and functions used to calculate the potential energy of a system is commonly referred to as force field. Atomistic force fields provide parameters usually for every type of atom in a system (hydrogen included) but also united atom representations are often used in which non-polar hydrogens are neglected. In contrast, coarse-grained potentials are a cruder representation of the inter- and intra-molecular interactions. Regarding the latter, their parametrization follows two main routes: Hierarchical (*bottom-up*) and pragmatic (*top-down*) coarse-graining [39].

The key idea of hierarchical coarse-graining is that, the interactions at a less detailed level are the result of the collective interac-





**Fig. 2.** Examples of integrative structures determined by partial/full coarse-grained/hybrid computational approaches as archived in the PDB-dev database [120,121] (pdb-dev.wwpdb.org). Pictures were generated with ChimeraX [136]. The experimental information used for the modeling (if included) has been omitted for visualization purposes. Models can be directly opened in ChimeraX from the command line as: `open [model_number] from pdbdev ignoreCache true`.

tions at the more detailed level [40]. As an illustration, in the 1975 abovementioned study by Levitt and Warshel [8], the interactions between coarse-grained sites were derived in a *bottom-up* way by explicitly summing up all microscopic interactions of an atomistic model. One obvious limitation of these models is that the quality of the coarse-grained model highly depends on the accuracy of the underlying atomistic one. Similarly, the seminal force-matching (FM) method proposed by Ercolessi and Adams [41] and further developed by Voth and co-workers [42,43] under the name of MS-CG (multiscale coarse-graining) uses atomistic-level interactions to derive coarse-grained potentials. In short, those potentials are systematically fitted to atomistic forces by minimizing the mean-square errors between them. Much like iterative Boltzmann (IB) derived models [44], these force fields are usually more accurate as compared to more generic ones. However, they are typically less transferable and require more parametrization effort. These methods, and their extensions [45], have been recently applied to coarse-grained models for proteins such as the UNRES model [46].

Pragmatic force fields, however, are designed in such a way that they reproduce a given chosen (experimental) property [47]. The earlier lattice models (such as HP) represent a well-studied example of *top-down* coarse-graining. These models are typically cheaper to parametrize, easily transferable (to similar systems) and use rather simple analytical potentials [48]. In a similar way and as shown in Fig. 1, methodologies based on reproducing thermodynamical properties have been extensively applied in different

branches of chemistry such as physical and organic chemistry. Equations of State (EoS), which are mathematical relationships between the thermodynamic variables of a given system, have been shown appropriate to accurately link the macroscopic properties of the system and the force field parameters [49]. As an example, the powerful SAFT- $\gamma$  EoS, a variation of the Statistical Associating Fluid Theory (SAFT), has been used to estimate the coarse-grained potentials of the Mie force field [50]. This force field has been recently used to calculate solvation free energies of aromatic compounds, which are broadly used in the pharmaceutical industry for drug design purposes [51].

### 3.2.2. Machine learning-based parametrization

Machine learning, and especially deep learning, is revolutionizing in the last years many areas of science and technology. Certainly, the most significant breakthrough of the decade in the field of protein folding has been the development of AlphaFold [52]. DeepMind, an artificial intelligence company affiliated to Google, has designed a deep learning-based method that represents a substantial advance as compared to classical modeling techniques [53,54]. These machine learning methods have been also applied in the development of force fields and are usually purely based on existing data. A general approach to design a machine (deep) learning-based force field typically includes: The generation of reference atomic configurations and forces (QM calculations), the identification of specific signatures, the selection of

training and test datasets, the mapping of selected signatures to forces using specific algorithms and the assessment of the resulting predictive model [55]. Deep neural networks [56], adversarial machine learning models [57] and genetic algorithm [58] have been recently shown appropriate for the development coarse-grained force fields. Altogether, machine learning-based parametrization methodologies represent an emerging trend to automatize analytical model building from more complex data, which can deliver faster and perhaps more accurate results with minimal human intervention.

### 3.3. Combining different levels of resolution

An exhaustive, yet accurate, sampling of the conformational landscape is crucial in attempts to model biomolecular interactions and evaluate the underlying energetics. The use of simplified representations offers an effective way of sampling the landscape. However, the reduced accuracy due to the inherent simplifications still limits the systems and processes that can be studied by CG approaches. Hybrid approaches, which typically couple coarse-grained and atomistic-level representations, aim to overcome these limitations by combining different levels of resolution [59]. These combined approaches might be very helpful for quantitative studies (e.g. free energy calculations of large systems [60,61]), while still reducing the computational cost. They are also particularly useful to include components of a system for which no or only low-resolution structural data are available. A key challenge in hybrid modeling is to integrate the different levels of resolution and to describe the AA/CG interactions. Standard mixing rules [62] have been historically very successful for this task. In short, Lennard-Jones and electrostatic interactions for mixed systems can be averaged and combined with an optimal scaling parameter depending on the size of the system [63]. Besides energetics, it still remains unclear how the interaction between two atoms might be affected by a coarse-grained surrounding as compared to its “native” environment and vice versa [64].

There are several hybrid schemes proposed in the literature, with MARTINI as a popular choice for the coarse-grained representation. One example is the PACE force field [65,66], which pairs MARTINI (water and lipids) with a united-atom protein model. In this case, the AA/CG parameters are optimized against specific thermodynamic data, which somehow limits its direct applicability to other systems. GROMOS/MARTINI coupling [64] has also been described as a potential alternative. In this work, cross-resolution interactions are calculated via virtual interactions sites on relevant atomistic groups and the standard CG beads, an approach that might lead to unbalanced electrostatics behaviors. For this reason, Wasenaar and coworkers [67] introduced an explicit electrostatic AA/CG coupling on the coarse-grained side. More recently, the CHARMM/PRIMO coupling has been proposed for single hybrid simulation purposes [68]. In the model proposed by Kar and Feig [68], the atomistic segment of the hybrid model was found to structurally deviate more than its corresponding one in a full atomistic model. This suggests that proper mixing of resolutions remains a difficult problem.

In the context of integrative modeling, the integration of experimental data at the various possible levels might have a crucial role for hybrid representations of the system. At the sampling level, data can be used to narrow the conformational search so that binding incompetent and/or irrelevant regions are discarded *a priori*. This strategy has been shown to be best suited compared to post-simulation filtering approaches. It not only outperforms the scenario where data is solely used to discard models with a high degree of uncertainty, but also reduces significantly the computational cost [69]. Data can be also incorporated at the scoring level via a numerical penalty term or as restraining energy potential

[70]. As an example, in HADDOCK [71] the distance restraints are incorporated into the scoring scheme via a soft-harmonic potential where the potential becomes linear for violations longer than 2 Å [72], effectively avoiding large forces for high restraints violations. Therefore, the incorporation of data in the modeling might work as a firewall and somewhat reduce the impact of inaccuracies of hybrid schemes in terms of intra- and inter-molecular interactions.

### 3.4. Sampling and scoring schemes

Decreasing the computational cost, as well as the complexity of the system, is a major goal of coarse-grained modeling. By lowering the resolution, the energy landscape becomes smoother and it is therefore, in principle, easier to identify the global minimum. In the context of integrative modeling with HADDOCK, we recently showed that introducing the MARTINI coarse-grained force field results in a substantial increase (8–30%) in the number of near-native models generated [73]. We also find CG sampling schemes in ATTRACT [74–76] (also hybrid scoring), CABS-dock [77,78] (also scoring), FRODOCK2.0 [79], InterEvDock2 [80,81] (also scoring), LZerD [82,83], MAXDo [84], MCDNA [85] (also scoring), MDockPP [86] and RosettaDock [87] (also scoring in RosettaDock 4.0 [88]). Some of the methods used by these software to sample the conformational landscape includes: Rigid-body energy minimization, Fast Fourier Transformation (FFT) or Molecular Dynamics (Monte Carlo). For the purpose of scoring, coarse-grained molecular dynamics simulations have been also evaluated on a heterogeneous benchmark of protein–protein docking models [89]. Other modeling software such as: DOCK/PIERR [90], GALAXY [91,92], LightDock [93], MEGADOCK 4.0 [94,95], PPI3D [96,97], pyDock [98,99] and V-D<sup>2</sup>OCK [100] incorporate, to some extent, coarse-grained/hybrid scoring approaches for (quasi)atomistic models.

IMP [101] and PyRy3D (genesilico.pl/pyry3d) are examples of ready-to-use hybrid modeling software for predicting (sampling and scoring) biomolecular assemblies allowing to incorporate experimental data into their calculations. The Integrative Modeling Platform leans on the concept that the resolution of the representation depends on the quantity and quality of the available information. This information is also encoded in a scoring function, whose ultimate goal is to evaluate the uncertainty of the generated models. Andrej Sali and co-workers [2] understand the modeling as an endless cyclic process driven by the continuous acquisition of data. In IMP, the different subunits are represented as a combination of spherical beads of varying sizes (different levels of coarseness). The same subunits can be also be represented as 3D Gaussians (for EM map fitting) and thus combine different resolution scales simultaneously [102]. During the conformational sampling, the relative distances from all the CG beads and Gaussians are either constrained (in rigid bodies) or restrained (in flexible bodies) by the sequence connectivity. For very high degrees of coarse-graining, only geometric considerations, e.g. exclude volume, might be used in the computations. PyRy3D allows for building low-resolution models of large macromolecular assemblies. In the software developed by Kasprzak and Bujnicki (genesilico.pl/pyry3d), proteins and nucleic acids can be represented as rigid-bodies or as flexible shapes. A spatial restraints-driven Monte Carlo approach is used to bring the components together followed by an evaluation via a simple scoring function. For a more detailed list of software that allow for building structural models of multi-subunit macromolecular complexes refer to Table 1.

### 3.5. Backmapping from coarse-grained to atomistic resolution

The inherent loss of accuracy of coarser representations is a limiting factor when analyzing integrative models of biomolecular complexes. Atomic details, such as specific contacts, are usually

**Table 1**

Available software for building structural models of protein, peptide and/or DNA complexes that incorporates a coarse-grained/hybrid approach into their protocols. Most of the listed software are available as webserver and/or standalone package.

Modeling platform	System(s)	Characteristics	Link	Reference(s)
ATTRACT	Protein, peptide and DNA	CG sampling and hybrid scoring	attract.ph.tum.de	[74–76]
CABS-dock	Peptide	CG sampling and scoring	biocomp.chem.uw.edu.pl	[77,78]
DOCK/PIERR	Protein	Hybrid scoring	clsbweb.oden.utexas.edu *	[90]
FRODOCK2.0	Protein	3D grid potential maps	frodock.chaconlab.org	[79]
GALAXY	Peptide	Hybrid scoring	galaxy.seoklab.org	[91,92]
HADDOCK	Protein, peptide and nucleic acids	CG sampling	bianca.science.uu.nl/haddock2.4	[73,103,104]
IMP	Protein and DNA	Hybrid sampling and scoring	integrativemodeling.org	[101]
InterEvDock2	Protein	Sampling by FRODOCK2.0 and CG scoring	bioserv.rpbs.univ-paris-diderot.fr/services/InterEvDock2	[80,81]
LightDock	Protein, peptide and DNA	Hybrid scoring	lightdock.org	[93]
LZerD**	Protein and peptide	3DZD representation and hybrid scoring	kiharaalab.org/proteindocking	[82,83]
MAXDo	Protein	CG sampling	lcqb.upmc.fr/CCDMintseris	[84]
MCDNA	Protein and DNA	CG sampling and scoring	mmb.irbbarcelona.org/MCDNA	[85]
MDockPP	Protein	CG sampling	zoulab.dalton.missouri.edu	[86]
MEGADOCK 4.0	Protein	Hybrid scoring	bi.cs.titech.ac.jp	[94,95]
PPI3D	Protein	Voronoi tessellation-based scoring	bioinformatics.ibt.lt/ppi3d	[96,97]
pyDock	Protein	CG scoring	life.bsc.es/pid/pydockweb	[98,99]
PyRy3D	Protein and DNA	Hybrid sampling and scoring	genesilico.pl/pyry3d	–
RosettaDock	Protein	CG sampling and scoring	rosettacommons.org	[87,88]
V-D <sup>2</sup> OCK	Protein	CG scoring	bioinsilico.org/cgi-bin/VD2OCK/	[100]

\* Submission to DOCK/PIERR webserver is no longer supported.

\*\* LZerD has a specific protocol for modeling unstructured protein–protein interactions [83].

essential to understand molecular recognition and it is therefore crucial to accurately reconstruct atomistic models from their CG counterparts [105]. This process is commonly referred in the literature as reverse transformation, inverse mapping or backmapping. There is currently a number of different backmapping protocols proposed, which mostly follow two different stages: (1) The generation of an atomistic structure based on the coarse-grained coordinates, and (2) a relaxation step of the generated AA structure.

For the first step, geometrical interpolation [23,106,107], random placement [108] and fragment-based methods [87,109–111] are the most used ones. All these methods perform sufficiently well according to backbone deviations (<1.0 Å in general) but side chain reconstruction seems more problematic [112]. Side chain optimization has been extensively studied as it directly applies for protein designing purposes. The most successful methods discretize possible side chain conformations into rotamers and usually require of an exhaustive search algorithm (e.g. Monte Carlo, simulated annealing...) and an effective scoring function for selecting the proper side chain conformation. The backmapped atomistic structures can then be further improved by energy minimization [73,113] and/or more sophisticated molecular dynamics-based approaches [114]. In HADDOCK, the CG generated models are converted into atomistic resolution by using distance restraints between the atoms and their corresponding coarse-grained beads. Using those restraints, the all-atom models of the individual components of a complex are morphed onto the coarse grained complex by a series of energy minimizations and Cartesian molecular dynamics [73].

#### 4. Application examples of integrative modeling of protein interactions

Ultimately, the true value of any biomolecular model is in the structural information and insights that it provides. When speaking about integrative modeling here, we refer to the branch of structural biology whose aim is to gain structural insights into biomolecular complexes by integrating a wide variety of experimental information into computational calculations. There are var-

ious challenges associated with the incorporation and use of that information for the modeling of assemblies. However, a detailed overview of those is beyond the scope of this manuscript and have been reviewed in depth elsewhere [115–117]. The relevance of integrative models is underscored by the fact that the Protein Data Bank [118,119] has now started to collect them in a new integrative model database (PDB-dev; pdb-dev.www.pdb.org) [120,121], which ultimately should be merged into the current PDB database. Since 2014, it is possible to archive structural models obtained by combining traditional structural experimental techniques such as NMR spectroscopy, electron microscopy (3DEM), small angle scattering (SAS), atomic force microscopy (AFM), chemical cross-linking, Förster resonance energy transfer (FRET), electron paramagnetic resonance (EPR), mass spectrometry (MS), Hydrogen/Deuterium exchange (HDX) and various bioinformatic approaches, with computational methods. In this section we highlight several examples of integrative structures of protein complexes that have been determined by combining coarse-grained/hybrid computational approaches with experimental information.

Among all archived structures, we find a number of them determined by coarse-grained/hybrid computational methods in combination with a wide variety of structural data (see Fig. 2). Integrative structures derived from chemical cross-linking data are by far the most abundant ones, including models of the heptameric module of NPC [122], the exosome complex [123], the Complement C3 (H2O) [124], the EGAP/UBE3A-p53 enzyme-substrate complex [125], Pol II(G) [126], the Proteasome-Ecm29 complex [127] and the canonical/non-canonical human COP9 Signalosome [128]. Protein cross-links have been also combined with other types of experimental information such as three/two-dimensional Electron Microscopy (2DEM/3DEM) and/or SAS to determine structures like the yeast Mediator complex [129] or the native BBSome [130]. Other sources of information such as mutagenesis and NMR data [131] and single molecule FRET data [132] have been also used.

There are also multiple examples of integrative structures, not deposited in the PDB-dev database, which have been modelled by integrative coarse-graining methods. One of those is the ATP synthase membrane motor. Leone and Faraldo-Gómez [133] pro-



posed a computational integrative model based on chemical cross-links, a cryo-EM map ( $\sim 7\text{\AA}$  of resolution) and evolutionary couplings. The initial homology models of either subunits were refined against the experimentally determined cryo-EM map using Rosetta, which starts its conformational exploration in coarse-grained resolution. The computationally generated models were further validated with co-evolutionary and cross-linking data and revealed important mechanistic insights into the function of the ATP synthase. Another representative example is the ISWI ATPase complex. Using upper bound distance restraints based on BS<sup>3</sup>, BS<sup>2</sup>G and UV cross-links, Harrer and coworkers [134] modelled the complex with ATTRACT, which performs a rigid-body energy minimization driven by a coarse-grained force field [109] and the distance restraints provided. The top scoring ISWI models were validated against SAXS data.

The Nuclear Pore Complex (NPC) is probably the largest protein assembly determined by an integrative structural approach to date. It constitutes an eight-fold symmetrical cylindrical complex of 552 copies of 32 different nucleoporin proteins (Nups) [135]. With respect to the computational modeling, the NPC was represented in a multiscale fashion including multiple levels of coarseness. As an illustration, all rigid bodies derived from X-ray, NMR and integrative structures were coarse-grained into two different resolutions. They either mapped single residues or consecutive portions of up to ten different amino acids into larger beads. The modeling was performed using the integrative modeling platform software (IMP) (integrativemodeling.org) [101]. The experimental information available included chemical cross-links, a cryo-ET density map, immuno-electron microscopy localizations, excluded volume, sequence connectivity, the shape of the pore membrane, symmetry and SAXS data, which were used to benefit the sampling, to improve the scoring, to filter out inconsistent models and/or validation purposes. By putting all these data together, they were able to fully describe, at sub-nanometer precision, the structure of the entire NPC.

## 5. Concluding remarks

Over the past decades, coarse-grained/hybrid modeling has been demonstrated as a powerful approach to model biomolecules and their interactions. It extends the capabilities of traditional atomistic protocols. There are multiple models to simplify the three-dimensional representation of biomolecules, each of those specifically designed to answer a specific research question. The choice between different representations directly affects the sampling and scoring capabilities of current modeling approaches. In other words, the smaller the number of pseudo-atoms or beads, the higher the increase in speed but the lower the accuracy of the resulting models. For cases where higher level of resolution is required, multiscale/hybrid modeling might help to alleviate the inherent loss of accuracy of pure coarse-grained models as demonstrated, for instance, in the modeling of the nuclear pore complex. Nevertheless, there is still an urgent need for improving interaction schemes. Coarse-grained force fields derived from classical molecular mechanics are not easily transferable and therefore, very much system-dependent. On the contrary to *bottom-up* strategies, *top-down* approaches aim to generalize structural patterns that have been seen in thousands of known structures and/or to reproduce thermodynamic quantities. Likely, a combination of *bottom-up* and *top-down* approaches is a better option. In other words, improving *top-down* models by inferring additional interaction terms derived by *bottom-up* coarse-graining might have the most impact in future designs, increasing both their accuracy and applicability range to wider, larger and more complex assemblies. We are now approaching a time where, taking advantage of all scien-

tific and technological advances, one might expect to build reasonable three-dimensional models of cells, which might provide insights into still unknown cellular mechanisms.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

This work was supported by the Dutch Foundation for Scientific Research (NWO) (TOP Grant 718.015.001) and by the BioExcel CoE (www.bioexcel.eu), a project funded by the European Union Horizon 2020 program under grant agreements 675728 and 823820.

The authors acknowledge all members from the Computational Structural Biology group at Utrecht University with special mention to Dr. Brian Jiménez-García for fruitful discussions and Dr. Zuzana Jandová for carefully reviewing this manuscript.

## Appendix A. Supplementary data

There is no supplementary data associated to the present work.

## References

- [1] Berggård T, Linse S, James P. Methods for the detection and analysis of protein-protein interactions. *Proteomics* 2007;7:2833–42.
- [2] Rout MP, Sali A. Principles for Integrative Structural Biology Studies. *Cell* 2019;177:1384–403.
- [3] Koukos PI, Bonvin AMJJ. Integrative Modelling of Biomolecular Complexes. *J. Mol. Biol.* In press; 2019.
- [4] Braitbard M, Schneidman-Duhovny D, Kalisman N. Integrative Structure Modeling: Overview and Assessment. *Annu Rev Biochem* 2019;88:113–35.
- [5] Singla J et al. Opportunities and challenges in building a spatiotemporal multi-scale model of the human pancreatic  $\beta$  Cell. *Cell* 2018;173:11–9.
- [6] Phillips DC. Symposium on Three-Dimensional Structure of Macromolecules of Biological Origin. By Invitation of the Committee on Arrangements for the Autumn Meeting. Presented before the Academy on October 19, 1966. Chairman, Walter Kauzmann. *Proc Natl Acad Sci* 1967;57:483–95.
- [7] Warshel A. Multiscale modeling of biological functions: From enzymes to molecular machines (nobel lecture). *Angew Chemie – Int Ed* 2014;53:10020–31.
- [8] Levitt M, Warshel A. Computer simulation of protein folding. *Nature* 1975;253:694–8.
- [9] Levinthal C. How to fold graciously. *Mössbauer Spectrosc Biol Syst Proc* 1969;24:22–4.
- [10] Chothia C, Janin J. Principles of protein-protein recognition. *Nature* 1975;256:705–8.
- [11] Wodak SJ, Janin J. Computer analysis of protein-protein interaction. *J Mol Biol* 1978;124:323–42.
- [12] Fersht AR et al. Analysis of Enzyme Structure and Activity by Protein Engineering. *Angew Chemie Int Ed English* 1984;23:467–73.
- [13] Kmiecik S et al. Coarse-grained protein models and their applications. *Chem Rev* 2016;116:7898–936.
- [14] Lau KF, Dill KA. A lattice statistical mechanics model of the conformational and sequence spaces of proteins. *Macromolecules* 1989;22:3986–97.
- [15] Dill KA et al. Principles of protein folding – A perspective from simple exact models. *Protein Sci* 1995;4:561–602.
- [16] Sali A, Shakhnovich E, Karplus M. Kinetics of protein folding: A lattice model study of the requirements for folding to the native state. *J Mol Biol* 1994;235:1614–36.
- [17] Dinner AR, Sali A, Karplus M. The folding mechanism of larger model proteins: Role of native structure. *Proc Natl Acad Sci U S A* 1996;93:8356–61.
- [18] Locker CR, Hernandez R. A minimalist model protein with multiple folding funnels. *Proc Natl Acad Sci U S A* 2001;98:9074–9.
- [19] Kaya H, Chan HS. Towards a consistent modeling of protein thermodynamic and kinetic cooperativity: How applicable is the transition state picture to folding and unfolding? *J Mol Biol* 2002;315:899–909.
- [20] Kolinski A, Skolnick J. Reduced models of proteins and their applications. *Polymer (Guildf)* 2004;45:511–24.
- [21] Kolinski A, Skolnick J. Monte carlo simulations of protein folding. I. Lattice model and interaction scheme. *Proteins Struct Funct Bioinforma* 1994;18:338–52.

- [22] MacKerell AD, Feig M, Brooks CL. Improved treatment of the protein backbone in empirical force fields. *J Am Chem Soc* 2004;126:698–9.
- [23] Gopal SM, Mukherjee S, Cheng YM, Feig M. PRIMO/PRIMONA: A coarse-grained model for proteins and nucleic acids that preserves near-atomic accuracy. *Proteins Struct Funct Bioinforma* 2010;78:1266–81.
- [24] Pasquali S, Derreumaux P. HiRE-RNA: A high resolution coarse-grained energy model for RNA. *J Phys Chem B* 2010;114:11957–66.
- [25] Darré L et al. SIRAH: A structurally unbiased coarse-grained force field for proteins with aqueous solvation and long-range electrostatics. *J Chem Theory Comput* 2015;11:723–39.
- [26] Dans PD, Zeida A, MacHado MR, Pantano S. A coarse grained model for atomic-detailed DNA simulations with explicit electrostatics. *J Chem Theory Comput* 2010;6:1711–25.
- [27] Darré L, MacHado MR, Dans PD, Herrera FE, Pantano S. Another coarse grain model for aqueous solvation: WAT FOUR?. *J Chem Theory Comput* 2010;6:3793–807.
- [28] Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, De Vries AH. The MARTINI force field: Coarse grained model for biomolecular simulations. *J Phys Chem B* 2007;111:7812–24.
- [29] Monticelli L et al. The MARTINI coarse-grained force field: Extension to proteins. *J Chem Theory Comput* 2008;4:819–34.
- [30] Lee H, De Vries AH, Marrink SJ, Pastor RW. A coarse-grained model for polyethylene oxide and polyethylene glycol: Conformation and hydrodynamics. *J Phys Chem B* 2009;113:13186–94.
- [31] Gobbo C et al. MARTINI model for physisorption of organic molecules on graphite. *J Phys Chem C* 2013;117:15623–31.
- [32] López CA et al. Martini coarse-grained force field: Extension to carbohydrates. *J Chem Theory Comput* 2009;5:3195–210.
- [33] Yesylevskyy SO, Schäfer LV, Sengupta D, Marrink SJ. Polarizable water model for the coarse-grained MARTINI force field. *PLoS Comput Biol* 2010;6:1–17.
- [34] López CA, Sovova Z, Van Eerden FJ, De Vries AH, Marrink SJ. Martini force field parameters for glycolipids. *J Chem Theory Comput* 2013;9:1694–708.
- [35] Uusitalo JJ, Ingólfsson HI, Akhshi P, Tieleman DP, Marrink SJ. Martini coarse-grained force field: extension to DNA. *J Chem Theory Comput* 2015;11:3932–45.
- [36] Uusitalo JJ, Ingólfsson HI, Marrink SJ, Faustino I. Martini coarse-grained force field: extension to RNA. *Biophys J* 2017;113:246–56.
- [37] López CA et al. MARTINI coarse-grained model for crystalline cellulose microfibrils. *J Phys Chem B* 2015;119:465–73.
- [38] De Jong DH et al. Improved parameters for the martini coarse-grained protein force field. *J Chem Theory Comput* 2013;9:687–97.
- [39] Noid WG. Perspective: Coarse-grained models for biomolecular systems. *J Chem Phys* 2013;139.
- [40] Saunders MG, Voth GA. Coarse-graining methods for computational biology. *Annu Rev Biophys* 2013;42:73–93.
- [41] Ercolessi F, Adams JB. Interatomic potentials from first-principles calculations: The force-matching method. *EPL* 1994;26:583–8.
- [42] Izvekov S, Parrinello M, Burnham CJ, Voth GA. Effective force fields for condensed phase systems from ab initio molecular dynamics simulation: A new method for force-matching. *J Chem Phys* 2004;120:10896–913.
- [43] Izvekov S, Voth GA. A Multiscale Coarse-Graining Method for Biomolecular Systems. *J Phys Chem B* 2005;109:2469–73.
- [44] Soper AK. Empirical potential Monte Carlo simulation of fluid structure. *Chem Phys* 1996;202:295–306.
- [45] Lu L, Dama JF, Voth GA. Fitting coarse-grained distribution functions through an iterative force-matching method. *J Chem Phys* 2013;139:121906.
- [46] Liwo A, Czaplewski C. Extension of the force-matching method to coarse-grained models with axially symmetric sites to produce transferable force fields: Application to the UNRES model of proteins. *J Chem Phys* 2020;152:054902.
- [47] Ingólfsson HI et al. The power of coarse graining in biomolecular simulations Wiley Interdisciplinary Reviews. *Comput Mole Sci* 2014;4:225–48.
- [48] Potoyan, D., Papoian G. A. The need for computational Speed: State of the art in DNA coarse graining. In: *Coarse-Grained Modeling of Biomolecules 1st ed.* Boca Raton: CRC Press, 271–297 (2017).
- [49] Mejía A, Herdes C, Müller EA. Force fields for coarse-grained molecular simulations from a corresponding states correlation. *Ind. Eng. Chem. Res.* 2014;53:4131–41.
- [50] Müller EA, Jackson G. Force-field parameters from the SAFT- $\gamma$  equation of state for use in coarse-grained molecular simulations. *Annu Rev Chem Biomol Eng* 2014;5:405–27.
- [51] Matos IQ, Abreu CRA. Evaluation of the SAFT- $\gamma$  Mie force field with solvation free energy calculations. *Fluid Phase Equilib* 2019;484:88–97.
- [52] Senior AW et al. Improved protein structure prediction using potentials from deep learning. *Nature* 2020;577:706–10.
- [53] Senior AW et al. Protein structure prediction using multiple deep neural networks in the 13th Critical Assessment of Protein Structure Prediction (CASP13). *Proteins Struct Funct Bioinforma* 2019;87:1141–8.
- [54] Kryshchukovych A, Schwede T, Topf M, Fidelis K, Moulton J. Critical assessment of methods of protein structure prediction (CASP) – Round XIII. *Proteins Struct Funct Bioinforma* 2019;87:1011–20.
- [55] Botu V, Batra R, Chapman J, Ramprasad R. Machine learning force fields: Construction, validation, and outlook. *J Phys Chem C* 2017;121:511–22.
- [56] Wang J et al. Machine learning of coarse-grained molecular dynamics force fields. *ACS Cent Sci* 2019;5:755–67.
- [57] Durumeric AEP, Voth GA. Adversarial-residual-coarse-graining: Applying machine learning theory to systematic molecular coarse-graining. *J Chem Phys* 2019;151:124110.
- [58] Chan H et al. Machine learning coarse grained models for water. *Nat Commun* 2019;10:379.
- [59] Ayton GS, Noid WG, Voth GA. Multiscale modeling of biomolecular systems: in serial and in parallel. *Curr Opin Struct Biol* 2007;17:192–8.
- [60] König G, Hudson PS, Boresch S, Woodcock HL. Multiscale free energy simulations: An efficient method for connecting classical MD simulations to QM or QM/MM free energies using non-Boltzmann Bennett reweighting schemes. *J Chem Theory Comput* 2014;10:1406–19.
- [61] Lee S, Liang R, Voth GA, Swanson JM. Computationally efficient multiscale reactive molecular dynamics to describe amino acid deprotonation in proteins. *J Chem Theory Comput* 2016;12:879–91.
- [62] Scott R, Allen MP, Tildesley DJ. Computer simulation of liquids. *Math Comput* 1991;57:442.
- [63] Michel J, Orsi M, Essex JW. Prediction of partition coefficients by multiscale hybrid atomic-level/coarse-grain simulations. *J Phys Chem B* 2008;112:657–60.
- [64] Rzeplia AJ, Louhivuori M, Peter C, Marrink SJ. Hybrid simulations: Combining atomistic and coarse-grained force fields using virtual sites. *Phys Chem Chem Phys* 2011;13:10437–48.
- [65] Wan CK, Han W, Wu YD. Parameterization of PACE force field for membrane environment and simulation of helical peptides and helix-helix association. *J Chem Theory Comput* 2012;8:300–13.
- [66] Ward MD, Nangia S, May ER. Evaluation of the hybrid resolution PACE model for the study of folding, insertion, and pore formation of membrane associated peptides. *J Comput Chem* 2017;38:1462–71.
- [67] Wassenaar TA, Ingólfsson HI, Prieß M, Marrink SJ, Schäfer LV. Mixing MARTINI: Electrostatic coupling in hybrid atomistic-coarse-grained biomolecular simulations. *J Phys Chem B* 2013;117:3516–30.
- [68] Kar P, Feig M. Hybrid all-atom/coarse-grained simulations of proteins by direct coupling of CHARMM and PRIMO force fields. *J Chem Theory Comput* 2017;13:5753–65.
- [69] Roel-Touris J, Bonvin AMJJ, Jiménez-García B. LightDock goes information-driven. *Bioinformatics* 2020;36:950–2.
- [70] Vangone A et al. Sense and simplicity in HADDOCK scoring: Lessons from CASP-CAPRI round 1. *Proteins Struct Funct Bioinforma* 2017;85:417–23.
- [71] Dominguez C, Boelens R, Bonvin AMJJ. HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* 2003;125:1731–7.
- [72] Brünger AT et al. Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr Sect D Biol Crystallogr* 1998;54:905–21.
- [73] Roel-Touris J, Don CG, Honorato RR, Rodrigues JPGLM, Bonvin AMJJ. Less Is More: coarse-grained integrative modeling of large biomolecular assemblies with HADDOCK. *J Chem Theory Comput* 2019;15:6358–67.
- [74] De Vries S, Zacharias M. Flexible docking and refinement with a coarse-grained protein model using ATTRACT. *Proteins Struct Funct Bioinforma* 2013;81:2167–74.
- [75] De Vries SJ, Rey J, Schindler CEM, Zacharias M, Tuffery P. The pepATTRACT web server for blind, large-scale peptide-protein docking. *Nucleic Acids Res* 2017;45:W361–4.
- [76] Setny P, Bahadur RP, Zacharias M. Protein-DNA docking with a coarse-grained force field. *BMC Bioinform* 2012;13.
- [77] Kolinski A. Protein modeling and structure prediction with a reduced representation. *Acta Biochim Pol* 2004;51:349–71.
- [78] Kurcinski M, Jamroz M, Blaszczyk M, Kolinski A, Kmiecik S. CABS-dock web server for the flexible docking of peptides to proteins without prior knowledge of the binding site. *Nucleic Acids Res* 2015;43:W419–24.
- [79] Ramírez-Aportela E, López-Blanco JR, Chacón P. FRODOCK 2.0: Fast protein-protein docking server. *Bioinformatics* 2016;32:2386–8.
- [80] Andreani J, Faure G, Guerois R. InterEvScore: A novel coarse-grained interface scoring function using a multi-body statistical potential coupled to evolution. *Bioinformatics* 2013;29:1742–9.
- [81] Quignot C et al. InterEvDock2: An expanded server for protein docking using evolutionary and biological information from homology models and multimeric inputs. *Nucleic Acids Res* 2018;46:W408–16.
- [82] Esquivel-Rodriguez J, Filos-Gonzalez V, Li B, Kihara D. Pairwise and multimeric protein-protein docking using the IZerD program suite. *Methods Mol Biol* 2014;1137:209–34.
- [83] Peterson LX, Roy A, Christoffer C, Terashi G, Kihara D. Modeling disordered protein interactions from biophysical principles. *PLoS Comput Biol* 2017;13.
- [84] Sacquin-Mora S, Carbone A, Lavery R. Identification of protein interaction partners and protein-protein interaction sites. *J Mol Biol* 2008;382:1276–89.
- [85] Walther J et al. A multi-modal coarse grained model of DNA flexibility mappable to the atomistic level. *Nucleic Acids Res* 2020;48.
- [86] Huang SY, Zou X. MDockPP: A hierarchical approach for protein-protein docking and its application to CAPRI rounds 15–19. *Proteins Struct Funct Bioinforma* 2010;78:3096–103.
- [87] Gray JJ et al. Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J Mol Biol* 2003;331:281–99.
- [88] Roy Burman SS et al. Novel sampling strategies and a coarse-grained scoring function for docking homomers, flexible heteromers, and oligosaccharides



- using Rosetta in CAPRI rounds 37–45. *Proteins Struct. Funct. Bioinforma.* In press; 2019.
- [89] Hou Q, Lensink MF, Heringa J, Feenstra KA. CLUB-MARTINI: Selecting favourable interactions amongst available candidates, a coarse-grained simulation approach to scoring docking decoys. *PLoS One* 2016;11.
- [90] Viswanath S, Ravikant DVS, Elber R. DOCK/PIERR: Web server for structure prediction of protein-protein complexes. *Methods Mol Biol* 2014;1137:199–207.
- [91] Shin W-H, Lee GR, Heo L, Lee H, Seok C. Prediction of protein structure and interaction by GALAXY protein modeling programs. *Biol Des* 2014;2:01–11.
- [92] Lee H, Heo L, Lee MS, Seok C. GalaxyPepDock: A protein-peptide docking tool based on interaction similarity and energy optimization. *Nucleic Acids Res* 2015;43:W431–5.
- [93] Jiménez-García B et al. LightDock: A new multi-scale approach to protein-protein docking. *Bioinformatics* 2018;49:34–55.
- [94] Ohue M, Matsuzaki Y, Ishida T, & Akiyama Y. Improvement of the protein-protein docking prediction by introducing a simple hydrophobic interaction model: An application to interaction pathway analysis. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 7632 LNBI, 178–187 (2012).
- [95] Ohue M et al. MEGADOCK 4.0: An ultra-high-performance protein-protein docking software for heterogeneous supercomputers. *Bioinformatics* 2014;30:3281–3.
- [96] Olechnovič K, Venclovas Č. Voronota: A fast and reliable tool for computing the vertices of the Voronoi diagram of atomic balls. *J Comput Chem* 2014;35:672–81.
- [97] Dapkunas J et al. The PPI3D web server for searching, analyzing and modeling protein-protein interactions in the context of 3D structures. *Bioinformatics* 2017;33:935–7.
- [98] Jiménez-García B, Pons C, Fernández-Recio J. pyDockWEB: A web server for rigid-body protein-protein docking using electrostatics and desolvation scoring. *Bioinformatics* 2013;29:1698–9.
- [99] Solernou A, Fernandez-Recio J. PyDockCG: New coarse-grained potential for protein-protein docking. *J Phys Chem B* 2011;115:6032–9.
- [100] Segura J, Marín-López MA, Jones PF, Oliva B, Fernandez-Fuentes N. VORFFIP-driven dock: V-D2OCK, a fast, accurate protein docking strategy. *PLoS One* 2015;10.
- [101] Russel D et al. Putting the pieces together: Integrative modeling platform software for structure determination of macromolecular assemblies. *PLoS Biol* 2012;10(1):e1001244.
- [102] Webb B et al. Integrative structure modeling with the Integrative Modeling Platform. *Protein Sci* 2018;27:245–58.
- [103] Van Zundert GCP et al. The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes. *J Mol Biol* 2016;428:720–5.
- [104] Honorato RV, Roel-Touris J, Bonvin AMJJ. MARTINI-Based Protein-DNA Coarse-Grained HADDOCKing. *Front Mol Biosci* 2019;6.
- [105] Badaczewska-Dawid AE, Kolinski A, Kmiecik S. Computational reconstruction of atomistic protein structures from coarse-grained models. *Comput Struct Biotechnol J* 2020;18:162–76.
- [106] Wassenaar TA, Pluhackova K, Böckmann RA, Marrink SJ, Tieleman DP. Going backward: A flexible geometric approach to reverse transformation from coarse grained to atomistic models. *J Chem Theory Comput* 2014;10:676–90.
- [107] Machado MR, Pantano S. SIRAH tools: Mapping, backmapping and visualization of coarse-grained models. *Bioinformatics* 2016;32:1568–70.
- [108] Rzepiela AJ et al. Software news and update reconstruction of atomistic details from coarse-grained structures. *J Comput Chem* 2010;31:1333–43.
- [109] Zacharias M. Protein-protein docking with a reduced protein model accounting for side-chain flexibility. *Protein Sci* 2003;12:1271–82.
- [110] Shimizu M, Takada S. Reconstruction of atomistic structures from coarse-grained models for Protein-DNA complexes. *J Chem Theory Comput* 2018;14:1682–94.
- [111] Heath AP, Kaviraki LE, Clementi C. From coarse-grain to all-atom: Toward multiscale analysis of protein landscapes. *Proteins Struct Funct Genet* 2007;68:646–61.
- [112] Lombardi LE, Marti MA, Capece L. CG2AA: Backmapping protein coarse-grained structures. *Bioinformatics* 2016;32:1235–7.
- [113] Joosten RP, Joosten K, Murshudov GN, Perrakis A. PDB-REDO: Constructive validation, more than just looking for errors. *Acta Crystallogr Sect D Biol Crystallogr* 2012;68:484–96.
- [114] Peng J, Yuan C, Ma R, Zhang Z. Backmapping from multiresolution coarse-grained models to atomic structures of large biomolecules by restrained molecular dynamics simulations using bayesian inference. *J Chem Theory Comput* 2019;15:3344–53.
- [115] London N, Raveh B, Schueler-Furman O. Peptide docking and structure-based characterization of peptide binding: from knowledge to know-how. *Curr Opin Struct Biol* 2013;23:894–902.
- [116] Rodrigues JPGLM, Bonvin AMJJ. Integrative computational modeling of protein interactions. *FEBS J* 2014;281:1988–2003.
- [117] Nithin C, Ghosh P, Bujnicki J. Bioinformatics tools and benchmarks for computational docking and 3D structure prediction of RNA-Protein complexes. *Genes (Basel)* 2018;9:432.
- [118] Joosten RP et al. A series of PDB related databases for everyday needs. *Nucleic Acids Res* 2011;39.
- [119] Touw WG et al. A series of PDB-related databanks for everyday needs. *Nucleic Acids Res* 2015;43:D364–8.
- [120] Sali A et al. Outcome of the first wwPDB hybrid/integrative methods task force workshop. *Structure* 2015;23:1156–67.
- [121] Berman HM et al. Federating structural models and data: outcomes from a workshop on archiving integrative structures. *Structure* 2019;27:1745–59.
- [122] Shi Y et al. Structural characterization by cross-linking reveals the detailed architecture of a coatomer-related heptameric module from the nuclear pore complex. *Mol Cell Proteomics* 2014;13:2927–43.
- [123] Shi Y et al. A strategy for dissecting the architectures of native macromolecular assemblies. *Nat Methods* 2015;12:1135–8.
- [124] Chen ZA et al. Structure of complement C3(H2O) revealed by quantitative cross-linking/mass spectrometry and modeling. *Mol Cell Proteomics* 2016;15:2730–43.
- [125] Sailer C et al. Structural dynamics of the E6AP/UBE3A-E6-p53 enzyme-substrate complex. *Nat Commun* 2018;9:4441.
- [126] Jishage M et al. Architecture of Pol II(G) and molecular mechanism of transcription regulation by Gdown1. *Nat Struct Mol Biol* 2018;25:859–67.
- [127] Wang X et al. The proteasome-interacting Ecm29 protein disassembles the 26S proteasome in response to oxidative stress. *J Biol Chem* 2017;292:16310–20.
- [128] Gutierrez C et al. Structural dynamics of the human COP9 signalosome revealed by cross-linking mass spectrometry and integrative modeling. *Proc Natl Acad Sci U S A* 2020;117:4088–98.
- [129] Robinson PJ et al. Molecular architecture of the yeast Mediator complex. *Elife* 2015;4.
- [130] Chou HT et al. The molecular architecture of native BBSome obtained by an integrated structural approach. *Structure* 2019;27:1384–1394.e4.
- [131] Bender BJ et al. Structural model of ghrelin bound to its G protein-coupled receptor. *Structure* 2019;27:537–544.e4.
- [132] Dai G, Aman TK, DiMaio F, Zagotta WN. The HCN channel voltage sensor undergoes a large downward motion during hyperpolarization. *Nat Struct Mol Biol* 2019;26:686–94.
- [133] Leone V, Faraldo-Gómez JD. Structure and mechanism of the ATP synthase membrane motor inferred from quantitative integrative modeling. *J Gen Physiol* 2016;148:441–57.
- [134] Harrer N et al. Structural architecture of the nucleosome remodeler ISWI determined from cross-linking, mass spectrometry, SAXS, and modeling. *Structure* 2018;26:282–294.e6.
- [135] Kim SJ et al. Integrative structure and functional anatomy of a nuclear pore complex. *Nature* 2018;555:475–82.
- [136] Goddard TD et al. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci* 2018;27:14–25.