



**HAL**  
open science

## Building Biological Relevance Into Integrative Modelling of Macromolecular Assemblies

Anne-Elisabeth Molza, Yvonne Westermaier, Magali Moutte, Pierre Ducrot,  
Claudia Danilowicz, Veronica Godoy-Carter, Mara Prentiss, Charles H.  
Robert, Marc Baaden, Chantal Prévost

► **To cite this version:**

Anne-Elisabeth Molza, Yvonne Westermaier, Magali Moutte, Pierre Ducrot, Claudia Danilowicz, et al.. Building Biological Relevance Into Integrative Modelling of Macromolecular Assemblies. *Frontiers in Molecular Biosciences*, 2022, 9, pp.826136. 10.3389/fmolb.2022.826136 . hal-03826944

**HAL Id: hal-03826944**

**<https://hal.science/hal-03826944v1>**

Submitted on 24 Oct 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# Building Biological Relevance Into Integrative Modelling of Macromolecular Assemblies

Anne-Elisabeth Molza<sup>1,2†</sup>, Yvonne Westermaier<sup>3</sup>, Magali Moutte<sup>4</sup>, Pierre Ducrot<sup>3</sup>, Claudia Danilowicz<sup>5</sup>, Veronica Godoy-Carter<sup>6</sup>, Mara Prentiss<sup>5</sup>, Charles H. Robert<sup>1,2</sup>, Marc Baaden<sup>1,2</sup> and Chantal Prévost<sup>1,2\*</sup>

<sup>1</sup>CNRS, Université Paris-Cité, UPR 9080, Laboratoire de Biochimie Théorique, Paris, France, <sup>2</sup>Institut de Biologie Physico-Chimique-Fondation Edmond de Rothschild, PSL Research University, Paris, France, <sup>3</sup>Biophysics and Modelling Department/In Vitro Pharmacology Unit-IDRS (Servier Research Institute), Croissy-sur-Seine, France, <sup>4</sup>Servier Monde, Suresnes, France, <sup>5</sup>Department of Physics, Harvard University, Cambridge, MA, United States, <sup>6</sup>Department of Biology, Northeastern University, Boston, MA, United States

## OPEN ACCESS

### Edited by:

Paulo Ricardo Batista,  
Oswaldo Cruz Foundation (Fiocruz),  
Brazil

### Reviewed by:

Tahsin F. Kellici,  
Evotec, United Kingdom  
Lorenzo Stella,  
University of Rome Tor Vergata, Italy

### \*Correspondence:

Chantal Prévost  
chantal.prevost@ibpc.fr

### †Present address:

Anne-Elisabeth Molza,  
DSIMB, UMR-S1134 INSERM,  
Faculté des Sciences et Technologies,  
Université de La Réunion, Saint Denis,  
La Réunion, France

### Specialty section:

This article was submitted to  
Structural Biology,  
a section of the journal  
Frontiers in Molecular Biosciences

Received: 30 November 2021

Accepted: 21 February 2022

Published: 11 April 2022

### Citation:

Molza A-E, Westermaier Y, Moutte M,  
Ducrot P, Danilowicz C,  
Godoy-Carter V, Prentiss M,  
Robert CH, Baaden M and Prévost C  
(2022) Building Biological Relevance  
Into Integrative Modelling of  
Macromolecular Assemblies.  
*Front. Mol. Biosci.* 9:826136.  
doi: 10.3389/fmolb.2022.826136

Recent advances in structural biophysics and integrative modelling methods now allow us to decipher the structures of large macromolecular assemblies. Understanding the dynamics and mechanisms involved in their biological function requires rigorous integration of all available data. We have developed a complete modelling pipeline that includes analyses to extract biologically significant information by consistently combining automated and interactive human-guided steps. We illustrate this idea with two examples. First, we describe the ryanodine receptor, an ion channel that controls ion flux across the cell membrane through transitions between open and closed states. The conformational changes associated with the transitions are small compared to the considerable system size of the receptor; it is challenging to consistently track these states with the available cryo-EM structures. The second example involves homologous recombination, in which long filaments of a recombinase protein and DNA catalyse the exchange of homologous DNA strands to reliably repair DNA double-strand breaks. The nucleoprotein filament reaction intermediates in this process are short-lived and heterogeneous, making their structures particularly elusive. The pipeline we describe, which incorporates experimental and theoretical knowledge combined with state-of-the-art interactive and immersive modelling tools, can help overcome these challenges. In both examples, we point to new insights into biological processes that arise from such interdisciplinary approaches.

**Keywords:** integrative modelling, biological function, large macromolecular assemblies, molecular dynamics simulations, normal modes, ryanodine receptor, homologous recombination

## 1 INTRODUCTION

Many biological processes depend on the formation of large transient or permanent macromolecular assemblies. These assemblies may harbor signaling functions (e.g., membrane receptor proteins), act as motors for cell motility, membrane crossing, DNA maintenance, or ATP synthesis, or form scaffolds for the cell wall or for communication networks. In recent years, extraordinary advances have been made in both technology and information processing, leading to an avalanche of experimental 3D structures of large biomolecular complexes (Nogales, 2016; Murata and Wolf, 2018; Ziegler et al., 2021), many with resolutions better than 4 Å (Ghanim et al., 2021). The

availability of these structures has enabled unprecedented advances in understanding many important biological processes (Verkhivker et al., 2021). It seems that access to high-resolution structures of all large multimolecular edifices involved in complex biological processes is only a matter of time. Does this mean that we will be able to understand their mechanisms? This question highlights a fundamental challenge for understanding functional biological assemblies: how to consistently integrate experimental structure determination into mechanistic and functional models. From this point of view, integrative modelling can be seen as a tool that goes beyond the structural level to explicitly include questions of biological function and mechanism.

Mechanistic understanding of the function of proteins and other biological macromolecules is usually based on the availability of three-dimensional structural information. The earliest example of this is probably the replication of DNA based on an atomic model of the double helix (Watson and Crick, 1953). Shortly thereafter, the first crystal structure of a protein, myoglobin, appeared (Kendrew et al., 1958), sparking a movement to determine more and more macromolecular structures at ever higher resolution by X-ray diffraction, neutron diffraction, NMR, and cryo-electron microscopy. Currently, there are nearly 200,000 experimental 3D structures of individual proteins and small complexes in the Protein Data Bank. Analysis of this rich source of high-resolution structural data increasingly contributed to the creation of models of proteins for which no experimental structure existed (e.g., Webb and Sali, 2021). This culminates today in the successful application of artificial intelligence to the automatic generation of highly accurate 3D models (Jumper et al., 2021), reflecting the tremendous progress that has been made in the last decade in predicting protein structures from sequences. Taken together, this detailed structural information has helped to elucidate the mechanisms of myriad catalytic reactions and biochemical processes carried out by macromolecules in the cell.

Over the same period, and largely thanks to advances in cryoEM (Nogales, 2016; Murata and Wolf, 2018) and its coupling with integrative modelling (Ziegler et al., 2021), this inexorable push towards high-resolution 3D structural data has attained the scale of large-scale assemblies in the cell. It would thus appear that understanding the mechanisms of these processes is indeed at hand. Yet deciphering the function of large assemblies in many cases defies the “structure illuminates function” logic. This may occur because 3D data is missing for particular regions of otherwise well-resolved structures, thus preventing one from fully exploiting them, or from the absence of essential cofactors or partner proteins needed for their function. Similarly, structural disorder in the complex may be intrinsic to its function, so that a structural average or snapshot provides little insight. The availability of the structures of the endpoints of a given process (reactants, products) and of a few intermediates may not be sufficient *per se* to directly address the underlying biological mechanisms. This could be because the stable intermediates are structurally too far apart to simply infer the transition between them, or because new structural information reveals new gaps in our knowledge. Structural sets

could reflect different experimental conditions or different levels of resolution and thus not readily provide a meaningful overview of the biology in question. In such cases, filling the structural gaps requires the additional effort of integrating complementary experimental and numerical approaches.

The term *integrative modelling* is frequently used to describe obtaining 3D structures of macromolecular assemblies based upon medium-to low-resolution data such as cryo-EM maps or SAXS profiles. Additional information obtained from mutagenesis, NMR, or mass-spectrometry and cross-linking (Faini et al., 2016) has frequently been used to reduce structural ambiguities. Docking has been helpful in predicting 3D structures of large macromolecular complexes starting from those of their known constituents (Inbar et al., 2005; Schneidman-Duhovny and Wolfson, 2020), although many such cases remain a challenge due to complex combinatorial problems. Platforms for integrating such diverse information into structural models propose user-friendly software pipelines that take advantage of the known high-resolution 3D-structures of complex components or their homologs sharing a sufficient degree of sequence similarity (Russel et al., 2012; van Zundert et al., 2016; Mirabello and Wallner, 2017). The outcome of such modelling is often considered to represent metastable conformational substates that exhibit robust stability when subjected to hundreds of nanoseconds of molecular dynamics simulations. However, these criteria are not always sufficient to approach functional questions.

Here, we focus on two representative examples of assemblies where structural information alone is not sufficient to obtain mechanistic insights. The first is the ryanodine receptor (RyR). To date, RyRs are the largest known ion channels, with a molecular mass of over 2.3 MDa. Each subunit of the homotetrameric channel contains more than 5,000 residues. The number of RyR cryo-EM datasets is steadily increasing (Baldwin et al., 2018; Spurgeon et al., 2021). However, although many maps are available that span different conformational substates obtained under different conditions, there are still no complete models of this receptor. The maps lack homogeneity and the necessary coherence that would permit gaining insight into the structure-dynamics-function relationships in this critical channel protein and help uncover missing biological significance. The second case concerns the nucleoprotein filament complex that is active in homologous recombination (HR)— a fundamental biological process that aims at faithfully repairing broken DNA strands (Prentiss et al., 2015; Bell and Kowalczykowski, 2016). This complex comprises a long filament of recombination proteins assembled on DNA strands. The whole assembly undergoes very rapid dynamic evolution that is directly linked to its function, hence the difficulty in obtaining reliable structural information on intermediate states.

We illustrate how these challenges can be addressed using a modelling pipeline (**Table 1**) which begins with relatively simple structural modelling and gradually incorporates steps that require more sophisticated methodologies. The process is explicitly oriented towards arriving at dynamic information which may be used simply to verify the metastability of a state, for example, or to explore changes of the system in response to a perturbation

**TABLE 1** | Integrative modelling tasks applied to two representative examples.

	RyR1	RecA*
generation of missing atom coordinates	yes	no
side chain reconstruction	yes	no
backbone modelling	yes	yes
<i>ab initio</i> structure prediction	yes	no
interactive loop modelling	yes	yes
interactive engineering	no	yes
cofactor/molecular docking	no	yes
MD refinement	yes	yes
MD (meta-)stability verification	yes	no
MD/NM preliminary (or model) exploration	yes	yes

or relaxation process. A given step in **Table 1** may or may not be applied depending on the biological context. In the first system presented, the ryanodine receptor RyR1, we focus on the modelling of “building blocks” consisting of a single macromolecule but in different conformational states, for which coherent modelling, validation and comparisons can be performed. The second example, a RecA-DNA-polymerase complex that executes the final steps of the HR process, concerns the assembly and coordination of building blocks that are already available into functional biological machines that can be used to predict mechanistic aspects of the process. Although this is conceptually more complex than the construction of the building blocks themselves, the two examples share similar logic at different levels, and employ dynamic exploration as part of the modelling and preliminary exploration steps.

## 2 METHODS

The elements described here allowed performing the integrative modelling tasks in our modelling pipeline, which are summarized in **Table 1**. We also detail the protocols we developed, introduce nomenclature, and explain key decisions. The overall methodology is general, but the actual details are specific to each system.

### 2.1 Full-Length RyR1 Models

#### 2.1.1 Multiple Alignments

Three mammalian isoforms of RyR exist in cells: RyR1 predominates in skeletal muscle; RyR2 is the most abundantly expressed isoform in cardiac muscle; and RyR3 is expressed differentially in brain (Takeshima et al., 1989; Nakai et al., 1990; Fill and Copello, 2002; Lanner et al., 2010), endocrine cells and other tissues. At least two of these isoforms, RyR1 and RyR2, share the same structural organization with an overall mushroom-like shape as observed by electron microscopy (Van Petegem, 2015; Zalk et al., 2015; Georges et al., 2016; Peng et al., 2016; Dhindwal et al., 2017). These isoforms have a high percentage (ca. 70%) of sequence identity.

We selected reviewed RyR sequences from the UniProtKB database (Boutet et al., 2016). The entries used for this study were: P11716, P21817, E9PZQ0, P30957, Q92736, E9Q401, Q9TS33,

Q15413, and A2AGL3. ClustalO (Sievers et al., 2011) was used for alignments, WebLogo (Crooks et al., 2004) and Skyalign (Wheeler et al., 2014) were used to assess the presence of conserved residues.

#### 2.1.2 Secondary Structure and Transmembrane Domain Predictions

RyR1 homologues have a high percentage of sequence identity (on the order of 70%), and their secondary structures are well conserved. We were particularly interested in the interspersed disordered region between residues 4254 and 4539, which separates the cytoplasmic from the transmembrane domains. To obtain information on this region, which we call “Big Loop” (BL), we first examined its sequence conservation and residue composition. Since we focus on the RyR1 isoform, we used five homologous protein sequences and performed a multiple sequence alignment using the Clustal Omega software. Secondary structures were predicted using several web servers, such as PSIPRED and PSSPRED (Jones, 1999; Yan et al., 2013). Transmembrane regions were predicted using the TOPCONS and CCTOP servers (Dobson et al., 2015; Tsigirigos et al., 2015). The ion binding sites were predicted via the IonCom program (Hu et al., 2016). For the BL, we used PEP-Fold3 (Lamiabile et al., 2016), particularly for residues 4320 to 4345. Finally, we assessed the presence of an auxiliary transmembrane helix (or helices) and other secondary structure elements by visual inspection of the cryo-EM density maps.

#### 2.1.3 Starting Structures Used for Modelling

The partial structural models deposited in the PDB (Georges et al., 2016) were used as initial data for a series of modelling steps (**Table 2**). Since then, additional structures for the rabbit RyR1 isoform were released, which we may use in subsequent work to compare to the models that we produced. We briefly examined this new structural data, which only became available after our work was completed, and concluded that there was no imperative to repeat the procedure for the purpose of illustrating the integrative modelling pipeline.

#### 2.1.4 Initial Structure Preparation

Due to the inherent conformational flexibility of some regions of RyR1, the cryo-EM structures contain several gaps and are subject to many uncertainties regarding the coordinates of specific residues.

We first examined the coordinates of the “unknown residue” (UNK) annotations in the PDB files. We assumed that the C $\alpha$  coordinates in the PDB files were acceptable starting positions for these residues. These amino acids were assigned initial positions with ROSETTA based on their sequence and then checked visually one by one against the density map. Further, missing side chains and residues of known sequence were modelled with the *de novo* modelling method using the program ROSETTA remodel (Huang et al., 2011).

At the end of the model building process, we checked newly rebuilt residues visually, especially the helical domains such as the junctional solenoid (J-solenoid), using selected buried residues as

**TABLE 2** | RyR1 putative functional states, conformations, ligands, PDB and EMDB identifiers used for this work. # AAs stands for number of amino acids in the corresponding sequence. CFF stands for caffeine. All data are for rabbit RyR1.

State	Conformation	Ligands	PDB Id	EMDB Id	# AAs
apo	closed	-	5TB0	8391	18096
primed	closed-like	Ca <sup>2+</sup>	5T15	8342	19136
Activated/intermediate	intermediate	ATP/CFF	5TAP	8381	18096
activated	open	ATP/CFF/Ca <sup>2+</sup>	5TA3	8377	18096
locked	open	Ryanodine	5TAW	8387	18096

a guide to assess whether the side chains of the residues fit well into the density maps.

Before refining the entire model, several loops or side chains were optimized by a combination of side-chain reorientation and loop placement using in-house routines for flexible and interactive molecular dynamics implemented in the BioSpring tool (Molza et al., 2014) followed by energy-minimization steps using the YASARA Structure software (version 17.12.24) (Krieger et al., 2002). This procedure was particularly useful for refining interfacial loops between adjacent subunits (residues 4290–4299, 3086–3120, 3067–3075, and 4346–4426).

### 2.1.5 Model Refinement

For refining the coordinates of the modelled residues and verifying the side-chain placements, Molecular Dynamics Flexible Fitting (MDFF) simulations were performed using the NAMD software (Phillips et al., 2005; Trabuco et al., 2008). Our initial models were refined using the real-space refinement procedure from the PHENIX package (Afonine et al., 2013; Urzhumtsev et al., 2016) in order to avoid the propagation of geometry errors from the input structure (PDB and *de novo* modelling output files) during the MDFF procedure.

The secondary structures of residues folded into  $\alpha$ -helices or  $\beta$ -sheets were preserved during simulations by imposing harmonic restraints with force constants of 20 kcal mol<sup>-1</sup> rad<sup>-2</sup> for dihedral angles and hydrogen bonds, involving backbone atoms of the same subset of residues using the ssrestraints command.

The NAMD configuration files for the MDFF simulations were generated using the mdff setup command. The MDFF simulation was performed *in vacuo* with NAMD 2.12 using the CHARMM36 force field (Lee et al., 2014). 10,000 minimization steps were performed, followed by slow heating to 300 K and 4 M production steps (4 ns) of molecular dynamics, with a time step of 1 fs/step. A scaling factor of 1 kcal mol<sup>-1</sup> was used to adjust the strength of the influence of the electron density map on the tetramer model during the fitting process. The remaining parameters were defined based on default values in the mdff\_template.namd file. The MD calculations were performed on the ADA supercomputer at the French IDRIS Supercomputing Center.

For reducing remaining outliers, additional refinement steps were performed using PHENIX and YASARA minimization steps under restraints.

### 2.1.6 Side-Chain Refinement in Ligand Binding Sites

As starting positions, we used the coordinates of the calcium and zinc ions, ATP and caffeine molecules, as previously determined by electron microscopy (Georges et al., 2016). For ryanodine, we used a Ryd model provided by Ngo and the coordinates from their MD study of ligand placement (Ngo et al., 2017). The models were created using YASARA. For each model, the most likely protonation state was calculated at pH 7.0. The program YASARA was used for atom typing and hydrogen atom assignment, followed by virtual titration. Sidechains of residues forming ligand binding sites and ligands were then refined using a YASARA macro within the VINALS method (VINA with Local Search) for local ligand docking plus minimization steps. The simulation cell was enlarged by 10 Å in this step.

### 2.1.7 Coarse-Grained Molecular Dynamics Simulations

Coarse-grained Molecular Dynamics (CGMD) simulations were performed using the Martini force field (<http://www.cgmartini.nl/>). The setup uses version v2.2 of the force field (de Jong et al., 2013) in combination with the ELNeDin elastic network (Periole et al., 2009) and the GROMACS 2016.3 simulation engine (Abraham et al., 2015). Proteins were embedded in a membrane composed of DOPC/DOPE at a ratio of 5:3, and the systems were neutralized and then solvated in 150 mM NaCl. More than 100K pseudo atoms were simulated for two systems containing the channel core (with or without the BL region). More than 438K pseudo atoms were simulated for the entire protein. The calculations were performed at the French GENCI “IDRIS” Supercomputing Center and on local GPU clusters.

### 2.1.8 Normal Mode Analyses

Due to the large size of the RyR1 system, we carried out Normal Modes Analysis (NMA) using the Non-Linear Rigid Block (NOLB by Hoffmann and Grudinin (2017)) software, which uses a coarse-grained block approach (Durand et al., 1994). In addition to efficiently analyzing large assemblies, NOLB decomposes the block motions into instantaneous rotational and translational components, whose nonlinear extrapolation allows the computation of periodic trajectories for visualization that preserve structural integrity better than standard linear approaches. The 100 lowest frequency vibrational modes were calculated for each of the superimposed RyR1 conformations using an all-atom elastic network model with an interaction cutoff distance of 10 Å.



Ligand atoms, when present, were included in the analyses. Vibrational modes were sorted in order of frequency  $\sqrt{\lambda_i}$ , with  $\lambda_i$  being the eigenvalues of the mass-weighted stiffness matrix (Hoffmann and Grudin, 2017).

Two NM sets *a* and *b* were compared by examining the subspace overlap, or sum of squared projection of each mode vector from set *a* onto the space spanned by the modes of set *b* (see, e.g., Batista et al., 2010). Only  $C_\alpha$  atom movements were used for these comparisons. The comparison subspace was obtained by orthonormalizing the  $C_\alpha$  mode vectors from set *b* using standard techniques. The value of each vector overlap lies between 0 (no subspace overlap) and 1 (perfect overlap).

NOLB was also used to compute transition paths between pairs of RyR1 conformations by minimizing the RMSD between two input structures (an initial and a target structure) obtained by the linear combination of the calculated modes; these paths were achieved using all 100 computed modes. Scripts were written in Python (general numerical analysis and Chimera scripts) and Tcl (in the VMD environment) to analyze the NM results.

## 2.2 Atomic Model of a RecA\*-DinB Complex

### 2.2.1 Starting Point for the Present Study

The starting point for this system was the results of a preliminary coarse-grained study of the densely packed association complex between the RecA nucleoprotein filament (RecA\*) and DNA-polymerase IV (DinB), which had been obtained by interactive simulations using the BioSpring software also used for the RyR1 modelling above (Section 2.1.4), and which represents each component of a complex system by a variable network of harmonic springs, allowing interactive dynamic testing of possible component orientations and deformations (Molza et al., 2014; Tashjian et al., 2019). In this case the components were the nucleoprotein filament (RecA and DNA) resulting from our earlier strand exchange simulations (Yang et al., 2015) and the crystal structure of DinB bound to its cognate DNA (PDB code 4IRC) (Figure 1B).

### 2.2.2 Initial Model Refinement

The result of the coarse-grained BioSpring flexible assembly process required refinement, as much of the DinB secondary structure was distorted due to the spring network adaptation to the crowded environment of the RecA\* filament. We thus superimposed individual DinB secondary structures taken from the crystal structure of DNA-bound DinB onto the corresponding regions in the preliminary model. One helix showed a partial steric clash with the terminal RecA monomer; the clash was released by laterally displacing the helix as a rigid body by 2 Å with respect to the corresponding BioSpring-positioned segment. Where possible, the loops linking secondary structure regions in the crystal structure were included in the model when this resulted in no steric clashes with RecA proteins; otherwise the BioSpring loop structure was used. All disruptions of the peptide chain structure after this process were removed using energy minimization with NAMD, which restored the standard covalent bond geometry.

### 2.2.3 Introduction of a Guanine Nucleotide Tag

We introduced a guanine nucleotide (resid 11 in the template strand, noted  $G_{11}^{\text{templ}}$ ) as a tag in the template strand in order to be able to identify the site of nucleotide (CTP) addition. The final simulated complex was formed by 17 RecA monomers, the DinB protein and three DNA strands forming a D-loop inside the RecA filament: the ssDNA or primer, with sequence (5'-3') (dT)<sub>53</sub>, the template strand (dA)<sub>10</sub>dG(dA)<sub>63</sub> partly dissociated from the homologous dsDNA, and the displaced strand (dT)<sub>74</sub> from the dsDNA.

### 2.2.4 All-Atom Molecular Dynamics Simulation

MD simulations were performed with NAMD 2.10 (Phillips et al., 2005) and the CHARMM 27 force field with CMAP corrections (Mackerell et al., 2004). A 2 fs integration timestep was used with the SHAKE algorithm, long-range electrostatics were accounted for using the particle-mesh Ewald method, and a Nosé-Hoover-Langevin piston was used for pressure maintenance. The complex was solvated in a TIP3P water box with 0.15 mMol NaCl and progressively heated to 300K. During heating,  $C_\alpha$  and P atoms were harmonically restrained to their initial position with a force constant of 0.5 kcal mol<sup>-1</sup> Å<sup>-2</sup> that was gradually released during 20 ns equilibration followed by a 270 ns production phase. Calculations were performed at the French GENCI CINES Supercomputing Center.

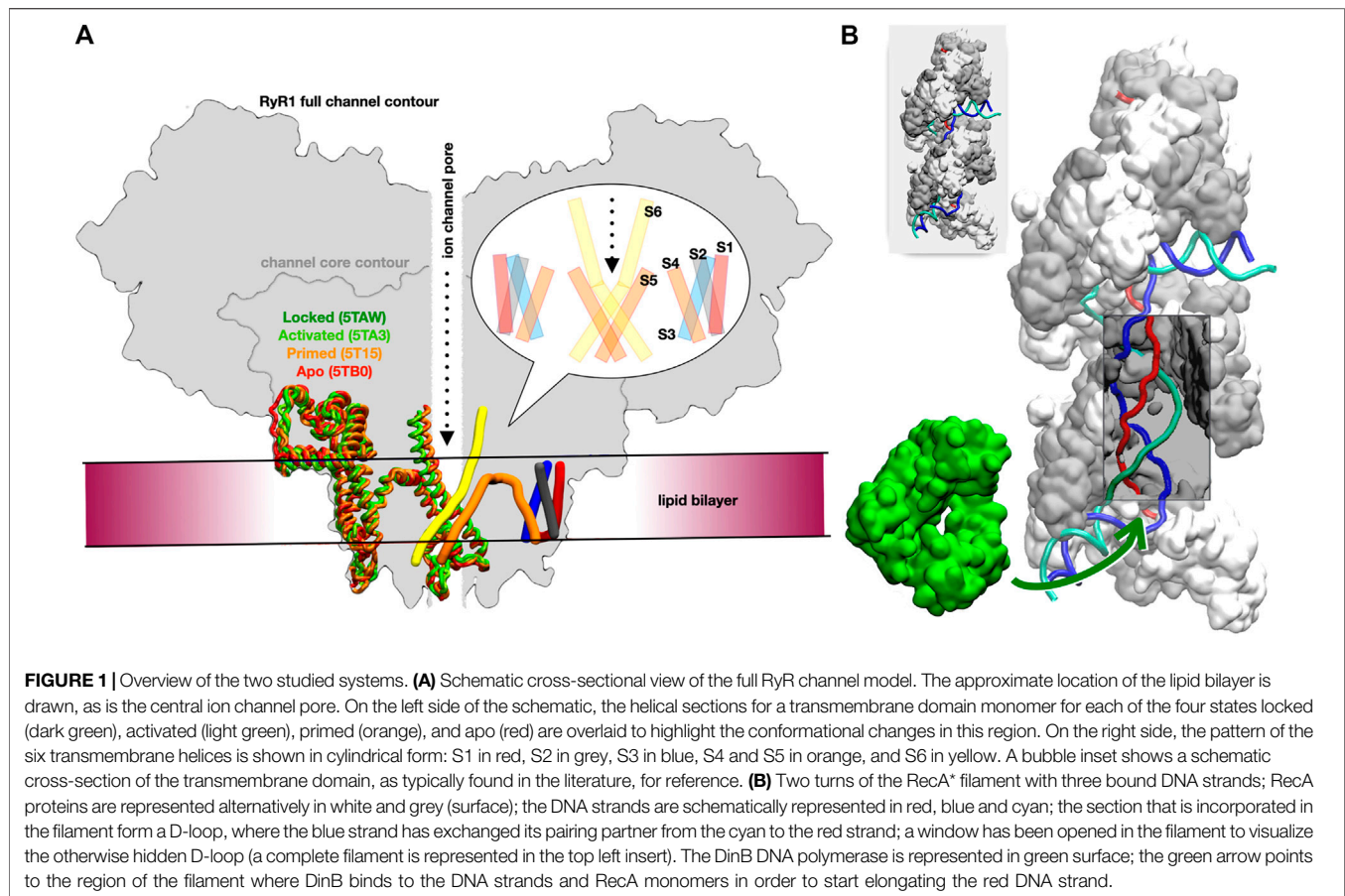
## 3 RESULTS

In this section we describe how we created models using our integrative modelling pipeline by detailing the modelling tasks summarized in Table 1 employed for each of our two example systems. We then present an initial set of interpretations of the models' properties to show how they can provide biologically useful new information. The first "use case," RyR, illustrates how consistent integrative models of a single macromolecular building block can be created. The second use case, RecA, documents how to create larger assemblies from such building blocks and the pitfalls to avoid to create a functionally relevant model. The overall approach, however, is general. The details provided for each system are intended to support the usefulness of the models and to provide avenues for further exploration of these biological systems. These may be reported in future work.

### 3.1 Modelling the RyR1 Ryanodine Receptor Channel and Its Dynamics

Ryanodine Receptors (RyRs) regulate cytosolic calcium concentration, which is critical in numerous signaling pathways. Mutations in these receptors in muscle cells can lead to severe skeletal muscle and heart disease. RyRs belong to the six-transmembrane helix ion channel superfamily (Lanner et al., 2010) and are localized in the endoplasmic and sarcoplasmic reticulum.

Here, RyR1 provides a case study to demonstrate general aspects of the pipeline that lead from deposited cryo-EM



datasets to complete models that can be used for molecular modelling studies on mechanistic and structural aspects. We were particularly interested in 1) addressing issues related to the channel's large size and 2) benefitting from the availability of maps for multiple functional/conformational sub-states, which provide a good test for the usability of the final models and the level of detail and accuracy that can be expected from them.

We focused on the rabbit RyR1 isoform due to the availability of numerous structural and experimental studies on its gating and activation (Georges et al., 2016). Our goal was to test whether reliable atomistic models of the five resolved channel states (Table 2) could be constructed and mechanistically interpreted based on the experimental data. These five states are defined according to Georges et al. (2016): for the apo and primed states, the channel is in a closed conformation, whereas in the activated and ryanodine-locked states, it is in an open conformation. Figure 1A provides a schematic view of the different states of the receptor highlighting the transmembrane and adjacent regions (the intermediate state is not shown in order to reduce visual distraction).

### 3.1.1 Multiple Sequence Alignment

As previous comparative studies have shown (Yan et al., 2015; Efremov et al., 2015), RyR homologs have a significant percentage of sequence similarity, especially in structured regions such as the

solenoid domains, with the highest sequence identity in the transmembrane domain (TMD). For example, human and rabbit RyR1 sequences are found to have 96.6% identity. Clustal Omega analyses revealed a high degree of similarity for the RyR isoforms RyR1 to RyR3 with percentages of more than 73%, while the identities are slightly lower at about 66%. In the divergent regions observed in the multiple alignments, more than 100 residues in the RyR3 sequence between the SPRY2 and SPRY3 domains are absent. Interestingly, even within these divergent regions, some conserved amino acids were highlighted. For example, a glutamate-rich loop (residues 1875–1921), in which some charged residues are conserved (particularly in the sequences of the RyR1 isoform) (Supplementary Figure S1A) or the Big Loop (BL, residues 4254–4539) connecting the cytoplasmic and transmembrane domains (Supplementary Figure S2). DISOPRED (Ward et al., 2004) predicted the BL region to be unstructured. Moreover, it was not resolved in any cryo-EM electron density map available when constructing our models. These regions have a particular amino acid composition that we analyzed further. A sequence logo plot clearly shows pronounced conservation within the glutamate-rich region of the loop at residues 1875 to 1921. For the disordered BL loop (residues 4254–4539), the sequence logo shows a variety of features, for example, the presence of a putative PE<sub>5</sub> Ca<sup>2+</sup> binding site motif, hydrophobic and charged residues

with a polyarginine repeat (that are either conserved or share considerable similarity), and many glycine residues known for their flexibility (**Supplementary Figure S1B**). The multiple sequence alignment with Clustal Omega revealed a high percentage of residue conservation for the BL region in homologs: for example, a 94% sequence identity for RyR1\_BL human and RyR1\_BL rabbit.

### 3.1.2 Clues for an Additional Transmembrane Helix in the Big Loop Region

The multiple sequence alignment raises the question of the structural interpretation of the BL residue range (4254–4539). We wondered about the role of this disordered region connecting cytoplasmic and transmembrane parts. Since no experimental structural data was available, we performed an *in silico* study of this region. To this end, we analyzed the intrinsic disorder and predicted possible secondary structure elements as well as the transmembrane region, which indicated the possibility of an additional transmembrane helix in the BL region (**Supplementary Figure S3**). When we began this study, ambiguous information about the presence of a transmembrane helix was available from cryo-EM. Below, we review the various observations from our investigations that support such a finding.

#### 3.1.2.1 Secondary Structures and Propensity to Embed into the Membrane

PPSIPRED predicted five putative alpha-helices within the big disordered region (see **Supplementary Figures S3A,B**). The TOPCONS and CCTOP server predictions for the transmembrane region predicted multiple transmembrane parts. Interestingly, most tools predicted the same auxiliary transmembrane helix in the region of residues 4320 to 4336 as shown in **Supplementary Figure S3C**. To evaluate these results, we attempted to model some parts of this loop using a *de novo* approach with the PEP-FOLD3 package (Lamiable et al., 2016).

#### 3.1.2.2 Structural Predictions Through De Novo Folding

The Pep-Fold3 predictions indicate that the conserved regions (residues 4310–4318 and other parts of the BL region previously shown in the sequence logos in **Supplementary Figure S1B**) are most likely in helical conformation (**Supplementary Figure S3D**).

#### 3.1.2.3 Unravelling the Role of the BL Region

The results and predictions for the transmembrane region are consistent with the literature, which suggests that some residues in this region tend to fold into an alpha-helix and possibly embed into the membrane (Van Petegem, 2015; Zalk et al., 2015; Du et al., 2002; Van Petegem, 2017; Santulli et al., 2018; Efremov et al., 2015) (**Supplementary Figure S3C**). An additional arginine-rich helix could be localized at the surface of the bilayer (Hristova and Wimley, 2011; Ulmschneider et al., 2017). The BL region could be interacting with other partner molecules or phospholipids or even be implicated in the oligomeric interactions of the RyR. The propensity of some residues to fold into transient TM helices could also impact

function, but this aspect remains somewhat unclear. These hypotheses should be verified experimentally. Comparing the rabbit and human RyR sequences, we can speculate about the role of the BL region. Because of the high identity and similarity percentages, we could use the initial models for this region as structural templates for constructing full-length RyR2 and even RyR3 models.

### 3.1.3 Model-Building Procedure

Despite many available maps for RyR, complete models of the protein are not yet available. This lack of completed models hampers many efforts to obtain insight into the structure-dynamics-function relationship. In addition, RyR serves here as a use case to show how information from the pool of deposited cryo-EM datasets can be used to complete these models. Given that the number of relatively high-resolution datasets is continuously increasing, we hope that the model building procedures developed here will provide a general reference protocol that can be further adapted and optimized for exploring mechanistic and structural aspects of other systems under investigation.

**Figure 2** provides a global schematic overview of the entire modelling procedure. In the following sections, the individual steps for the reconstruction of the receptors are described in detail.

#### 3.1.3.1 Full-Length Models of the Receptor (RyR1) at the Atomic Level

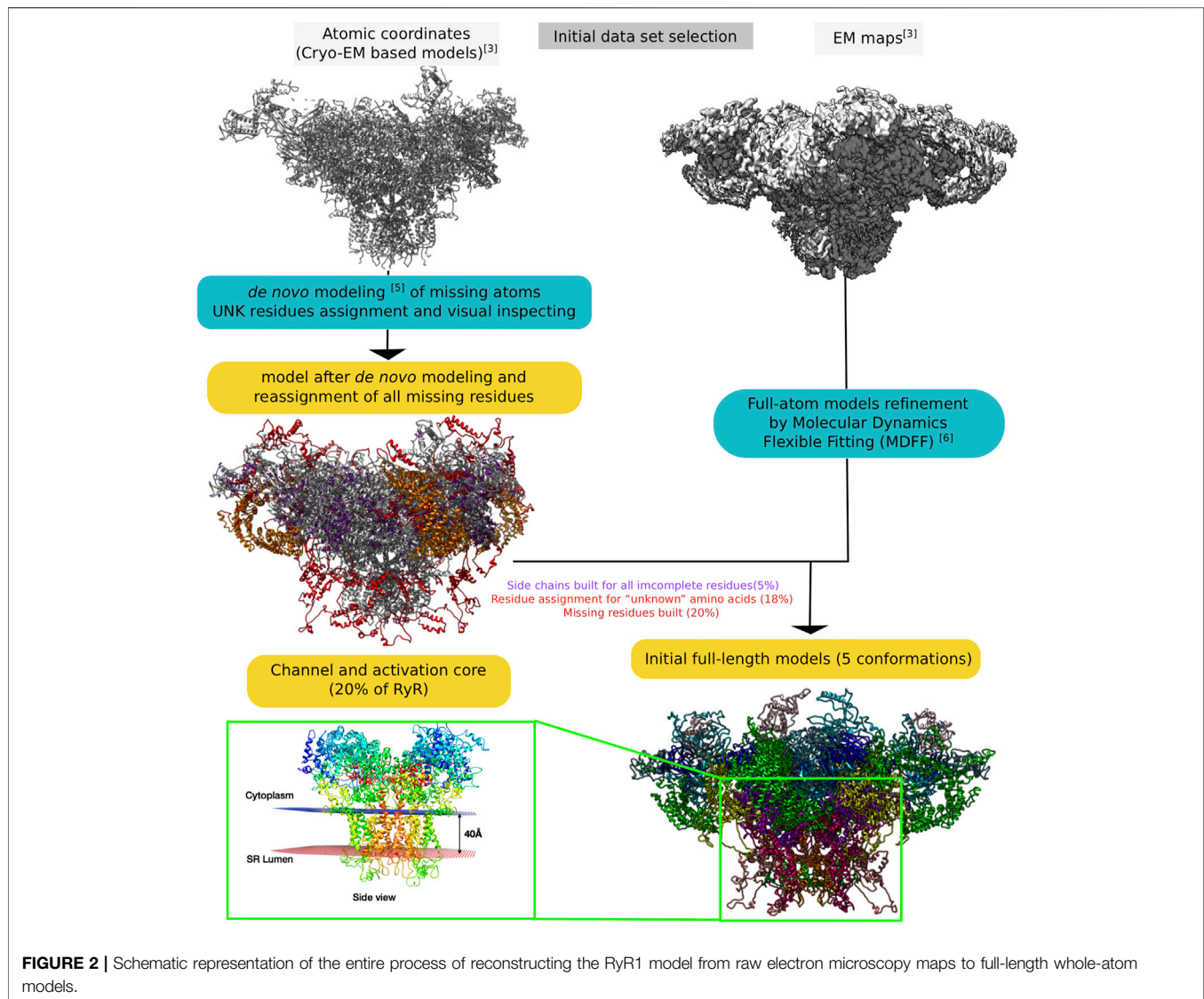
To model missing parts of the protein, several software packages were tested. For the specific use case of the RyR protein, given the challenges presented by its size, the best results were obtained by *de novo* modelling using Rosetta (<https://www.rosettacommons.org/>) to complete the models. The process is shown schematically in **Supplementary Figure S4**.

To optimize the geometry of these initial models for further computations after *de novo* modelling, we applied some model refinement routines using minimization steps within the Yasara software (Yasara Structure 17.12.24; <http://www.yasara.org/>) to remove possible clashes, knots, and some bond errors. We then refined the models for all atoms using the MDFF method (<https://www.ks.uiuc.edu/Research/mdff/>). This step is intended to optimize the newly created loops and side-chain coordinates based on the volumetric data from the experimental electron density maps.

We set up an MDFF routine that allows us to extract a partial volume corresponding to a monomer from a map to fit the model (**Figure 3**). After applying the MDFF procedure to a monomer, the different partial volumes fit well into the full map. However, we found that long loops (e.g., the region of residues 4254–4539) did not converge into the map despite many MDFF runs (**Figure 3** and **Supplementary Figure S5A**). We attribute this lack of convergence to either the large size of these loop regions or to missing densities. Using the BioSpring tool (Molza et al., 2014) at a later stage (see below) improved convergence.

A variant of the MDFF method was required due to the system's size and the computational resources needed. We experimented with a more computationally intensive protocol





**FIGURE 2** | Schematic representation of the entire process of reconstructing the RyR1 model from raw electron microscopy maps to full-length whole-atom models.

to fit the entire tetramer model into the map. Fitting each monomer into a subvolume is a good choice for a big system. However, if sufficient technical resources are available, fitting the whole system is preferable, as it also accounts for the interactions between the subunits.

The protocol is shown in **Figure 3**. It can be summarized for the monomer case as follows: Preparing the subvolume, fitting the monomer model to the subvolume and generating the tetramer model, checking that there is no overlap between subunits, and making any necessary corrections. For modelling the full tetramer (without the option of imposed symmetry), one first has to prepare the tetramer model and files for MDFF, then fit the model to the experimental map (in our experiments, this took about eight times longer than in the monomer case), check that there are no “aberrant” fits, and adjust the coordinates if necessary.

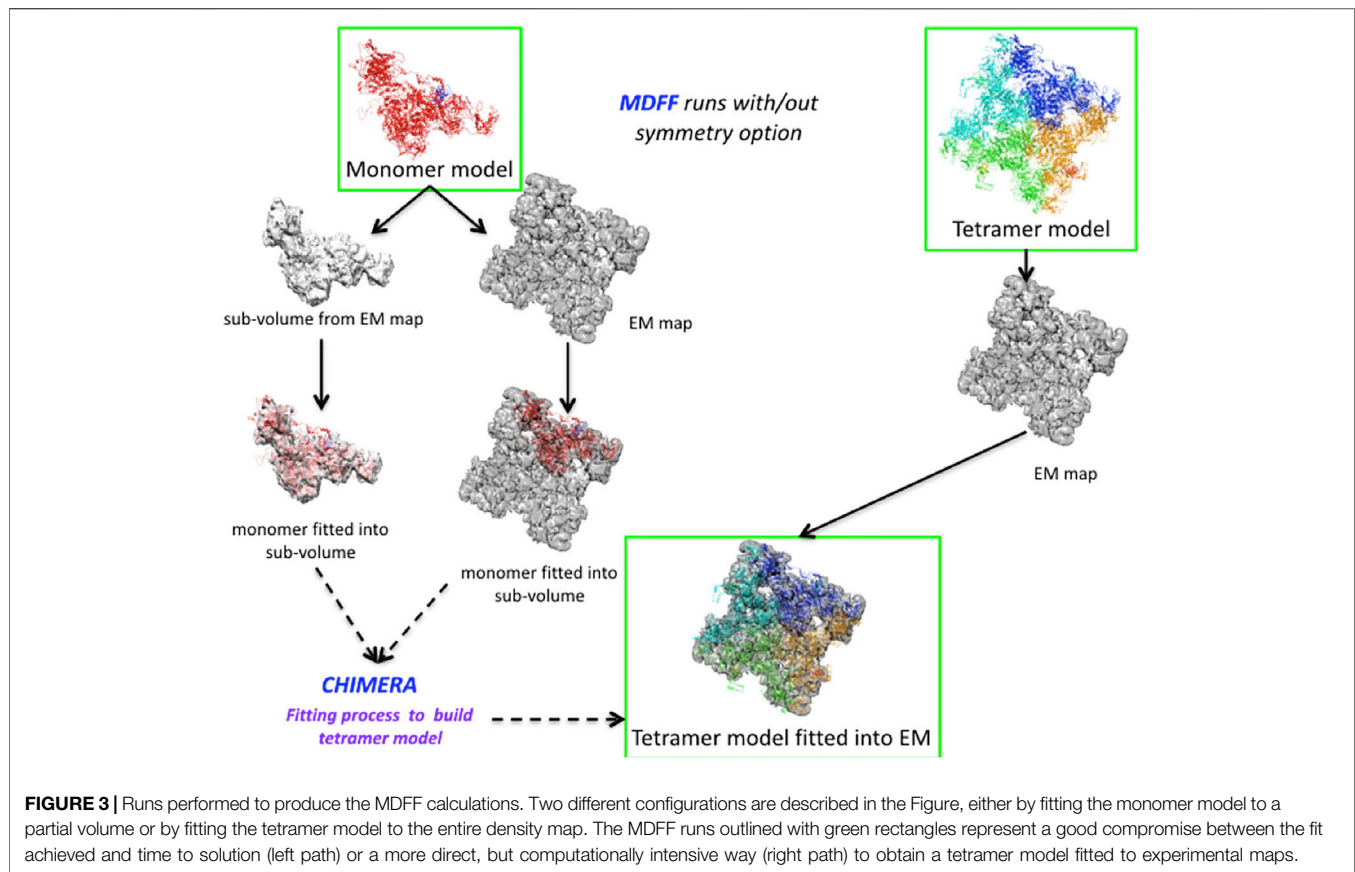
This step was followed by a visual inspection of the residues, especially residues modelled from “UNK” residue types. Some

limitations are due to low-density regions or difficulties in fitting long loops. When necessary, some loops (e.g., in residue regions 4290–4299, 3064–3133) were shifted using the interactive and flexible method BioSpring previously developed in the laboratory (Molza et al., 2014), as shown in **Supplementary Figure S5B**. Namely, these regions were aligned to the corresponding densities of the neighbouring subunits.

In our final models used for subsequent analysis, the additional transmembrane helix was included for the apo, activated and locked states based on the evidence we found in the cryo-EM densities, whereas there was no density to model it into the primed and activated/intermediate states, despite the fact that the helix was present in PDB 5TAP.

### 3.1.3.2 Assessment of Stereochemical Model Quality

Molprobit (Williams et al., 2018) was used to assess the structural quality of the models compared to the starting



**TABLE 3** | Molprobability validation results (% residues) for the models derived from the five structural RyR1 templates used.

State	Structure	Favored regions	Allowed	Outliers	Rotamer outliers	$C_{\beta}$ outliers
apo	5tb0	88.80	11.08	0.12	0.50	0
	noloops model	88.52	11.20	0.28	0.20	0
	full model	86.92	12.75	0.33	0.19	0
primed	5t15	88.24	11.49	0.27	0.45	0
	noloops model	91.25	8.51	0.24	0.06	0
	full model	89.42	10.33	0.25	0.05	0
activated/ intermediate	5tap	89.92	9.90	0.18	0.54	4
	noloops model	89.21	10.46	0.33	0.26	0
	full model	87.18	12.47	0.35	0.22	0
activated	5ta3	89.09	10.76	0.15	0.54	0
	noloops model	89.40	10.36	0.24	0.17	0
	full model	87.32	12.41	0.27	0.17	0
locked	5taw	89.53	10.30	0.17	0.39	0
	noloops model	90.85	8.95	0.20	0.21	0
	full model	88.96	10.77	0.27	0.23	0

PDB structures, distinguishing the full models from those (“no-loops”) without modelled loops (red regions of **Supplementary Figure S4**). The results are summarized in **Table 3** and show systematic improvement for some metrics, whereas others fluctuate or degrade. The  $C_{\beta}$  outliers in the 5tap conformation were corrected by our procedure, and rotamer outliers were systematically reduced for all models and fall within the recommended limit of 0.3. While the backbone

outliers would normally be less than 0.2%, we observed several cases with higher values. The percentage in favoured Ramachandran regions should be above 98%, but does not go beyond 91.25% in the deposited structures in the PDB or in our models. In summary, despite the addition of a large number of missing structural data, the model quality did not significantly degrade, and even systematically improved for some metrics.

### 3.1.3.3 Assessment of the Overall Assembly Through Coarse-Grained Simulations

The modelling steps described above resulted in all-atom structures of the RyR channel that were significantly enhanced with respect to the starting structures (**Supplementary Figure S6A**).

An important question is whether the molecular assemblies generated constitute a viable tetrameric channel in a membrane environment. To address this question, we performed Coarse-Grained Molecular Dynamics (CGMD) simulations. Two types of models of different sizes were used for this study: a **channel core model** centered on the RyR channel and a **full-length model** of the receptor. The channel core includes the core solenoid, the transmembrane (including the pore), and the C-terminal domains; the interaction of all these domains forms the ligand-binding sites of RyR1. This system comprises residues 3667 to 5037 (biological numbering) of each monomer. The complete model contains all residues in the RyR protein (residues 1 to 5037 for each monomer). To avoid bias in the coarse-grained simulations, we experimented with the presence or absence of the unstructured BL region that connects the cytoplasmic with the transmembrane domain. No connections or restraints were added between the monomers of the channel in either model.

Microsecond-scale CGMD simulations of the full RyR channel in DOPC/DOPE membrane and water simulations (500,000 particles, **Supplementary Figure S6B**) showed good stability over this period. A plot of the center-of-mass distance for all pairs of subunits is shown in **Supplementary Figure S6C**. Extended CGMD simulations (1–10  $\mu$ s) were performed for the RyR1 channel core (3667–5037 of each monomer), the channel core excluding the BL region (deleted residues: 4254–4539) for the five RyR channel states of **Table 2**, and the full-length apo-state model. The protein architecture remained stable during the simulations, as shown in **Supplementary Figure S6**, confirming the interactions between the subunits of the tetramer and thus indirectly validating the model reconstruction and monomer interfaces. Specifically, the six helices (**Figure 1**) maintained their integrity during MD, with the highest fluctuation amplitude in the loop regions, especially in the S1-S2 loop. The distance between each monomer of the channel remained stable, with the radius of gyration varying between 10.6 and 10.8 nm (the variation may be due to the interaction with lipids). The bilayer thickness fluctuated little ( $38.7 \pm 0.3$  Å). In the future, we plan to transfer the systems from CG to the atomistic level and perform molecular dynamics for all atoms to achieve an accurate structural characterization of the RyR1 dynamics.

## 3.1.4 Exploration of the Dynamics of the RyR Channel Models

### 3.1.4.1 Studying RyR1 Deformations Through Normal Mode Analyses

Normal modes analysis (NMA) allows one to examine the dynamics of a given structure given the 3D coordinates of its component atoms, which in this case are obtained from our

models. NMA is based on a harmonic approximation of the molecule's potential energy in the region of a minimum in the potential energy surface, which is expressed as a function of changes in the  $3N$  coordinates of the structure's  $N$  atoms. It provides a detailed description of the vibrational dynamics associated with small perturbations of the minimum-energy structure. NMA notably provides a set of normal mode vectors, each of which consists of  $3N$  components representing the amplitudes of atom movements for a single vibrational mode, plus the set of associated frequencies or deformation energies. In practice, the low-frequency (or low-energy) modes are of the most interest, as they have been extensively correlated with directions of conformational changes and dynamics in biological macromolecules (e.g., Tama and Sanejouand, 2001; Rueda et al., 2007).

Here we used NMA to better understand the dynamics of our models of the ryanodine receptor in its different conformational states and how these movements differ from one state to another. For such large molecules, the analyses were facilitated through the use of the NOLB software (Hoffmann and Grudinin, 2017). We performed NMA on the full-length RyR-1 models and the restricted channel regions after removing the modelled loop regions (structure “no-loops” in **Table 3**). These calculations were carried out on the five conformational states of RyR-1 (**Table 2**): the closed or closed-like states apo and primed, the intermediate state, and the open states activated and locked.

A variety of definitions for RyR's motions have been used in the literature. Here we defined specific geometric characteristics of protein subdomain movements that characterized RyR vibrational modes. Three types of movements in particular were used to categorize the motions observed when visualizing the NM vibrational modes, and are designated here as twisting, breathing, and blooming motions. **Supplementary Figure S7** shows schematic diagrams of these motions in RyR.

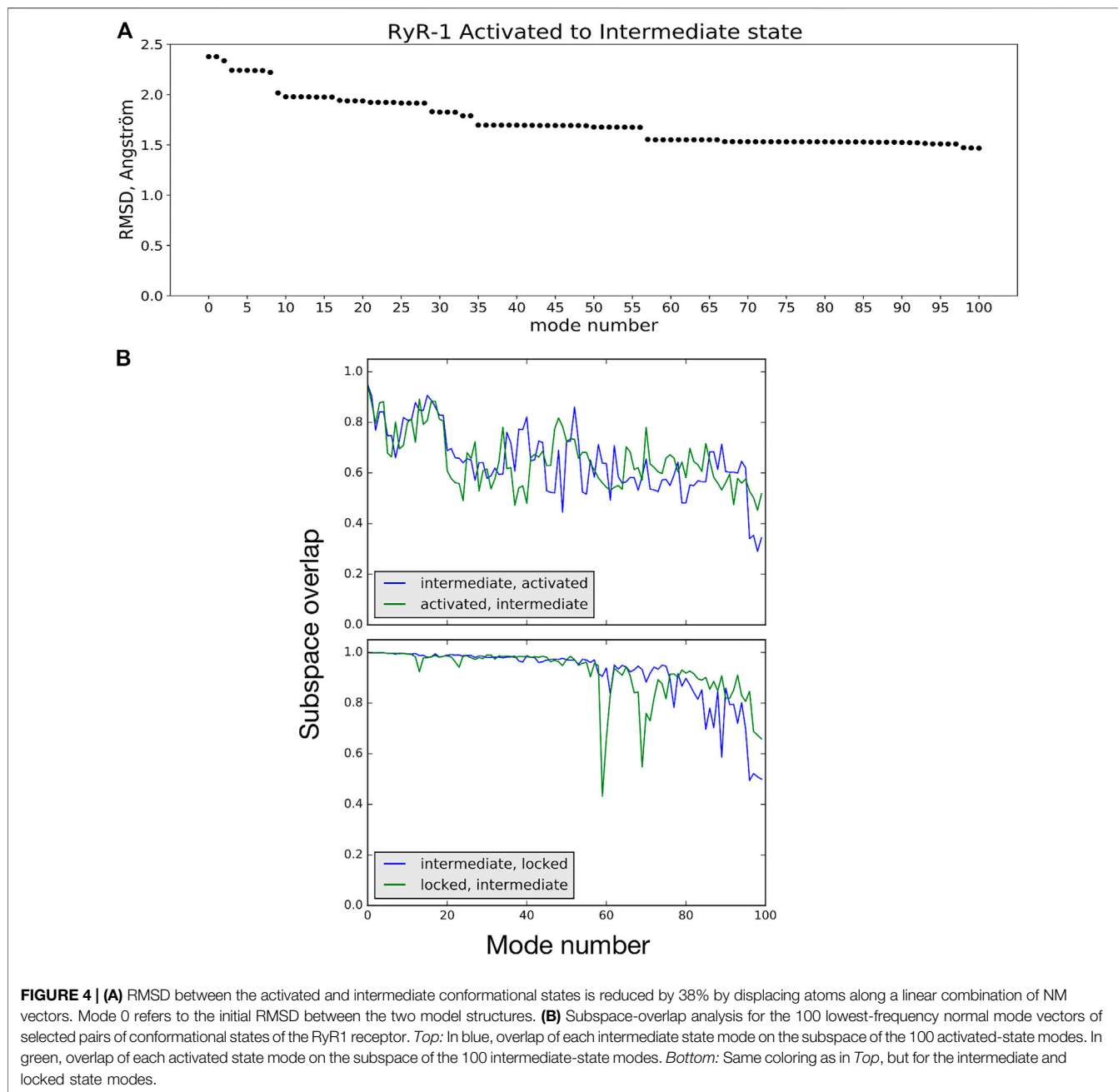
### 3.1.4.2 Contributions of NM Movements to RyR Conformational Transitions

For the following analyses, we used the models restricted to the channel region, which encompasses residues 3667 to 5037 of each chain. The structural similarity of the different channel states was assessed by the  $C_\alpha$  RMSD after superposition for all model structure pairs (**Table 4**). Of the channel regions of the five models, the primed and intermediate structures are the closest ( $1.4$  Å  $C_\alpha$  distance), followed by the models of the locked and activated states ( $1.6$  Å), respectively. All other model structure pairs are between  $2.4$  and  $2.9$  Å apart.

NMA suggests how the vibrational dynamics of the macromolecule may contribute to a given conformational transition. One way to see this is by calculating the linear combination of low-frequency vibrational mode vectors (atom movements) of the starting structure *A* that minimizes the least-squares distance to the final structure *B* (Hoffmann and Grudinin, 2017; Moal and Bates, 2010). **Figure 4A** shows an example of using this approach to studying the activated  $\rightarrow$  intermediate transition. The linear combination of normal-mode vectors for the 100 lowest-frequency modes of the activated conformation reduces the RMS distance to the intermediate

**TABLE 4** |  $C_{\alpha}$  rmsd in Angstroms for pairs of RyR model channel core models (calculated without the flexible BL region).

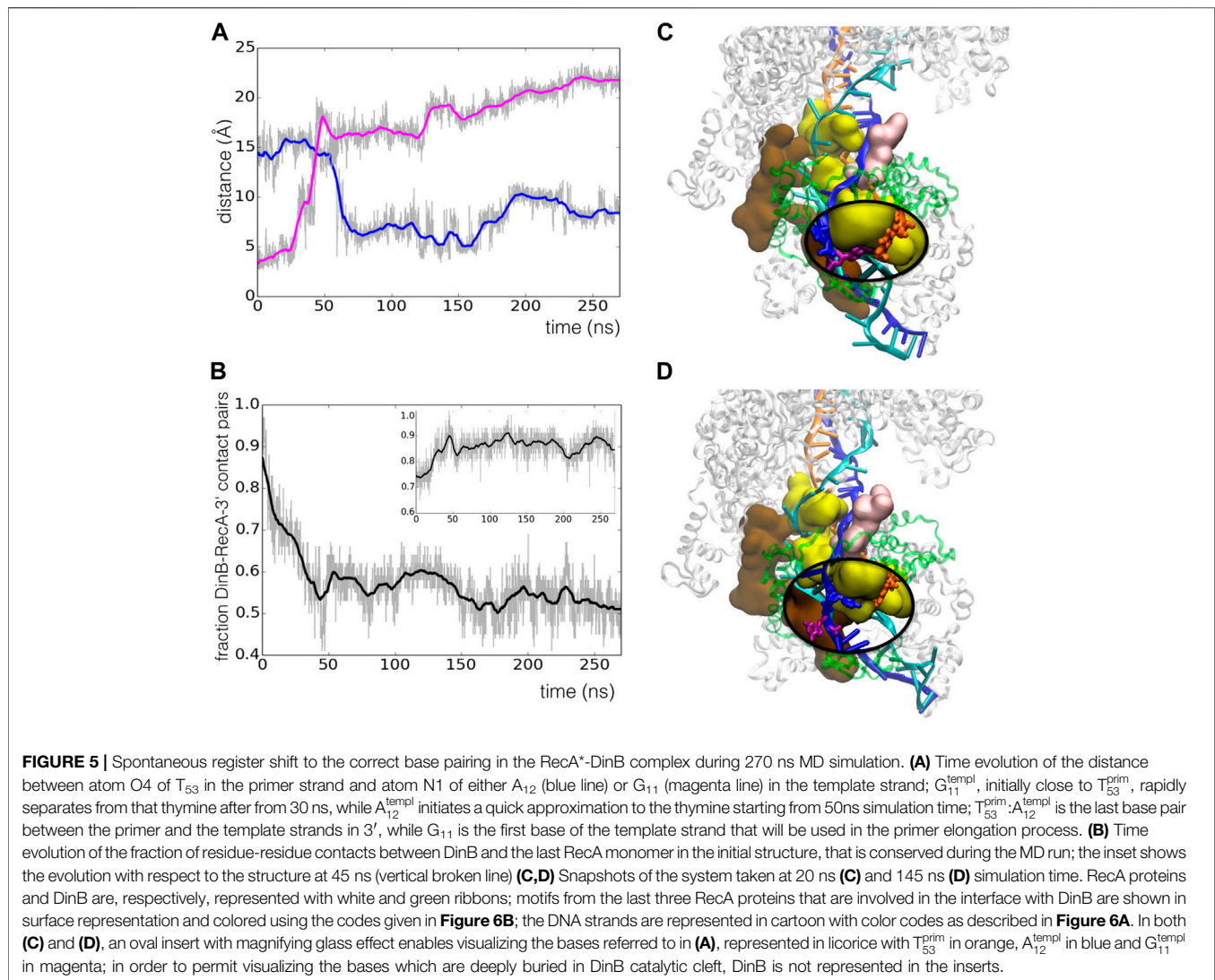
	Apo	Primed	Activated/intermediate	Activated	Locked
Apo	-	2.8	2.5	2.7	2.9
Primed		-	1.4	2.5	2.6
Activated/intermediate			-	2.4	2.5
Activated				-	1.6



conformation from 2.38 to 1.46 Å (38%), while the first 50 reduce the RMSD by 29%. As is often the case for conformational changes in biological macromolecules (Moal and Bates, 2010),

a relatively small number of low-frequency NM movements participate in this transition; mode 9 alone contributes 10% of the total RMSD decrease. A short animation of the motion





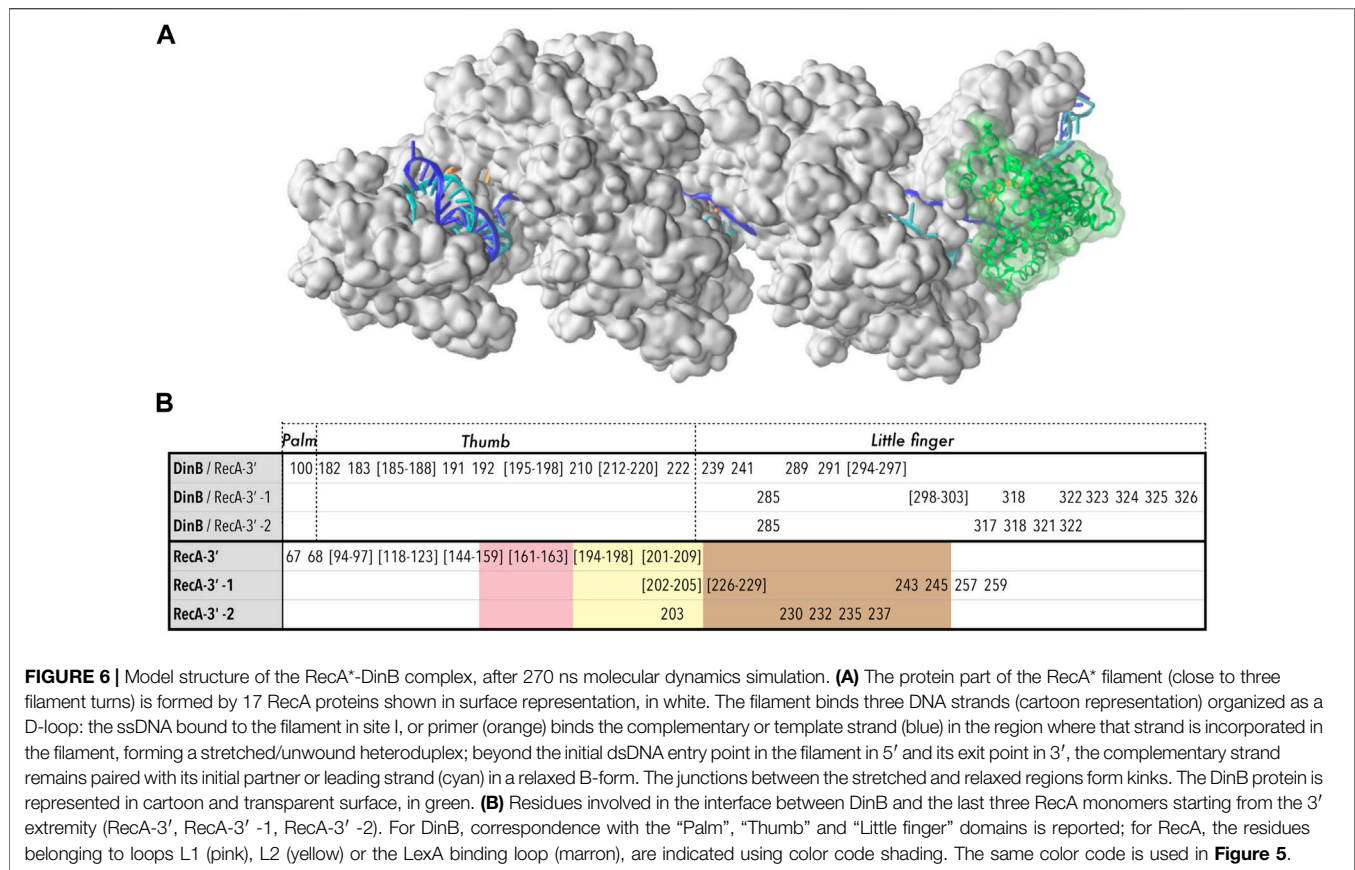
described by the low-frequency NM contributions to this transition is available (Robert and Molza, 2021). These NM contributions describe a longitudinal compression of the channel: bending of helix S6c and movement of the C-terminal domain act to reduce the opening on the cytoplasmic side, while the transmembrane domain twists about the channel axis and descends relative to the rest of the receptor, reducing the volume of the central cavity enclosed by helix S6 (breathing motion).

The NMA thus showed that the intrinsic, low-energy vibrational movements of the RyR channel in the active conformation already encode atom movements that tend to lead to a conformational transition to the intermediate state. This could be then used to obtain further insight into associated features such as domain movement, ligand-binding cavity opening/closing, etc. The NM motions thus reveal an energetic and dynamic rationale for previous depictions of conformational differences between cryo-EM structures (Georges et al., 2016; Willegems and Efremov, 2018), depictions that alone do not

provide information for judging the energetic or dynamic favorability of inferred movements.

### 3.1.4.3 Comparing Vibrational Motions of the RyR1 Channel as a Function of State

NMA can also be employed to assess the similarity of two conformational states of a protein in terms of their vibrational dynamics. To demonstrate this, we extracted the movements of common C<sub>α</sub> atoms from the all-atom modes of different channel models calculated using NOLB. These will be referred to as C<sub>α</sub> modes in the following. The plots shown in **Figure 4B** compare the intermediate and activated state models (top) and the intermediate and locked state models (bottom). In these plots, each of the 100 C<sub>α</sub> mode vectors from one model *i*, arranged in order of increasing frequency, is projected onto the space spanned by the first 100 vibrational modes of the second model *j*, and vice versa. The intermediate and active state models can be seen to share about 70% of their low-frequency mode subspaces. On the other hand, the intermediate and locked states share a higher



percentage of their low-frequency vibrational movements, particularly for the first 50 modes. This example demonstrates that comparing the vibrational dynamics of coherent models of the different states of the RyR receptor clearly provides a richer comparison than simple rmsd: the intermediate state is about the same rms distance from the activated state (2.4 Å) as it is from the locked state (2.5 Å), but the differences in vibrational dynamics of the models are more nuanced. Further investigation into comparisons of the RyR1 vibrational dynamics as a function of state are ongoing.

### 3.2 Integrative Modelling of DNA Polymerase Binding to RecA Nucleofilaments

We now provide an example in which refinement and initial validation of a structural model and dynamic exploration of the model go hand in hand. Here, the challenge is to validate the construction of an intermediate species in the final stage of homologous recombination (HR): the coupling of RecA-induced DNA strand exchange to DNA synthesis. As will be seen, the dynamic exploration is performed using classical molecular dynamics simulations rather than the harmonic dynamics (normal modes) approach employed above for the ryanodine receptor. Molecular dynamics appears to be well adapted to probing the modelled intermediate in this highly

dynamic process because of the strong mechanical coupling among the components.

It is useful at this point to briefly present the HR process, largely conserved from prokaryotes to humans, along with the key players of this process relevant to the present modelling. Homologous recombination is a highly complex, multicomponent process that intervenes in the cell to assure the repair of double-strand DNA breaks, which are typically lethal (Bell and Kowalczykowski, 2016). The two broken ends are first processed to form a single-stranded DNA tail (ssDNA) on either side of the break. To repair the break, a helical nucleoprotein filament is formed by the oligomerization of protein monomers on these ssDNA tails. In bacterial HR, a well-studied model system, this protein is RecA. The nucleoprotein filament systematically binds and searches double-stranded DNA (dsDNA) in the genome to find a sequence match with the ssDNA. Within the HR filament, the two DNA species (ssDNA and dsDNA) are positioned near each other, allowing mutual probing of their sequences for homology, followed by base pairing and strand exchange in the case of homology (Prentiss et al., 2015; Bell and Kowalczykowski, 2016). This probing and proofreading process takes advantage of frequent reversals of strand exchange (Danilowicz et al., 2021), but this phase terminates when a DNA polymerase takes over, pushing the strand exchange reaction towards irreversibility (Lu et al., 2019) by using the dsDNA complementary strand as a

template to start elongating the ssDNA at its 3' extremity. The subject of the modelling in this example is the interaction of the RecA-DNA filament with the DNA polymerase DinB.

### 3.2.1 Building-Block Scaffold

The modelling steps described here consist of the refinement and validation of a coarse-grained scaffold structure defined in a preliminary modelling study (Tashjian et al., 2019) of the RecA\*-DinB complex, which itself was built on experimental evidence that the coupling between recombination and synthesis occurs via direct binding of the DNA polymerase DinB to the RecA\* filament (RecA nucleoprotein filament with three bound DNA strands) (Henrikus et al., 2020) (see also Godoy et al., 2007). This structure was a coarse-grained model obtained using BioSpring, an interactive modelling facility also used in the RyR1 modelling presented above. The structure also integrated a structural building block RecA\* derived from the model of early RecA-ssDNA-dsDNA intermediate created in 2015 (Yang et al., 2015). Indeed, the striking similarity between that RecA\* model and the recently published Cryo-EM structure of short post-strand-exchange intermediates (Yang et al., 2020) provided further impetus for the current integrated modelling.

### 3.2.2 Coarse-Grained to All-Atom

While the coarse-grained representation used in BioSpring presents a direct correspondence to all-atom representation, which allows easy recovery of an all-atom model, that model needed refinement. This was notably the case for the DinB building block, for which the spring network used in BioSpring had allowed distortions in the secondary structures during the tight assembly process.

We reintroduced individual secondary structure elements and loops taken from the DinB crystal structure as described in the *Methods* section; thereafter energy minimization was sufficient to regularize their geometries.

### 3.2.3 Mechanical Coupling Within RecA\*-DinB Complex

We subjected the all-atom starting model to 270 ns of molecular dynamics simulation in explicit solvent. Over the timecourse of the simulation, the filament flexes somewhat, with the overall end-to-end distance reducing by 10%, but no tendency for dissociation of the polymerase was observed (**Figure 6A**); the total number of DinB-RecA\* interface contacts decreased by about 15% in the first 70 ns, then remained stable throughout the last 200 ns of the trajectory (**Supplementary Figure S8**). This indicates that the large majority of DinB/RecA contact regions from the preliminary modelled structure (Tashjian et al., 2019) were preserved in the highly crowded environment of the complex.

However, we observed spontaneous relaxation that took place during the simulation trajectory through concerted rearrangements of RecA motifs and DNA strands. The starting model presented a mismatch in the last heteroduplex base pair in the 3' end due to an artificial shift in the heteroduplex register resulting from the preliminary modelling process, which had shifted the last several bases of the template strand with respect to

the primer, such that  $G_{11}^{\text{templ}}$  interacted with  $T_{53}^{\text{prim}}$ , the last 3'-base of the primer strand (**Figure 5B**). From the point of view of sequence,  $G_{11}$  is the first base of the template strand that should assist the synthesis of the first added nucleotide in the polymerase catalytic site to extend the primer sequence (**Figure 5C**). **Figure 5A** shows how the mismatched  $T_{53}^{\text{prim}}:G_{11}^{\text{templ}}$  heteroduplex base pair spontaneously separated after about 50 ns, rapidly followed by the formation of the correct  $T_{53}^{\text{prim}}:A_{12}^{\text{templ}}$  base pair (shown in **Figure 5D**). This pairing shift comes with a sliding of the DNA template strand backbone (in blue in **Figures 5C,D** along the L2 loop of the last RecA monomer (in yellow in **Figures 5C,D**). This loop is deeply inserted in the DinB catalytic cleft. **Figure 5B** shows that this sliding occurs simultaneously with a shift in the interface between DinB and the terminal RecA monomer, that also takes place at 50 ns. At the same time, the interface loses up to 50% of its initial residue pair contacts, stabilizing to a new network of residue-residue contacts that is 86% conserved on average during the last 225 ns of the simulation (**Figure 5B**, insert). Recent studies, both on RecA filaments and on heterodimers, showed that stable interfaces typically conserve 70–90% of their interaction contact network when simulated at 300 K, due to thermal movements (Boyer et al., 2019; Prévost and Sacquin-Mora, 2021).

### 3.2.4 RecA\*-DinB Interface

**Figure 6B** lists the residues that participate in the interface between DinB and the last three RecA monomers at the filament 3' extremity. The list has been established from a MD structure at 120 ns simulation time, therefore after the reorganization took place. Two DinB domains mainly participate in the three DinB/RecA interfaces, namely the thumb and the little finger domains, with the palm domain showing only marginal contribution (for definitions of these domains see **Figure 6**). Decomposing the contributions of DinB domains and RecA motifs testifies to the intricate association of DinB within the filament interior. While the thumb domain only interacts with the last RecA monomer at the 3' end, the little finger domain contributes to all three DinB-RecA interfaces. Three residues of the little finger domain, namely Arg85, Lys318 and Trp322, simultaneously interact with the two RecA monomers situated 5' of the terminal RecA. On the RecA side, the interface between DinB and the terminal RecA encompasses most of the L1 and L2 loops, flexible loops that participate in the binding of the DNA strands in the HR filament. Although the L2 loop is deeply inserted in the DinB catalytic cleft, its presence remains compatible with that of the primer and template strands while allowing strand adjustments during the MD trajectory as described above. L2 loops of the two non-terminal RecA monomers also partly contribute to the interface with DinB, which also includes the so-called LexA-binding loop, a hairpin motif that contributes to forming the secondary DNA binding site in the filament.

The model we present here of a putative intermediate of association between the RecA nucleoprotein filament and the DinB DNA polymerase offers a link between the recombination and the DNA synthesis activities in HR. *In vitro* studies have previously demonstrated that DinB associates with isolated RecA

proteins, and the DinB interface with RecA has been characterized (Godoy et al., 2007) (See also **Supplementary Figure S9**). Most interestingly, the regions (or patches) of the DinB surface that interact strongly with RecA were mainly observed within the catalytic cleft, in the little finger (patch *Pa1*: 287–298) and in the palm domains (patch *Pa2*: 145–160). Our model of DinB interacting with RecA\*/DNA identified patch *Pa1* as a major contributor to the interface between DinB and the last RecA monomer (**Figure 6**). More exactly, *Pa1* interacts with the L2 loop of that monomer, which is deeply buried in the catalytic cleft. In addition, while patch *Pa2* does not directly contribute to the interface in our model, some of its residues such as Phe147 are found within 6 Å of that same L2 loop. Further approach of the *Pa2* patch to the L2 loop would appear to be sterically hindered by the neighboring RecA monomers. These observations suggest that DinB interacts with RecA in a similar way whether the RecA is part of the nucleoprotein filament or free in solution (Godoy et al., 2007).

## 4 DISCUSSION

The integrative modelling pipeline we describe here, as summarized in **Table 1**, aims at achieving coherence, completeness, and (meta-)stability in the generated models.

Coherence was a constant concern for building the RyR1 models starting from Cryo-EM data generated in different conditions. It was ensured by separately applying the modelling pipeline to each of the input data sets, resulting in five complete all-atom reference models of functional substates ranging from open to closed conformations. In order to take into account the enormous size of assemblies such as the RyR1 receptor, our protocol focuses on providing a general approach with computationally efficient steps such as interactive modelling and coarse-grained refinement. The pipeline also provides the possibility to limit the computational cost (e.g., by fitting to cryo-EM sub-maps) when performance might be critical. It should be noted that our protocol differs from that of Heinz et al. (2018), whose focus was on quantifying ion permeation and identifying pathways through the channel core. That work used molecular dynamics simulations as the main tool to support the refinement process, which is computationally expensive. It focused on modelling the closed state based on cryo-EM data from (Efremov et al., 2015; Yan et al., 2015; Zalk et al., 2015), while an open-state model restricted to the central and channel domains was generated from the closed-state model using MD guided by the cryo-EM data of Georges et al. (2016).

### 4.1 Missing Regions and Interactions

The RyR1 receptor also illustrates how this pipeline addresses the challenges of obtaining complete models from incomplete structural data. All available experimental density maps of RyR1 lack density for large regions of the sequence. This may indicate disorder or flexibility in these regions. However it is now well recognized that even intrinsically disordered, flexible protein regions may play important functional roles. In tackling this

problem, we observed that fully automated protocols could not handle such situations. For example, the use of MDFF was not always sufficient for generating the missing regions. Human interventions were required to make decisions and optimize the models, using available experimental data and additional sequence analyses. This phase of the modelling also highlighted the important question of the role of the BL region in RyR1, which exhibits distinct patterns and a high degree of conservation but for which no structural data is available. We hypothesize that this region contributes to the interaction with partner molecules, phospholipids, or other RyR1 subunits. These steps in the pipeline provided evidence in RyR1 for an additional transmembrane region, consistent with suggestions in the literature (Georges et al., 2016; Van Petegem, 2017). The additional TM helix was stable in the MD simulation step and appeared to reinforce intra- and inter-subunit packing of the transmembrane region by providing additional bridging interactions between neighboring subunits as well as between subunits and the membrane. Our initial models for this region could be used in further studies as structural templates for building full-length RyR2 and even RyR3 models.

We emphasize the power that interactive manipulation and visualization offers to this part of the integrative modelling pipeline, in addition to traditional modelling and visualization tools, when confronting such complex problems. Interactive approaches have existed for some time, but have not typically been present (or acknowledged) in integrative modelling studies. In cases such as the RyR1 and the RecA homologous recombination systems described here, the overall modelling pipeline involves results from automatic procedures that are explicitly coupled to, interpreted, and filtered through human interaction. For example, both the RyR1 receptor and the RecA-DNA-DinB complex required assembling entangled regions in the modelled complex. The release of steric clashes in many such cases is a tedious task but one that can be handled efficiently using interactive simulations. It might be pointed out that, in modelling, human input is ubiquitous and nearly unavoidable; a common example is the choice of restraints to be applied in modelling, even in such simple cases as energy minimization. The underlying hypothesis is that the ability to manipulate parts of the model as it is being built—to incorporate feedback on the modelling and to make decisions when there are multiple options or ambiguities—will enable the development of better models. Here, we made use of our in-house tool BioSpring at several points in the pipeline. Such approaches have occasionally been documented in the literature (Molza et al., 2014; Croll, 2018), but currently remain marginal. Since integrative modelling often means seeking the best compromise from different and sometimes conflicting sources of information about a given target, the pipeline we have developed explicitly acknowledges that the human operator in the modelling loop may be relied upon to make informed decisions in accordance with the biological data and the physical laws at play in the chosen modelling approach, e.g., in molecular dynamics simulation, constrained minimization, etc. (Lanrezac et al., 2021).



## 4.2 Roles of Assessment and Exploration of Model Dynamics

Simulating the dynamic nature of structural intermediates in the modelling process also plays an important role at different steps of our pipeline, where it serves multiple but distinct purposes. Molecular dynamics simulation, whether obtained using all-atom or coarse-grained representations, allows checking the (meta)-stability of a generated model, of a folded protein for example, over a chosen timescale, as in the RyR1 receptor modelling. It also permits identifying dynamic evolution of models, which reflects plausible local dynamics and structural changes, as proved valuable in the RecA\* system. However, even without costly exploratory MD simulations, normal modes analysis of model structures can provide insight into differences in their intrinsic vibrational dynamics depending on their functional state. For example, comparing the low-frequency vibrational modes for two different conformational states can provide information on large-amplitude, low-frequency mobility differences (e.g., Thomas et al., 1996; Batista et al., 2011). In the case of RyR1, we could examine such differences by exploiting the coherency obtained by modelling the different states of this receptor in order to establish comparisons. As shown, NMA also offers insight into the directions of conformational transitions between two conformational states, which reflects the power of looking to vibrational mode directions for understanding conformational change in biological macromolecules (Tama and Sanejouand, 2001) and predicting protein-ligand and protein-protein interactions (Moal and Bates, 2010).

Dynamic exploration by MD adds further value to integrative modelling through refinement of modelled structures, which can itself provide both better models and valuable insights into function. For example, when multiple building blocks, each refined independently of the others, are combined into a higher-order assembly, there is unavoidably a stress that develops at the interfaces. Such stress may be classified into two categories. The first is what might be termed a “residual” stress that must be resolved in order to obtain a stable interaction. Examples of this were seen in the steric clashes in the RyR receptor, and similarly in the interaction of the DinB crystal structure with RecA-DNA, when assembling subunits in which loops and other motifs in the interface regions had been modelled independently. In our work these stresses were resolved by interactive molecular dynamics approaches that induced little overall change in the geometry of the assembly itself. In contrast, the RecA system revealed a second category of stress that may be linked to the function of the RecA-DNA-DinB assembly (Lu et al., 2019). Evidence of this “mechanical stress” was seen in the resolution of the mismatch of the last basepair of the heteroduplex DNA bound to RecA through unbiased MD simulation, via a register shift in the DNA pairing which resulted from a collective movement that spontaneously reduced the mechanical stress on the stretched DNA strands in the complex. This stress has been shown to play a major role in the recombination process

through single molecule experiments (Conover et al., 2011; Danilowicz et al., 2012, 2014). A precedent for the functional role of stress was also seen in our earlier modelling study of the RecA-DNA interaction at the initial stage of the HR process, in which dynamic exploration of an integrative model resulted in spontaneous strand exchange (Yang et al., 2015), essentially mimicking the very fast event that stands at the heart of the HR process. Such functional conversion of mechanical stress could not have been deduced from static models such as the Cryo-EM structures recently published in Yang et al. (2020). The MD results obtained in the present study similarly suggest that tension-induced collective rearrangements in the DinB-RecA\* system may play a role during DNA synthesis in the last step of homologous recombination.

Our modelling pipeline is seen thus to place fluid boundaries between validation, refinement and dynamic exploration steps. Indeed, information obtained in the course of an integrative modelling workflow can be useful to orient further exploration of functional mechanisms related to the complex under study. This was the case for both systems presented here despite their very different characteristics. In the case of the RyR1 channel, validation of the ensemble of conformational substates through the NM analysis provides information about the dynamic relationship between these conformational states and evidence of transitions between them. This information is a step in understanding the nature of the transition pathways between the functional states. In the case of the HR complex, the observations that we made during refinement appear to provide important indications for further mechanistic exploration, while validating the model in terms of its capacity to demonstrate the transmission of stress in the DNA to other components of the assembly.

More generally, workflows such as the one presented here provide insights not only into the structure and sequence properties of studied complexes, but also into their dynamics and function, as is required for drug design. Models obtained in this way can be used, for example, to infer putative drug-binding pockets. But the dynamics of pocket accessibility affects its druggability, and thus dynamic information clearly adds insight, for example for the development of new molecules aimed at restoring the closure of the mutant RyR1 channel without altering its function.

## 5 CONCLUSION

The increasing flow of experimental structural data calls for a new class of integrative modelling workflows combining a broad range of tools. We have shown that integrative modelling of individual macromolecular building blocks, even for very large systems such as the RyR ion channel, is now possible. Ultimately, the results obtained from such approaches on the RyR family will pave the way to improve our understanding of allosteric long-range gating of channel opening and ligand binding effects, which are essential for drug development to treat RyR channelopathies. Further integrative modelling

applications involve how such building blocks are assembled into higher-level organizations. Our case study of the complex homologous recognition system shows that such approaches, requiring tailored modelling tools, can go very far in the exploration of structure-function questions in a reliable fashion, as the building blocks are mechanically coupled. This procedure permits emitting testable hypotheses for challenging systems that are difficult to address experimentally.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## AUTHOR CONTRIBUTIONS

MP, CD, VG-C and CP contributed to the conception of the RecA\*-DinB study, A-EM and CP produced the data and analyses on this complex; A-EM, YW, MM, PD, CR and MB contributed to the conception of the RyR study; A-EM and CR produced the data; A-EM, MB and CR performed the analyses on RyR; PD and MB co-supervised A-EM's work, with contributions by MM and YW. A-EM, MB, CHR and CP wrote sections of the manuscript; YW, PD, MM, CD, MP and VG-C participated in the writing; CR, MB and CP supervised the writing; all authors revised the manuscript and approved the submitted version.

## REFERENCES

- Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., et al. (2015). GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* 1-2, 19–25. doi:10.1016/j.softx.2015.06.001
- Afonine, P. V., Grosse-Kunstleve, R. W., Adams, P. D., and Urzhumtsev, A. (2013). Bulk-solvent and Overall Scaling Revisited: Faster Calculations, Improved Results. *Acta Crystallogr. D* 69, 625–634. doi:10.1107/S0907444913000462
- Baldwin, P. R., Tan, Y. Z., Eng, E. T., Rice, W. J., Noble, A. J., Negro, C. J., et al. (2018). Big Data in cryoEM: Automated Collection, Processing and Accessibility of EM Data. *Curr. Opin. Microbiol.* 43, 1–8. doi:10.1016/j.mib.2017.10.005
- Batista, P. R., Pandey, G., Bisch, P. M., Pascutti, P. G., Perahia, D., and Robert, C. H. (2011). Free Energy Profiles along Consensus normal Modes Provide Insight into HIV-1 Protease Flap Opening. *J. Chem. Theor. Comput.* 7, 2348–2352. doi:10.1021/ct200237u
- Batista, P. R., Robert, C. H., Maréchal, J.-D., Ben Hamida-Rebaï, M., Pascutti, P., Bisch, P. M., et al. (2010). Consensus Modes, a Robust Description of Protein Collective Motions from Multiple-Minima normal Mode Analysis—Application to the HIV-1 Protease. *Phys. Chem. Chem. Phys.* 12, 2850–2859. doi:10.1039/b919148h
- Bell, J. C., and Kowalczykowski, S. C. (2016). RecA: Regulation and Mechanism of a Molecular Search Engine. *Trends Biochem. Sci.* 41, 491–507. doi:10.1016/j.tibs.2016.04.002
- Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M., Bansal, P., Bridge, A. J., et al. (2016). UniProtKB/Swiss-Prot, the Manually Annotated Section of the UniProt KnowledgeBase: How to Use the Entry View. *Methods Mol. Biol.* 1374, 23–54. doi:10.1007/978-1-4939-3167-5\_2

## FUNDING

We acknowledge the Servier group for funding part of this project. HPC resources from GENCI were obtained via the grant numbers A0070701714, A0080707438 and A0100707438. This work was supported by the French “Initiative d’Excellence” program (Grants “DYNAMO”, ANR-11-LABX-0011, and “CACSIce”, ANR-11-EQPX-0008), the Chu Family Foundation to MP, and VG-C research was supported in part by a grant to the Northeastern University Skills for Capacity and Inclusion Program by the Howard Hughes Medical Institute through the Science Education Program. MB thanks “Sesame Ile-de-France” for co-funding the display wall used for data analysis.

## ACKNOWLEDGMENTS

We are grateful for the cryo-EM data provided to us by Andrew R. Marks. Moreover, we also thank Sergei Grudinin for useful help with and update on the NOLB software. Computational work was performed using HPC resources from GENCI and LBT-HPC (thanks to the support from Geoffrey Letessier).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2022.826136/full#supplementary-material>

- Boyer, B., Danilowicz, C., Prentiss, M., and Prévost, C. (2019). Weaving DNA Strands: Structural Insight on ATP Hydrolysis in RecA-Induced Homologous Recombination. *Nucleic Acids Res.* 47, 7798–7808. doi:10.1093/nar/gkz667
- Conover, A. J., Danilowicz, C., Gunaratne, R., Coljee, V. W., Kleckner, N., and Prentiss, M. (2011). Changes in the Tension in dsDNA Alter the Conformation of RecA Bound to dsDNA–RecA Filaments. *Nucleic Acids Res.* 39, 8833–8843. doi:10.1093/nar/gkr561
- Croll, T. I. (2018). ISOLDE: a Physically Realistic Environment for Model Building into Low-Resolution Electron-Density Maps. *Crystallogr. D* 74, 519–530. doi:10.1107/S2059798318002425
- Crooks, G. E., Hon, G., Chandonia, J.-M., and Brenner, S. E. (2004). WebLogo: a Sequence Logo Generator. *Genome Res.* 14, 1188–1190. doi:10.1101/gr.849004
- Danilowicz, C., Feinstein, E., Conover, A., Coljee, V. W., Vlassakis, J., Chan, Y.-L., et al. (2012). RecA Homology Search Is Promoted by Mechanical Stress along the Scanned Duplex DNA. *Nucleic Acids Res.* 40, 1717–1727. doi:10.1093/nar/gkr855
- Danilowicz, C., Peacock-Villada, A., Vlassakis, J., Facon, A., Feinstein, E., Kleckner, N., et al. (2014). The Differential Extension in dsDNA Bound to Rad51 Filaments May Play Important Roles in Homology Recognition and Strand Exchange. *Nucleic Acids Res.* 42, 526–533. doi:10.1093/nar/gkt867
- Danilowicz, C., Vietorisz, E., Godoy-Carter, V., Prévost, C., and Prentiss, M. (2021). Influences of ssDNA-RecA Filament Length on the Fidelity of Homologous Recombination. *J. Mol. Biol.* 433, 167143. doi:10.1016/j.jmb.2021.167143
- de Jong, D. H., Singh, G., Bennett, W. F. D., Arnarez, C., Wassenaar, T. A., Schäfer, L. V., et al. (2013). Improved Parameters for the Martini Coarse-Grained Protein Force Field. *J. Chem. Theor. Comput.* 9, 687–697. doi:10.1021/ct300646g
- Dhindwal, S., Lobo, J., Cabra, V., Santiago, D. J., Nayak, A. R., Dryden, K., et al. (2017). A Cryo-EM-Based Model of Phosphorylation- and FKBP12.6-mediated

- Allosterism of the Cardiac Ryanodine Receptor. *Sci. Signal.* 10, eaai8842. doi:10.1126/scisignal.aai8842
- Dobson, L., Reményi, I., and Tusnády, G. E. (2015). CCTOP: a Consensus Constrained TOPology Prediction Web Server. *Nucleic Acids Res.* 43, W408–W412. doi:10.1093/nar/gkv451
- Du, G. G., Sandhu, B., Khanna, V. K., Guo, X. H., and MacLennan, D. H. (2002). Topology of the Ca<sup>2+</sup> Release Channel of Skeletal Muscle Sarcoplasmic Reticulum (RyR1). *Proc. Natl. Acad. Sci. U S A.* 99, 16725–16730. doi:10.1073/pnas.012688999
- Durand, P., Trinquier, G., and Sanejouand, Y.-H. (1994). A New Approach for Determining Low-Frequency normal Modes in Macromolecules. *Biopolymers* 34, 759–771. doi:10.1002/bip.360340608
- Efremov, R. G., Leitner, A., Aebersold, R., and Raunser, S. (2015). Architecture and Conformational Switch Mechanism of the Ryanodine Receptor. *Nature* 517, 39–43. doi:10.1038/nature13916
- Faini, M., Stengel, F., and Aebersold, R. (2016). The Evolving Contribution of Mass Spectrometry to Integrative Structural Biology. *J. Am. Soc. Mass. Spectrom.* 27, 966–974. doi:10.1007/s13361-016-1382-4
- Fill, M., and Copello, J. A. (2002). Ryanodine Receptor Calcium Release Channels. *Physiol. Rev.* 82, 893–922. doi:10.1152/physrev.00013.2002
- Georges, A. D., Clarke, O. B., Zalk, R., Yuan, Q., Condon, K. J., Grassucci, R. A., et al. (2016). Structural Basis for Gating and Activation of RyR1. *Cell* 167, 145–157. e17. doi:10.1016/j.cell.2016.08.075
- Ghanim, G. E., Fountain, A. J., van Roon, A.-M. M., Rangan, R., Das, R., Collins, K., et al. (2021). Structure of Human Telomerase Holoenzyme with Bound Telomeric DNA. *Nature* 593, 449–453. doi:10.1038/s41586-021-03415-4
- Godoy, V. G., Jarosz, D. F., Simon, S. M., Abyzov, A., Ilyin, V., and Walker, G. C. (2007). UmuD and RecA Directly Modulate the Mutagenic Potential of the Y Family DNA Polymerase DinB. *Mol. Cell* 28, 1058–1070. doi:10.1016/j.molcel.2007.10.025
- Heinz, L. P., Kopec, W., de Groot, B. L., and Fink, R. H. A. (2018). In Silico assessment of the Conduction Mechanism of the Ryanodine Receptor 1 Reveals Previously Unknown Exit Pathways. *Sci. Rep.* 8, 6886. doi:10.1038/s41598-018-25061-z
- Henrikus, S. S., Henry, C., McGrath, A. E., Jergic, S., McDonald, J. P., Hellmich, Y., et al. (2020). Single-molecule Live-Cell Imaging Reveals RecB-dependent Function of DNA Polymerase IV in Double Strand Break Repair. *Nucleic Acids Res.* 48, 8490–8508. doi:10.1093/nar/gkaa597
- Hoffmann, A., and Grudin, S. (2017). NOLB: Nonlinear Rigid Block normal-mode Analysis Method. *J. Chem. Theor. Comput* 13, 2123–2134. doi:10.1021/acs.jctc.7b00197
- Hristova, K., and Wimley, W. C. (2011). A Look at Arginine in Membranes. *J. Membr. Biol.* 239, 49–56. doi:10.1007/s00232-010-9323-9
- Hu, X., Dong, Q., Yang, J., and Zhang, Y. (2016). Recognizing Metal and Acid Radical Ion-Binding Sites by Integrating Ab Initio Modeling with Template-Based Transferals. *Bioinformatics* 32, 3260–3269. doi:10.1093/bioinformatics/btw396
- Huang, P.-S., Ban, Y.-E. A., Richter, F., Andre, I., Vernon, R., Schief, W. R., et al. (2011). RosettaRemodel: a Generalized Framework for Flexible Backbone Protein Design. *PLoS One* 6, e24109. doi:10.1371/journal.pone.0024109
- Inbar, Y., Benyamini, H., Nussinov, R., and Wolfson, H. J. (2005). Combinatorial Docking Approach for Structure Prediction of Large Proteins and Multi-Molecular Assemblies. *Phys. Biol.* 2, S156–S165. doi:10.1088/1478-3975/2/4/S10
- Jones, D. T. (1999). Protein Secondary Structure Prediction Based on Position-specific Scoring Matrices. *J. Mol. Biol.* 292, 195–202. doi:10.1006/jmbi.1999.3091
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* 596, 583–589. doi:10.1038/s41586-021-03819-2
- Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H., and Phillips, D. C. (1958). A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-ray Analysis. *Nature* 181, 662–666. doi:10.1038/181662a0
- Krieger, E., Koraimann, G., and Vriend, G. (2002). Increasing the Precision of Comparative Models with YASARA NOVA—a Self-Parameterizing Force Field. *Proteins* 47, 393–402. doi:10.1002/prot.10104
- Lamiable, A., Thévenet, P., Rey, J., Vavrusa, M., Derreumaux, P., and Tufféry, P. (2016). PEP-FOLD3: Faster De Novo Structure Prediction for Linear Peptides in Solution and in Complex. *Nucleic Acids Res.* 44, W449–W454. doi:10.1093/nar/gkw329
- Lanner, J. T., Georgiou, D. K., Joshi, A. D., and Hamilton, S. L. (2010). Ryanodine Receptors: Structure, Expression, Molecular Details, and Function in Calcium Release. *Cold Spring Harb Perspect. Biol.* 2, a003996. doi:10.1101/cshperspect.a003996
- Landreac, A., Férey, N., and Baaden, M. (2021). Welding the Power of Interactive Simulations. *WIREs Comput. Mol. Sci.*, e1594. doi:10.1002/wcms.1594
- Lee, S., Tran, A., Allsopp, M., Lim, J. B., Hénin, J., and Klauda, J. B. (2014). CHARMM36 United Atom Chain Model for Lipids and Surfactants. *J. Phys. Chem. B* 118, 547–556. doi:10.1021/jp410344g
- Lu, D., Danilowicz, C., Tashjian, T. F., Prévost, C., Godoy, V. G., and Prentiss, M. (2019). Slow Extension of the Invading DNA Strand in a D-Loop Formed by RecA-Mediated Homologous Recombination May Enhance Recognition of DNA Homology. *J. Biol. Chem.* 294, 8606–8616. doi:10.1074/jbc.RA119.007554
- Mackerell, A. D., Jr, Feig, M., and Brooks, C. L., 3rd (2004). Extending the Treatment of Backbone Energetics in Protein Force fields: Limitations of Gas-phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulations. *J. Comput. Chem.* 25, 1400–1415. doi:10.1002/jcc.20065
- Mirabello, C., and Wallner, B. (2017). InterPred: A Pipeline to Identify and Model Protein–Protein Interactions. *Proteins* 85, 1159–1170. doi:10.1002/prot.25280
- Moal, I. H., and Bates, P. A. (2010). SwarmDock and the Use of normal Modes in Protein-Protein Docking. *Int. J. Mol. Sci.* 11, 3623–3648. doi:10.3390/ijms11103623
- Molza, A.-E., Férey, N., Czjzek, M., Le Rumeur, E., Hubert, J.-F., Tek, A., et al. (2014). Innovative Interactive Flexible Docking Method for Multi-Scale Reconstruction Elucidates Dystrophin Molecular Assembly. *Faraday Discuss.* 169, 45–62. doi:10.1039/c3fd00134b
- Murata, K., and Wolf, M. (2018). Cryo-electron Microscopy for Structural Analysis of Dynamic Biological Macromolecules. *BBA-Gen Subjects* 1862, 324–334. doi:10.1016/j.bbagen.2017.07.020
- Nakai, J., Imagawa, T., Hakamat, Y., Shigekawa, M., Takeshima, H., and Numa, S. (1990). Primary Structure and Functional Expression from cDNA of the Cardiac Ryanodine Receptor/calcium Release Channel. *FEBS Lett.* 271, 169–177. doi:10.1016/0014-5793(90)80399-4
- Ngo, V. A., Perissinotti, L. L., Miranda, W., Chen, S. R. W., and Noskov, S. Y. (2017). Mapping Ryanodine Binding Sites in the Pore Cavity of Ryanodine Receptors. *Biophys. J.* 112, 1645–1653. doi:10.1016/j.bpj.2017.03.014
- Nogales, E. (2016). The Development of Cryo-EM into a Mainstream Structural Biology Technique. *Nat. Methods* 13, 24–27. doi:10.1038/nmeth.3694
- Peng, W., Shen, H., Wu, J., Guo, W., Pan, X., Wang, R., et al. (2016). Structural Basis for the Gating Mechanism of the Type 2 Ryanodine Receptor RyR2. *Science* 354, aah5324. doi:10.1126/science.aah5324
- Periole, X., Cavalli, M., Marrink, S.-J., and Ceruso, M. A. (2009). Combining an Elastic Network with a Coarse-Grained Molecular Force Field: Structure, Dynamics, and Intermolecular Recognition. *J. Chem. Theor. Comput* 5, 2531–2543. doi:10.1021/ct9002114
- Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., et al. (2005). Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* 26, 1781–1802. doi:10.1002/jcc.20289
- Prentiss, M., Prévost, C., and Danilowicz, C. (2015). Structure/function Relationships in RecA Protein-Mediated Homology Recognition and Strand Exchange. *Crit. Rev. Biochem. Mol. Biol.* 50, 453–476. doi:10.3109/10409238.2015.1092943
- Prévost, C., and Sacquin-Mora, S. (2021). Moving Pictures: Reassessing Docking Experiments with a Dynamic View of Protein Interfaces. *Proteins* 89, 1315–1323. doi:10.1002/prot.26152
- Robert, C., and Molza, A.-E. (2021). Low-frequency normal Modes Contribution to the Activated to Intermediate State Transition in the Ryanodine Receptor RyR1. Available at: <https://figshare.com/s/c0ec203c4662f1a39c1f>.
- Rueda, M., Chacón, P., and Orozco, M. (2007). Thorough Validation of Protein normal Mode Analysis: a Comparative Study with Essential Dynamics. *Structure* 15, 565–575. doi:10.1016/j.str.2007.03.013
- Russel, D., Lasker, K., Webb, B., Velázquez-Muriel, J., Tjioe, E., Schneidman-Duhovny, D., et al. (2012). Putting the Pieces Together: Integrative Modeling Platform Software for Structure Determination of Macromolecular Assemblies. *PLoS Biol.* 10, 1–5. doi:10.1371/journal.pbio.1001244

- Santulli, G., Lewis, D., des Georges, A., Marks, A. R., and Frank, J. (2018). "Ryanodine Receptor Structure and Function in Health and Disease," in *Membrane Protein Complexes: Structure and Function*. Editors J. R. Harris and E. J. Boekema (Singapore: Springer Singapore), 87, 329–352. Series Title: Subcellular Biochemistry. doi:10.1007/978-981-10-7757-9\_11
- Schneidman-Duhovny, D., and Wolfson, H. J. (2020). Modeling of Multimolecular Complexes. *Methods Mol. Biol.* 2112, 163–174. doi:10.1007/978-1-0716-0270-6\_12
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, Scalable Generation of High-Quality Protein Multiple Sequence Alignments Using Clustal Omega. *Mol. Syst. Biol.* 7, 539. doi:10.1038/msb.2011.75
- Spurgeon, S. R., Ophus, C., Jones, L., Petford-Long, A., Kalinin, S. V., Olszta, M. J., et al. (2021). Towards Data-Driven Next-Generation Transmission Electron Microscopy. *Nat. Mater.* 20, 274–279. doi:10.1038/s41563-020-00833-z
- Takeshima, H., Nishimura, S., Matsumoto, T., Ishida, H., Kangawa, K., Minamino, N., et al. (1989). Primary Structure and Expression from Complementary DNA of Skeletal Muscle Ryanodine Receptor. *Nature* 339, 439–445. doi:10.1038/339439a0
- Tama, F., and Sanejouand, Y. (2001). Conformational Change of Proteins Arising from normal Mode Calculations. *Protein Eng.* 14, 1–6. doi:10.1093/protein/14.1.1
- Tashjian, T. F., Danilowicz, C., Molza, A.-E., Nguyen, B. H., Prévost, C., Prentiss, M., et al. (2019). Residues in the Fingers Domain of the Translesion DNA Polymerase DinB Enable its Unique Participation in Error-Prone Double-Strand Break Repair. *J. Biol. Chem.* 294, 7588–7600. doi:10.1074/jbc.RA118.006233
- Thomas, A., Field, M. J., and Perahia, D. (1996). Analysis of the Low-Frequency normal Modes of the R State of Aspartate Transcarbamylase and a Comparison with the T State Modes. *J. Mol. Biol.* 261, 490–506. doi:10.1006/jmbi.1996.0478
- Trabuco, L. G., Villa, E., Mitra, K., Frank, J., and Schulten, K. (2008). Flexible Fitting of Atomic Structures into Electron Microscopy Maps Using Molecular Dynamics. *Structure* 16, 673–683. doi:10.1016/j.str.2008.03.005
- Tsirigos, K. D., Peters, C., Shu, N., Käll, L., and Elofsson, A. (2015). The TOPCONS Web Server for Consensus Prediction of Membrane Protein Topology and Signal Peptides. *Nucleic Acids Res.* 43, W401–W407. doi:10.1093/nar/gkv485
- Ulmschneider, M. B., Ulmschneider, J. P., Freites, J. A., von Heijne, G., Tobias, D. J., and White, S. H. (2017). Transmembrane Helices Containing a Charged Arginine Are Thermodynamically Stable. *Eur. Biophys. J.* 46, 627–637. doi:10.1007/s00249-017-1206-x
- Urzhumtsev, A., Afonine, P. V., Van Benschoten, A. H., Fraser, J. S., and Adams, P. D. (2016). From Deep TLS Validation to Ensembles of Atomic Models Built from Elemental Motions. Addenda and Corrigendum. *Acta Crystallogr. D* 72, 1073–1075. doi:10.1107/S2059798316013048
- Van Petegem, F. (2017). Ligand Binding to Ryanodine Receptors Revealed through Cryo-Electron Microscopy. *Cell Calcium* 61, 50–52. doi:10.1016/j.ceca.2016.10.004
- Van Petegem, F. (2015). Ryanodine Receptors: Allosteric Ion Channel Giants. *J. Mol. Biol.* 427, 31–53. doi:10.1016/j.jmb.2014.08.004
- van Zundert, G., Rodrigues, J., Trellet, M., Schmitz, C., Kastrius, P., Karaca, E., et al. (2016). The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* 428, 720–725. doi:10.1016/j.jmb.2015.09.014
- Verkhivker, G. M., Agajanian, S., Oztas, D. Y., and Gupta, G. (2021). Comparative Perturbation-Based Modeling of the SARS-CoV-2 Spike Protein Binding with Host Receptor and Neutralizing Antibodies: Structurally Adaptable Allosteric Communication Hotspots Define Spike Sites Targeted by Global Circulating Mutations. *Biochemistry* 60 (19), 1459–1484. doi:10.1021/acs.biochem.1c00139
- Ward, J. J., McGuffin, L. J., Bryson, K., Buxton, B. F., and Jones, D. T. (2004). The DISOPRED Server for the Prediction of Protein Disorder. *Bioinformatics* 20, 2138–2139. doi:10.1093/bioinformatics/bth195
- Watson, J. D., and Crick, F. H. (1953). Genetical Implications of the Structure of Deoxyribonucleic Acid. *Nature* 171, 964–967. doi:10.1038/171964b0
- Webb, B., and Sali, A. (2021). Protein Structure Modeling with Modeller. *Methods Mol. Biol.* 2199, 239–255. doi:10.1007/978-1-0716-0892-0\_14
- Wheeler, T. J., Clements, J., and Finn, R. D. (2014). Skylin: a Tool for Creating Informative, Interactive Logos Representing Sequence Alignments and Profile Hidden Markov Models. *BMC Bioinformatics* 15, 7. doi:10.1186/1471-2105-15-7
- Willegens, K., and Efremov, R. G. (2018). Influence of Lipid Mimetics on Gating of Ryanodine Receptor. *Structure* 26, 1303–1313. e4. doi:10.1016/j.str.2018.06.010
- Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N., et al. (2018). MolProbity: More and Better Reference Data for Improved All-Atom Structure Validation. *Protein Sci.* 27, 293–315. doi:10.1002/pro.3330
- Yan, R., Xu, D., Yang, J., Walker, S., and Zhang, Y. (2013). A Comparative Assessment and Analysis of 20 Representative Sequence Alignment Methods for Protein Structure Prediction. *Sci. Rep.* 3, 2619. doi:10.1038/srep02619
- Yan, Z., Bai, X.-c., Yan, C., Wu, J., Li, Z., Xie, T., et al. (2015). Structure of the Rabbit Ryanodine Receptor RyR1 at Near-Atomic Resolution. *Nature* 517, 50–55. doi:10.1038/nature14063
- Yang, D., Boyer, B., Prévost, C., Danilowicz, C., and Prentiss, M. (2015). Integrating Multi-Scale Data on Homologous Recombination into a New Recognition Mechanism Based on Simulations of the RecA-ssDNA/dsDNA Structure. *Nucleic Acids Res.* 43, 10251–10263. doi:10.1093/nar/gkv883
- Yang, H., Zhou, C., Dhar, A., and Pavletich, N. P. (2020). Mechanism of Strand Exchange from RecA-DNA Synaptic and D-Loop Structures. *Nature* 586, 801–806. doi:10.1038/s41586-020-2820-9
- Zalk, R., Clarke, O. B., des Georges, A., Grassucci, R. A., Reiken, S., Mancina, F., et al. (2015). Structure of a Mammalian Ryanodine Receptor. *Nature* 517, 44–49. doi:10.1038/nature13950
- Ziegler, S. J., Mallinson, S. J., St. John, P. C., and Bomble, Y. J. (2021). Advances in Integrative Structural Biology: Towards Understanding Protein Complexes in Their Cellular Context. *Comput. Struct. Biotech. J.* 19, 214–225. doi:10.1016/j.csbj.2020.11.052

**Conflict of Interest:** MM was employed by Servier Monde.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Molza, Westermaier, Moutte, Ducrot, Danilowicz, Godoy-Carter, Prentiss, Robert, Baaden and Prévost. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.