Molecular dynamics of the human RhD and RhAG blood group proteins

Aline Floch[1,2]*, Tatiana Galochkina[3], France Pirenne[1,2], Christophe Tournamille[1,2] and Alexandre G. de Brevern[3]*

[1]University Paris Est Créteil, INSERM U955 Equipe Transfusion et Maladies du Globule Rouge, IMRB, Créteil, France, [2]Laboratoire de Biologie Médicale de Référence en Immuno-Hématologie Moléculaire, Etablissement Français du Sang Ile-de-France, Créteil, France, [3]Université Paris Cité and Université des Antilles and Université de la Réunion, Biologie Intégrée du Globule Rouge, UMR_S1134, BIGR, INSERM, DSIMB Bioinformatics team, Paris, France

Introduction: Blood group antigens of the RH system (formerly known as "Rhesus") play an important role in transfusion medicine because of the severe haemolytic consequences of antibodies to these antigens. No crystal structure is available for RhD proteins with its partner RhAG, and the precise stoichiometry of the trimer complex remains unknown.

Methods: To analyse their structural properties, the trimers formed by RhD and/or RhAG subunits were generated by protein modelling and molecular dynamics simulations were performed.

Results: No major differences in structural behaviour were found between trimers of different compositions. The conformation of the subunits is relatively constant during molecular dynamics simulations, except for three large disordered loops.

Discussion: This work makes it possible to propose a reasonable stoichiometry and demonstrates the potential of studying the structural behaviour of these proteins to investigate the hundreds of genetic variants relevant to transfusion medicine.

KEYWORDS

Rh blood group system, membrane proteins, molecular models, molecular dynamics simulation, protein structural elements, structural alphabet

1 Introduction

In transfusion medicine, blood group antigens from the RH system (formerly known as « Rhesus ») play a crucial role (Chou and Westhoff, 2010; Thornton and Grimsley, 2019). Firstly, because of their immunogenicity (ability to induce antibody formation in recipients) and secondly, because of the severe haemolytic consequences antibodies to these antigens can induce: post-transfusion haemolysis, haemolytic disease of the foetus and the newborn, transfusion impasse (Noizat-Pirenne, 2012; de Haas et al., 2015). Antibody formation may occur when an individual is exposed to foreign antigens through transfusion or pregnancy, but also depends on mostly unknown individual factors (Tormey and Hendrickson, 2019).

RH antigens are expressed on the red blood cell (RBC) membrane by two Rh proteins, i.e., RhD and RhCE. The proteins result from the expression of the highly homologous genes *RHD* and *RHCE*, situated on chromosome 1, at locus 1p36.11. They are associated with 5 main antigens: D, C, E, c and e (also named RH1, RH2, RH3, RH4 and RH5,

respectively). The *RHD* gene, when present, produces the RhD protein, which expresses the D antigen. The *RHCE* gene produces the RhCE protein with 4 common alleles resulting in the expression of the C, c, E and e antigens (each allele combines the expression of 2 antigens: C or c associated with E or e). Fifty other antigens have been defined in this blood group system, resulting from altered Rh proteins arising from point mutations, from hybrid alleles (through recombination events between *RHD* and *RHCE*), or from both (Chou and Westhoff, 2010).

RH antigens and phenotypes resulting from novel alleles can be characterized using experimental (serological data: testing with polyclonal antibodies resulting from alloimmunisation, monoclonal antibodies specific of certain epitopes, by flow cytometry …) and clinical data (if at least one carrier has produced antibodies to the standard antigen or if no antibodies have been detected despite many carriers having been exposed) (Chou and Westhoff, 2010). These methods provide only indirect information about the protein conformation and behaviour, and assumptions based on them are unreliable. Indeed, while several alleles were initially thought to present no alloimmunisation risk, individuals homozygous for these alleles have produced alloantibodies to the standard antigen, revealing that the antigen was, in fact, altered (St-Louis et al., 2011; Pham et al., 2013). Moreover, experimental data is unavailable for many alleles, especially the rare ones. In those cases, the assessment depends solely on whether the amino acid substitutions are exposed at the extracellular RBC membrane. Such a position is considered to surely alter the antigen's epitopes. We have proposed the most complete database dedicated to Rh variants. This dedicated database is named RHeference and contains entries for 710 *RHD* alleles, 11 *RHCE* alleles, 30 phenotype descriptions, it also covers also partly characterized alleles, haplotypes, and some miscellaneous entries, with molecular, phenotypic, serological, alloimmunization, haplotype, geographical, and other data, detailed for each source (Floch et al., 2021a).

Molecular modelling techniques have the potential to provide new and more direct information on Rh protein structure and dynamics (Burton and Daniels, 2011; de Brevern et al., 2018; Floch et al., 2021b), which is crucially important to better understand the expression of RH antigens and antibody formation in transfusion medicine. The first two major questions that could be addressed using molecular models are the three-dimensional structure of RhD monomers, and the structure and stability of RhD/RhAG trimers of different compositions. The monomer structure can be modelled by homology, i.e., mapping protein sequence to the structure of a resolved protein homolog used as a template. The composition and the stability of the trimers are more complex questions that can be addressed by running molecular dynamics (MD) simulations to reproduce protein movements at the atomistic level.

The main challenge in the development of reliable models of Rh/RhAG trimers for molecular dynamics is the choice of the structural template and identification of the trimer arrangement to be used as the initial conformation for the simulations. The Rh proteins RhD and RhCE and their main partner, RhAG ("Rhesus Associated Glycoprotein") (Chou and Westhoff, 2010), are highly homologous proteins, with 12 membrane-spanning domains. They are part of the Amt/MEP/Rh superfamily of ammonium
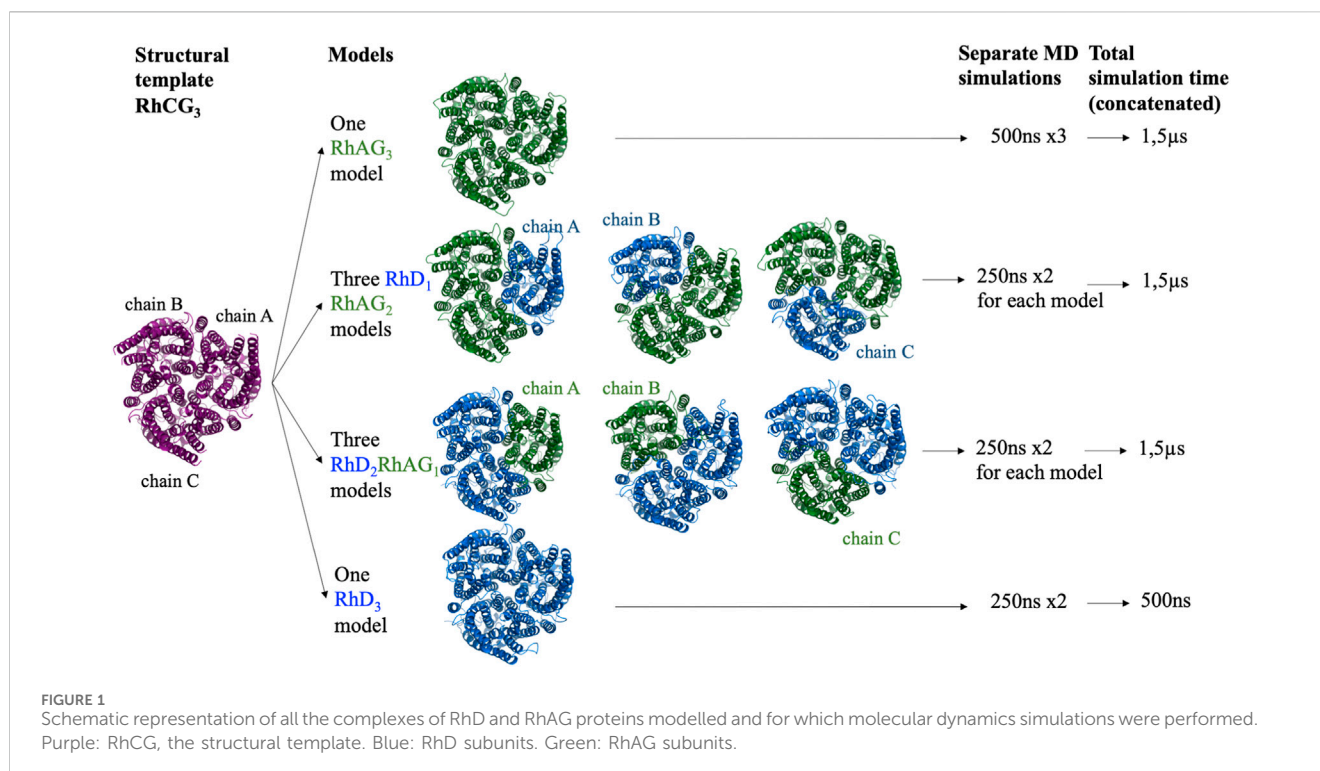
transporters, with a trimeric structure (Khademi et al., 2004; Li et al., 2007; Lupo et al., 2007; Gruswitz et al., 2010). No crystal structure has been published for the human Rh blood group proteins at the time when this project was initiated, but the structure of the human RhCG protein (PDB ID: 3HD6) (Gruswitz et al., 2010) and of the bacterial homolog NeRh50 from *Nitrosomonas europaea* (PDB ID: 3B9W) (Li et al., 2007; Lupo et al., 2007) have been resolved. RhCG shares 31.65% sequence identity with RhD, and 52.32% with RhAG, while NeRh50 has 24.86% sequence identity with RhD, and 35.75% with RhAG. Experimental data suggests that RhD and RhCE monomers are not associated in the same trimer (Avent et al., 1988), and that the composition of the human Rh trimer is likely to be 2 RhAG monomers for 1 RhD or RhCE monomer (Mouro-Chanteloup et al., 2002; Floch et al., 2021b). Homotrimers of RhAG can also be detected experimentally, when the *RHD* and *RHCE* genes are both altered. In contrast, when the *RHAG* gene is altered, no RhD or RhCE proteins are expressed (Chou and Westhoff, 2010), which reveals that RhD$_3$ and RhCE$_3$ homotrimers do not exist. The 37th residue of RhAG is glycosylated but neither RhD nor RhCE are glycosylated (Anstee and Tanner, 1993).

A few 3D models for the study of Rh proteins have been proposed previously by homology modelling. These propositions have been based on bacterial homologues, and/or on automated modelling methods (Khademi et al., 2004; Conroy et al., 2005; Callebaut et al., 2006; Lupo et al., 2007; Flegel et al., 2008; Tilley et al., 2010; Silvy et al., 2012). Our team has recently proposed, first a new multi-template model of the RhD monomer (de Brevern et al., 2018) based on the human RhCG and NeRh50 as templates, followed by a model of the RhD-RhAG trimer (Floch et al., 2021b), based on the human RhCG (the multi-template approach did not lead to better models), using state-of-the-art methods specific for transmembrane proteins. A recent electron microscopy structure was made available for RhCE in complex with RhAG and ankyrine [PDB id 7uzq (Vallese et al., 2022)] and used to propose monomer models (Trueba-Gómez et al., 2023). A few teams have modelled the trimer that Rh proteins form with RhAG, but none have performed molecular dynamics (MD) simulations for these proteins. This study focuses on the trimer(s) formed by RhD and RhAG proteins. We propose the best RhAG$_3$ and RhAG$_2$RhD models by homology modelling and study the molecular dynamics behaviour of the proteins through MD simulations. In recent years, deep learning approaches have had a major impact on the development of 3D protein models (Jumper et al., 2021), so we will also explore the potential benefits of AlphaFold2, for which some limitations have been also underlined (He et al., 2023; Tourlet et al., 2023).

2 Materials and methods

2.1 Comparative modelling of RhAG/RhD trimers

Since no structure for the human Rh proteins (RhD, UniProt id Q02161 and RhCE, UniProt id P18577) (Consortium, 2015) had been resolved at the time of our study, Rh trimers were modelled by homology (de Brevern, 2010) as previously published for the RhD

FIGURE 1
Schematic representation of all the complexes of RhD and RhAG proteins modelled and for which molecular dynamics simulations were performed. Purple: RhCG, the structural template. Blue: RhD subunits. Green: RhAG subunits.

monomer (de Brevern et al., 2018). Sequence's conservation was assessed with Consurf server (Ashkenazy et al., 2016), then homologous protein sequences for which protein structures are available were found by a query of the sequences of the target proteins (here, RhAG and RhD) with PSI-BLAST (Altschul et al., 1997) and HHpred (Hildebrand et al., 2009), using the Protein DataBank (PDB) database (Berman et al., 2002). The selected structures were analysed with MolProbity (Chen et al., 2010) and ProCheck (Laskowski et al., 1993). The sequences were carefully aligned with Clustal Omega (Sievers et al., 2011). Protein structures and/or structural models were superimposed using the McLachlan algorithm (McLachlan, 1982) as implemented in the program ProFit 3.1, and using iPBA (Gelly et al., 2011), and TM-Align (Zhang and Skolnick, 2005). Secondary structure was predicted by several specific methods, such as HMMTOP (Tusnády and Simon, 2001) and MEMSAT (Jones, 2007). Hundreds of structural models were generated with Modeller 9.12 (Sali and Blundell, 1993; Martí-Renom et al., 2000). These models were analysed and compared with TM-Align (Zhang and Skolnick, 2005), the Discrete Optimized Protein Energy (DOPE) (Shen and Sali, 2006) potential implemented in Modeller, the Hybrid Protein Model (HPM, dedicated to transmembrane proteins) (de Brevern and Hazout, 2003; Esque et al., 2015; Téletchéa et al., 2023) webserver and the MAIDEN approach (Postic et al., 2016a). The positioning of the models within the lipid bilayer were performed by OPM (Lomize et al., 2006; Lomize et al., 2011) and OREMPRO (Postic et al., 2016b). The best structural model was selected.

Using a symmetrical homotrimer of human RhCG (PDB ID: 3hd6), whose multimeric state was built with PDBePISA (Proteins, Interfaces, Structures and Assemblies) (Krissinel and Henrick, 2007), trimers of all the potential compositions were modelled:

one $RhAG_3$ homo-trimer, three RhD_1RhAG_2 trimers (with the RhD monomer positioned either as chain A, B or C), three RhD_2RhAG_1 (with the RhAG monomer positioned either as chain A, B or C) and one RhD_3 homo-trimer (Figure 1). RhCG (PDB id 3hd6) was used as a structural template. The recent structure of RhCE in complex with RhAG and ankyrine [PDB id 7uzq (Vallese et al., 2022)] and AlphaFold2 model (Jumper et al., 2021) will be discussed in Discussion section.

2.2 Molecular dynamics simulations protocol

The membrane systems were prepared using the CHARMM-GUI webserver (Wu et al., 2014). Each trimer was inserted into a rectangular box, with a palmitoyl-oleoyl phosphatidyl choline (POPC) bilayer, solvated with TIP3P water model (Jorgensen et al., 1983) and neutralized with Na^+ and Cl^- counter ions at a concentration of 0.15 M. The molecular dynamics (MD) simulations were performed with GROMACS 2016.4 (Van Der Spoel et al., 2005; Abraham et al., 2015) using all-atom CHARMM36 forcefield (Huang et al., 2013).

The system geometry was optimized by minimizing the energy with a steepest-descent algorithm for 5,000 steps, keeping the positions of the heavy atoms of the proteins fixed. As recommended by the CHARMM-GUI webserver output (Lee et al., 2000), the system was then equilibrated in six short runs from 25 ps to 100 ps each, using the Berendsen algorithm for temperature and pressure control (Berendsen et al., 1984), with time coupling constraints of 0.1 ps, and gradually releasing constraints keeping the positions of the heavy atoms of the proteins fixed. The

resulting equilibrated systems were used as an initial condition for the production runs of 250–500 ns with all the constraints removed.

The MD production runs were performed with the temperature fixed at 310,15 K (to approximate body temperature). Separate temperature coupling was applied for the protein, lipids and solvent molecules using the Nose-Hoover algorithm (Nosé, 1984; Hoover, 1985), with a coupling constant of 1.0 ps. The MD runs were performed with the pressure fixed at 1 bar, with semi-isotropic coupling using the Parrinello-Rahman barostat (Parrinello and Rahman, 1981) with a coupling constant of 5.0 ps. Periodic boundary conditions were used. Long-range electrostatic interactions were handled with the Particle Mesh Ewald (PME) algorithm (Darden et al., 1999). The lengths of the covalent bonds between hydrogen and heavy atoms were fixed using the LINCS algorithm (Hess et al., 1997), which enabled the use of an integration step of 2 fs.

Three independent 500 ns runs were performed for the RhAG$_3$ system, for a cumulative length of 1.5 μs? Two independent 250 ns runs were performed for each of the other systems (see Figure 1), for a cumulative length of 1.5μs for RhD$_1$RhAG$_2$ systems (regardless of the chain which was RhD), 1.5μs for RhD$_2$RhAG$_1$ systems (regardless of the chain which was RhAG), and only 500ns for the RhD$_3$ system, since the existence of the latter system *in vivo* is unlikely, as explained above.

2.3 MD simulation analysis

For the 8 trimer models considered, we obtained MD trajectories with conformations saved every 100 ps. The GROMACS software was used for the analyses of MD simulation trajectories, with in-house Python and R scripts. Root mean square deviations (RMSD) and root mean square fluctuations (RMSF) were computed on Cα atoms. We computed normalized RMSFs for each chain (Bornot et al., 2011).

To analyse local conformations, we used a structural alphabet composed of 16 local prototypes, called Protein Blocks (PBs) (de Brevern et al., 2000; Léonard et al., 2014), to describe the protein backbone conformation based on the dihedral angles of 5 consecutive residues. PBs are labelled from *a* to *p*. PBs *a* to *j* are specific to coils. PBs primarily representing β-strands are: PB *d* which can be roughly described as the prototype for central β-strand; PBs *a* to *c* which represent β-strand N-caps, and PBs *e* and *f* which represent β-strand C-caps. PBs primarily representing α-helices are: PB *m*, which can be roughly described as the prototype for α-helix, PBs *k* and *l* which are specific to α-helix N-caps and PBs *n* to *p* which are specific to α-helix C-caps. PB assignment was carried out for every residue obtained from every snapshot extracted from MD simulations using the PBxplore tool (Barnoud et al., 2017). The flexibility of each position was quantified with the N_{eq} (for equivalent number of PBs), a statistical measurement similar to entropy. It represents the average number of PBs a residue adopts at a given position (de Brevern et al., 2000).

$$N_{eq} = exp\left(-\sum_{x=1}^{16} f_x \ln f_x\right)$$

N_{eq} is calculated as shown here, where, f_x is the frequency of PB *x* at the position of interest. An N_{eq} value of 1 indicates that only one type

of PB is observed, while a value of 16 is equivalent to an equal probability for each of the 16 states, *i.e.*, a random distribution. An N_{eq} of 1.0 is, by definition, a rigid region of a protein. We have determined that disordered regions have an N_{eq} of 8.0 or more, while flexible regions have an N_{eq} around 4.0 (Akhila et al., 2020).

The analysis of conformational variations in terms of PBs and normalized RMSF was performed on the concatenated trajectories. Monomers coming from trimers of identical composition were analysed together (e.g., all trajectories for the RhD subunit within trimers composed of 2 RhAG and 1 RhD subunits).

To compare the local conformational variability between two different trajectories of the same protein, we calculated the total difference between PB frequencies (ΔPB).

$$\Delta PB = \sum_{x=1}^{16} |f_x{}^1 - f_x{}^2|$$

To compare the local conformational variability between two different trajectories of the same protein, ΔPB, the total difference between PB frequencies, was computed using this formula, where $f_x{}^1$ and $f_x{}^2$ are frequencies of PB *x*, in the first and the second trajectory respectively for the same position. The 50 ns at the beginning of each trajectory were truncated to compute average PB and RMSF only at the steady state (once the RMSD reached a plateau).

3 Results

3.1 Structure of RhAG/RhD model trimers

At the time of our study, out of the potential structural templates identified, human RhCG (PDB id 3hd6) (Gruswitz et al., 2010) was the best template, with the highest sequence identity to both with RhD (31.65%), and with RhAG (52.32%, see Supplementary Table S1). The most conserved part of the sequence corresponds to the transmembrane regions, which is consistent with the evolutionary conservation assessed by Consurf (Ashkenazy et al., 2016) (see Supplementary Figures S1, S2). The target sequence of each trimer formed by RhD and/ or RhAG subunits in all compositions was aligned to the sequences of both templates using Clustal Omega (Sievers et al., 2011).

Proposing trimer structure models is challenging. However, the difference in the models generated in our case seemed to be very small. For instance, for the RhAG$_3$ trimers composed of a total of 1,227 residues, the average RMSD of the models was 1.56 ± 0.16 Å and ranged from 0.97 to 2.12 Å for the whole trimer, with a minimum coverage of 1,202 residues, as measured by TM-Align (Zhang and Skolnick, 2005). The results were similar between the different models with the same composition (Figure 1). The RMSD for RhD$_1$RhAG$_2$ trimer models (1,235 residues) was 1.17 ± 0.23 Å and ranged from 0.61 to 1.93 Å, with a minimum coverage of 1,195 residues, and the RMSD for RhD$_2$RhAG$_1$ trimer models (1,243 residues) was 1.03 ± 0.24 Å and ranged from 0.54 to 1.73 Å, with a minimum coverage of 1,208 residues.
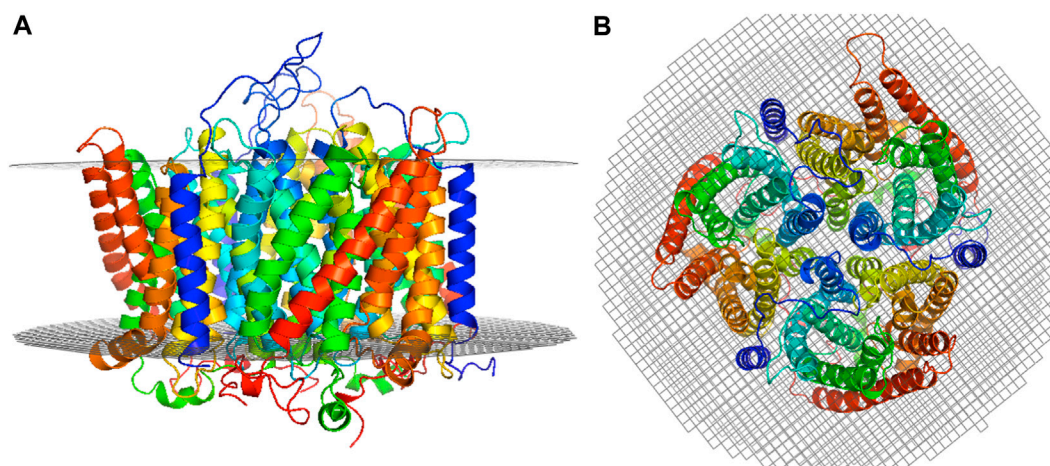
FIGURE 2
Selected RhD₁RhAG₂ trimer model. The RhD monomer is positioned as chain A (relative to the numbering of the structural template human RhCG, PDB ID: 3HD6) and the trimer positioned in a lipid bilayer using OREMPRO (Postic et al., 2016b). The extracellular compartment is at the top of figure. **(A)** Lateral view and **(B)** top view.

The best trimer for the RhD₁RhAG₂ with the RhD monomer positioned as chain A (relative to the numbering of the structural template human RhCG, PDB id: 3hd6, see Figure 1) is shown in Figure 2, positioned in a lipid bilayer using OREMPRO (Postic et al., 2016b).

3.2 Global flexibility of Rh trimers in molecular dynamics simulations

We have performed 1.5 μs of MD simulations for each of the trimer compositions: RhAG₃, RhD₁RhAG₂ and RhD₂RhAG₁ (see Figure 1). In addition, 500ns of MD simulations for the non-physiological RhD₃ system was done (see Figure 1). For all systems, the RMSD with respect to the initial conformation reached a plateau after 50 ns of simulation, indicating stabilisation of the protein structure (see Supplementary Figure S3).

The main behaviours of RhAG and RhD are summarized in Figures 3, 4. The analysis with Protein Blocks (PB) (de Brevern et al., 2000; Barnoud et al., 2017) reveals that most transmembrane domains of both RhD and RhAG (in Figures 3B, 4B) are α-helices, associated to PB *m* (see Figures 3A, 4A). They are highly rigid as their N_{eq} values are equal to 1.0 (Akhila et al., 2020) (see Figures 3C, 4C), i.e., only one PB is observed during the whole simulation. Nonetheless, we observe local variations in several helices, especially in the middle of helices 8 and 9 in RhAG (see Figures 3B, C) and in helix 9 in RhD (see Figures 4E, F). For these residues, N_{eq} reaches 2.0. This value remains relatively low because the conformation remains predominantly helical. In Figure 5, we show several examples of different conformations illustrating the PB transitions in these regions. Increased conformation variability is also observed at the extremities of several helices, with a higher N_{eq} (see Figures 3C, 4C, and the orange and yellow colours in Figures 3A, 4A). This applies in particular to the N-terminal ends of helices

2, 3, 6, 11 and 12 of both RhAG and RhD, often associated with the PB series *cklm,* with a reasonably low N_{eq} of 2.0. Similarly, variations are observed at the C-terminal extremities of helices 2, 3, 7, 8, 9 and 11 of RhAG and helices 2, 3, 6, 7, 9 and 11 of RhD, often associated to the PB series *mop* (see Figures 3A, 4A). These variations of helical extremities are mainly directed by connecting loops (de Brevern et al., 2002).

3.3 Conformational behaviour of the most mobile protein regions

In fact, loop analysis shows which loops are disordered, flexible, or relatively rigid (Akhila et al., 2020). In both proteins, 3 loops are disordered, with a N_{eq} of 8.0 or more: loops 2, 4 and 12 in RhAG (the first, second and last extracellular loops) and loops 2, 7 and 12 in RhD (respectively the first extracellular loop, the fourth intracellular loop and the last extracellular loop, respectively). Several loops are not disordered but can be considered flexible, with an N_{eq} between 4.0 and 8.0: loops 6 and 7 in RhAG (the third extracellular and fourth intracellular loops), and loops 3, 6, 8 and 10 (respectively the second intracellular, third extracellular, fourth extracellular and sixth intracellular loops, respectively). The more rigid loops, with an N_{eq} less than 4.0 were: loops 3, 5, 8, 9, 10 and 11 in RhAG (all very small, according to Stride's (Frishman and Argos, 1995) secondary structure assignment), and loops 4 and 8 in RhD (both small intracellular loops). Within the loops, several short alignments of PBs or structural words (de Brevern et al., 2002) show a β-sheet conformation: loops 2 and 4 in both RhAG (near residues 26 and 107) and in RhD (near residues 32 and 105). Most interestingly, the seventh loop of RhAG has a well-defined helical structure around residue 180. This loop is the fourth intracellular loop and the persistence of such local conformations may play a role in protein function.
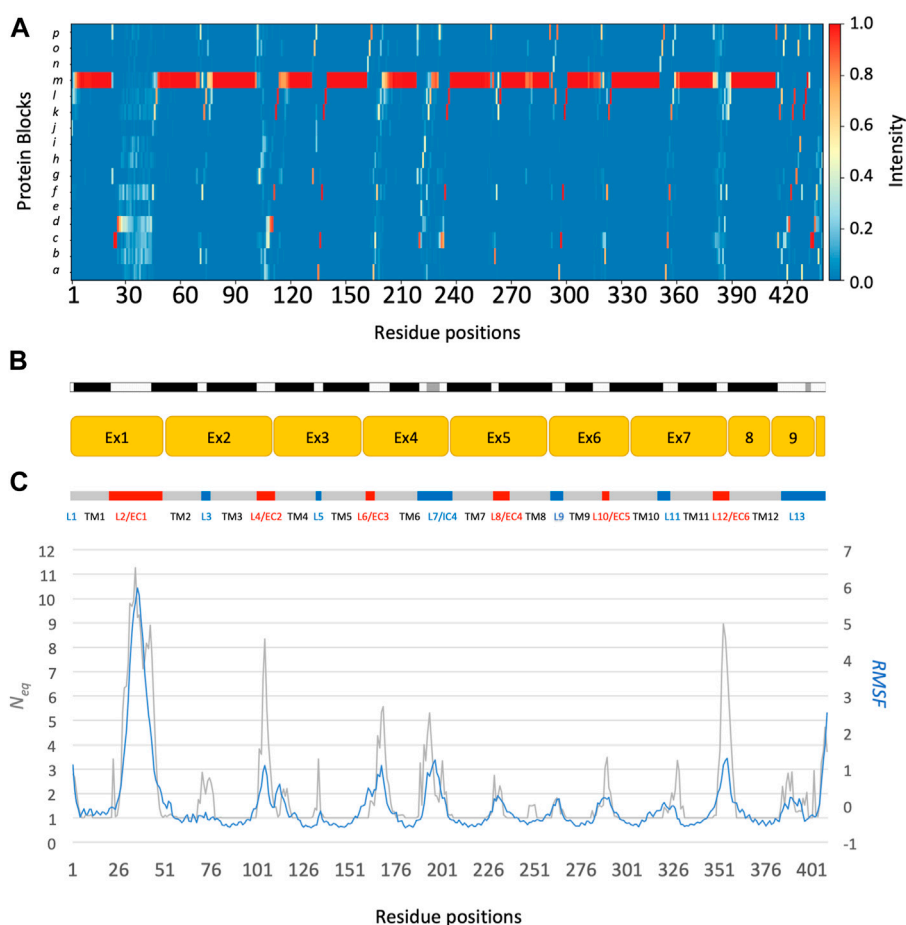
FIGURE 3
RhAG protein. The residues range from 1 to 409, from left to right. **(A)** Protein Blocks (de Brevern et al., 2000)assignment for the analysis of local conformations, extracted from each snapshot of the molecular dynamics simulations using the PBxplore tool (Barnoud et al., 2017). **(B)** Stride secondary structure assignment (Frishman and Argos, 1995), with α-helices in black, and 3_{10} helices in grey; Limits of the 10 exons of the protein; Position relative to the lipid bilayer, with intracellular domains (IC) in blue, transmembrane (TM) domains in grey, and extracellular domains in red (EC). **(C)** N_{eq}, for equivalent number of Protein Blocks, in grey (with the axis to the left) and average, normalized RMSF, in blue (with the axis to the right).

3.4 Comparison of the different flexibility/disorder descriptors for protein dynamics analysis

The correlation between normalized B-factor values and normalized RMSF in both RhD and RhAG was high (Pearson's coefficient $r = 0.85$ for RhAG loops, $r = 0.83$ for RhD loops). The obtained correlation is higher than previously observed in datasets of varied, representative globular proteins (Bornot et al., 2011; Narwani et al., 2018). However, some differences between RMSF and N_{eq} were observed. It should be recalled that the RMSF reflects a "global" structural variation, measuring the difference of all protein conformations with respect to a reference average conformation, whereas the N_{eq} is a local quantification, describing the protein backbone conformation based on the dihedral angles of 5 consecutive residues, i.e., PBs. It is therefore possible to have an N_{eq} of 1 (ultra-rigid) associated with a very high RMSF, i.e., a mobile region surrounded by deformable regions. The opposite (a low

RMSF with a high N_{eq}) is observed less frequently and represents strong local variations/oscillations of a protein region located in a globally more stable environment. RhAG loops 4 and 12 (the second and sixth extracellular loops) are the most prominent examples of this. In RhD, the four loops previously mentioned as flexible (loops 3, 6, 8 and 10) had a low RMSF compared to their high N_{eq}.

3.5 Influence of the trimer composition on protein dynamics

The differences between the systems depending on the trimer composition were also analysed. The normalized RMSF for either RhAG or RhD showed only slight differences as a function of the trimer composition (see Supplementary Figure S4). A slightly lower RMSF was observed in RhAG for trimers with 3 RhAG subunits around residues 193 to 201 (see Supplementary Figure S4A), which are part of the large fourth intracellular loop (seventh loop). A very
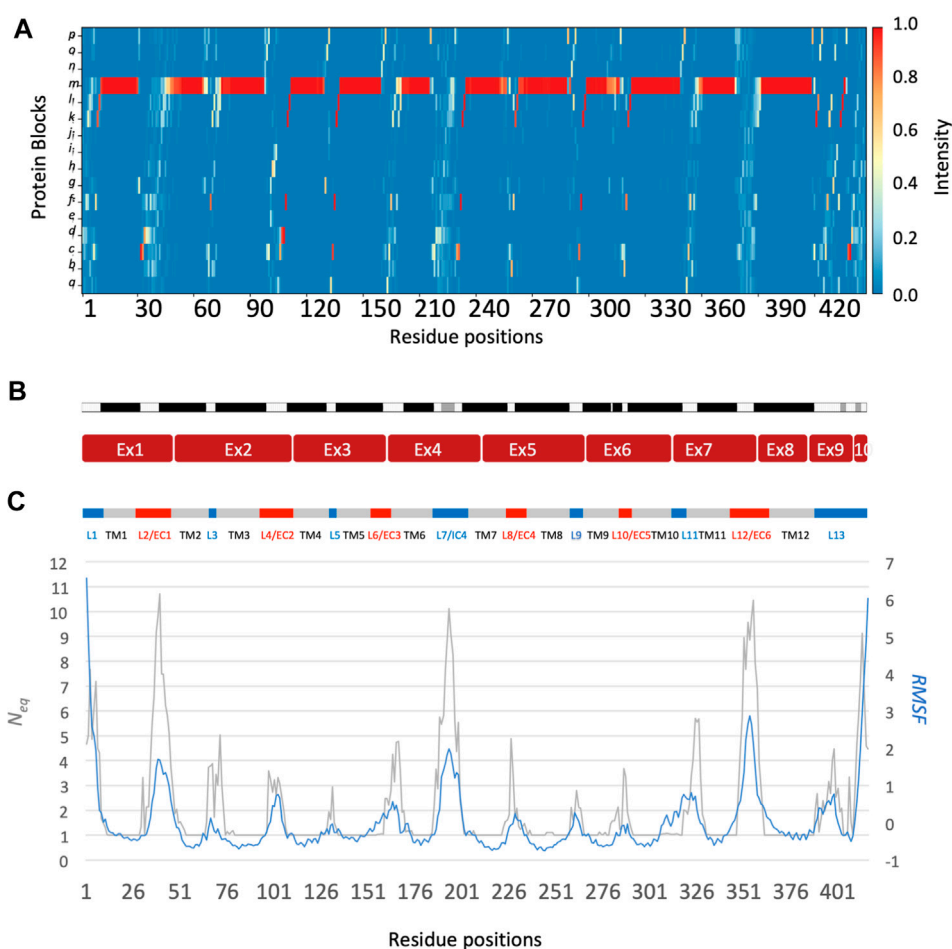
FIGURE 4
RhD protein. The residues range from 1 to 417, from left to right. See Figure 3 for legend.

small increase in RMSF was observed in RhD for trimers with 3 RhD subunits around residues 225 to 237, which is the fourth extracellular loop (eighth loop, see Supplementary Figure S4A). It is quite remarkable to see such comparability in a system of this size.

To further explore the variations in local conformations for RhAG and RhD, we calculated ΔPB to compare the PB frequencies observed for the systems with different RhAG and RhD composition. Figure 6A shows ΔPB between complexes with 1, 2 or 3 RhAG and Figure 6B shows similar information for systems with 1, 2 or 3 RhD. The comparison of the global distributions of PBs between the different systems does not reveal any particularities, i.e., no system had a region with very different conformational sampling. Outside of the disordered areas, there are no differences in ΔPB greater than 0.2. Strikingly, the trimer composition has no significant effect on the sampling of local protein conformations. This result may seem surprising, even counterintuitive. It should be clearly understood that the behaviour of a transmembrane protein complex constrained in a lipid environment is less labile than a globular protein. We thus have variations which are in fact quite classic, but do not show any strong change between the different systems.
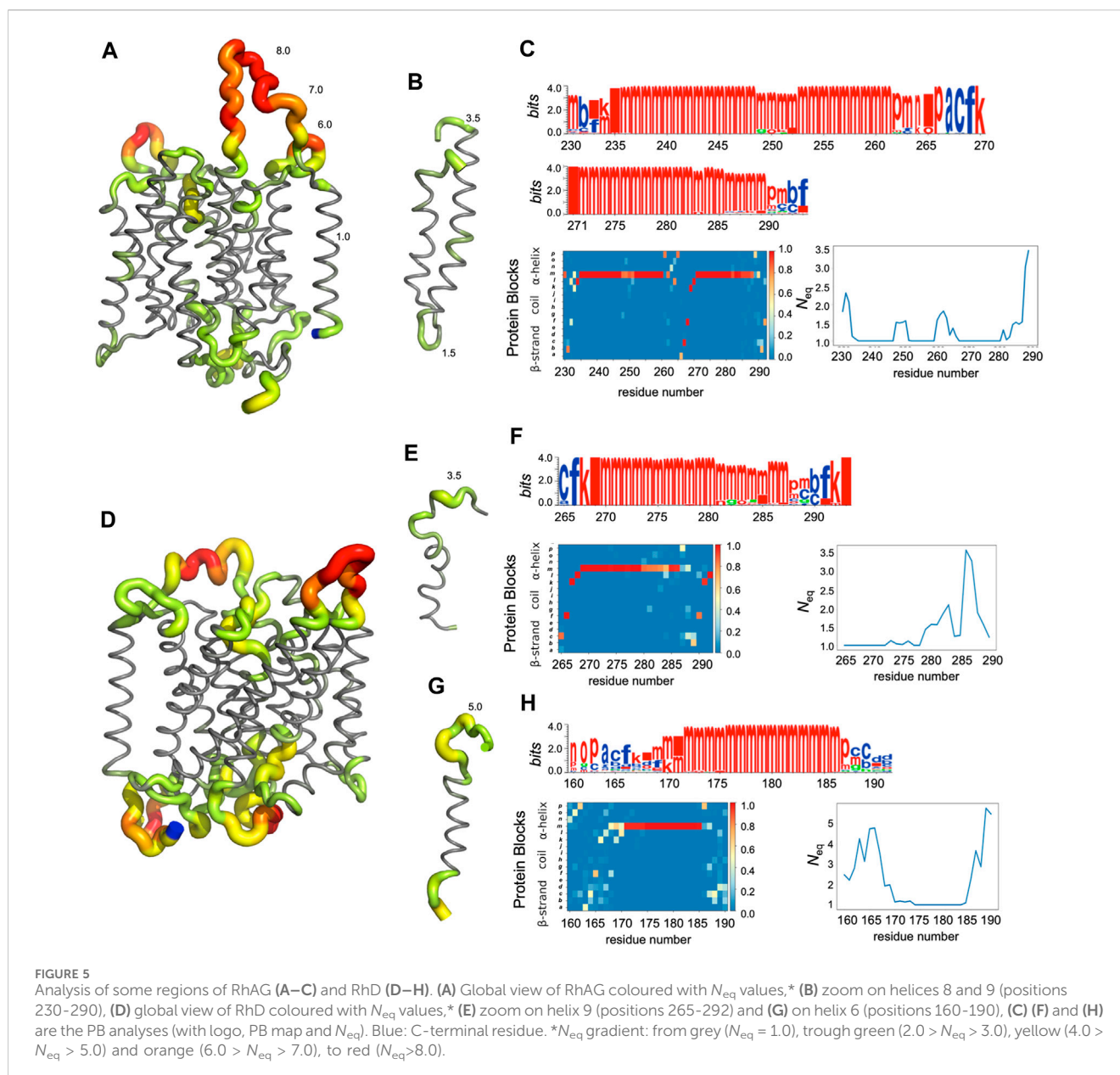
Significant variations in ΔPB were observed depending on the region of the protein. As expected, transmembrane regions with low

N_{eq} calculated for individual chains also had low ΔPB. Most of the loops had a limited ΔPB of 0.6 or less, i.e., the difference was less than 1/3 of the sampling. The maximum variability (see red arrows in Figure 6 for ΔPB greater than 1.0) was observed in the largest, disordered loops (with a N_{eq} of 8.0 or more): loops 2, and 12 for RhAG (respectively the first and the sixth extracellular loops, respectively), and loops 2, 7 and 12 for RhD (respectively the first extracellular loop, the fourth intracellular loop and the sixth extracellular loop, respectively). The last disordered loop with a N_{eq} of 8.0 was the 4th loop of RhAG; it was also associated with a ΔPB at the upper end of the ΔPB values, at 0.8.

4 Discussion and conclusion

In this work, we have proposed state-of-the-art trimer models for the human RhAG and RhD proteins, of major importance in transfusion medicine, and performed substantial molecular dynamics simulations (over 3μs total) to sample the local conformations of the RhAG and RhD proteins.

All the trimer structural models used in the molecular dynamics study were built by homology, using the same RhCG structure as a template (Figure 1). This choice was justified by a preliminary

FIGURE 5
Analysis of some regions of RhAG **(A–C)** and RhD **(D–H)**. **(A)** Global view of RhAG coloured with N_{eq} values,* **(B)** zoom on helices 8 and 9 (positions 230-290), **(D)** global view of RhD coloured with N_{eq} values,* **(E)** zoom on helix 9 (positions 265-292) and **(G)** on helix 6 (positions 160-190), **(C) (F)** and **(H)** are the PB analyses (with logo, PB map and N_{eq}). Blue: C-terminal residue. *N_{eq} gradient: from grey (N_{eq} = 1.0), trough green (2.0 > N_{eq} > 3.0), yellow (4.0 > N_{eq} > 5.0) and orange (6.0 > N_{eq} > 7.0), to red (N_{eq}>8.0).

estimation of the quality of the RhAG$_3$ trimer models, obtained using different approaches. Indeed, in our previous study (de Brevern et al., 2018), the best monomer model for RhD was obtained with a multi-template approach based on the structural templates of RhCG (Gruswitz et al., 2010) and NeRH50 (Li et al., 2007; Lupo et al., 2007). However, among the models of RhAG$_3$ trimer built using as a template NeRH50 or RhCG alone, or together in the multi-template approach, the best structural models were obtained using RhCG alone as a template according to both DOPE (Shen and Sali, 2006) and HPM (de Brevern and Hazout, 2003; Esque et al., 2015; Téletchéa et al., 2023) scores (Supplementary Figure S5). The high sequence identity of RhCG with both RhAG (52.32%) and RhD (31.65%) made this template appropriate to build all trimer models.

It is important to note that correct positioning in the membrane can be obtained only for complete trimer models. Attempts to insert RhD and RhAG monomers in the lipid bilayer using OPM (Lomize et al., 2006; Lomize et al., 2011) and OREMPRO (Postic et al., 2016b) result in a high variability and inconsistencies between the models (see Supplementary Figure S6A), clearly showing that the monomeric forms cannot exist in a membrane environment on their own. Conversely, trimers are positioned harmoniously in the bilayer (see Supplementary Figure S6B), further corroborating the necessity of modelling the whole trimer rather than only one subunit, for a relevant analysis of protein dynamics.

Two major events occurred during the course of this work. First, AlphaFold2 greatly advanced the field of molecular modelling (Jumper et al., 2021). On average, it has made it possible to obtain 25% more residues and to reach 10% of the additional human proteome (Tunyasuvunakool et al., 2021). However, it also has recognized limitations (Tourlet et al., 2023), particularly for transmembrane proteins (He et al., 2023; Karelina et al., 2023).
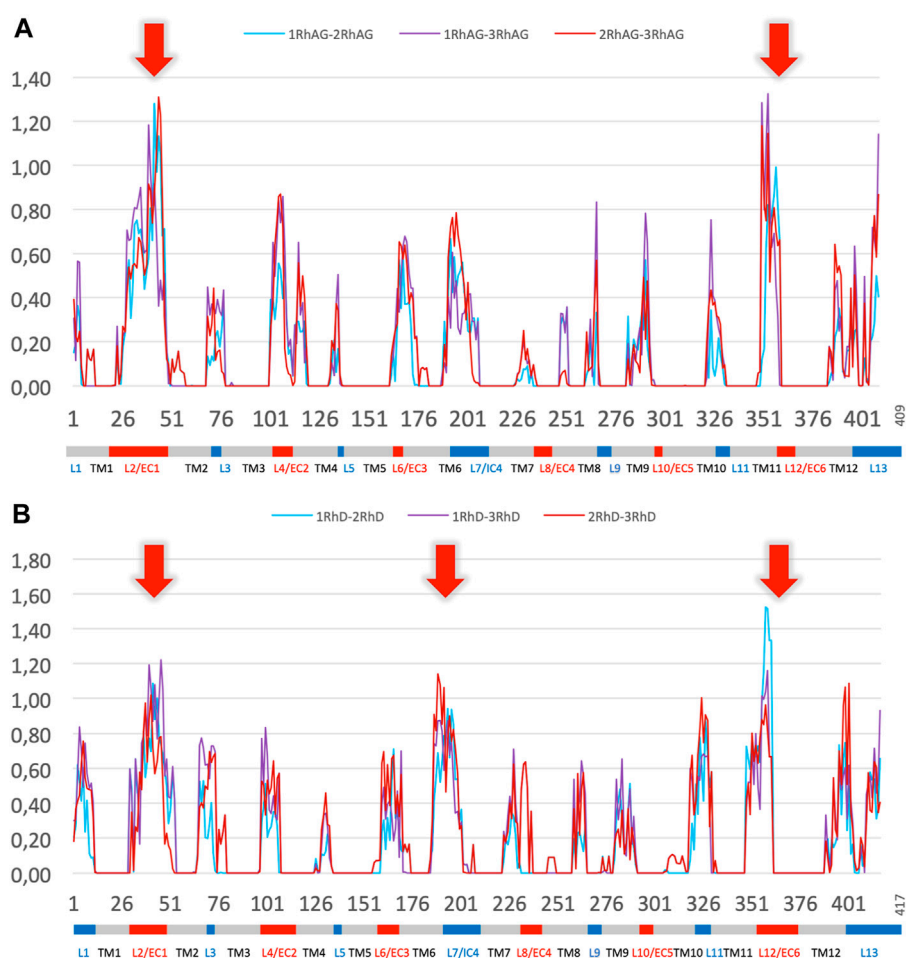
FIGURE 6
ΔPB for the different trimer compositions for RhAG and RhD. Panel **(A)** RhAG. Blue: ΔPB between trimers with 1 RhAG subunit and trimers with 2 RhAG subunits; Purple: ΔPB between trimers with 1 RhAG subunit and 3 RhAG subunits; Red: ΔPB between trimers with 2 RhAG subunits and 3 RhAG subunits. Red arrows indicate the highest ΔPB for RhAG, in loops 2 and 12. Panel **(B)** RhD. Blue: ΔPB between trimers with 1 RhD subunit and trimers with 2 RhD subunits; Purple: ΔPB between trimers with 1 RhD subunit and 3 RhD subunits; Red: ΔPB between trimers with 2 RhD subunits and 3 RhD subunits. Red arrows indicate the highest ΔPB for RhD, in loops 2, 7 and 12 for RhD.

The generation of an AlphaFold2 model to compare ours to was disappointing, with a lack of cohesion in the helical domain making insertion into a membrane, for example, inefficient. This underperformance of AlphaFold2 is consistent with what we have recently observed when modelling two other blood group proteins, BCAM and ERMAP (Floch et al., 2023a; Floch et al., 2023b). Very recently, the structure of RhCE complexed with RhAG was obtained by cryo-EM (Vallese et al., 2022). We built a model based on this new structure, using the methodology described above, and compared it to the model used in the present study. The RMS deviation was less than 0.2 Å on Cα, demonstrating the high reliability of the model we had previously built and were using, and allowing its use without reserve for future studies.

As underlined by a multiple studies, the AlphaFold2 model does not work very well for membrane proteins, and people are working on new AI-based approaches that could improve it in the future with already interesting results (Wang et al., 2022; Huang and Ecker, 2023). A high fraction of membrane proteins form oligomeric structures, however, by often predicting them as monomeric,

AlphaFold2 offers models that are not very useable (Dobson et al., 2023). Several approaches are being developed to overcome these limitations and also discover possible multiple conformations, particularly for globular proteins (Silva et al., 2023). A major advantage of artificial intelligence approaches is their speed of execution (after the learning phase) and this could be a great advantage compared to molecular dynamics simulation approaches, which are very computationally expensive (Tian et al., 2021; Zhang et al., 2022).

During MD simulations, the RhD and RhAG monomers behaved in a similar manner regardless of the trimer composition. The variations that were observed between the trimers of different compositions were similar to the variations that could be observed between the replicates for the exact same system. The strength of our work is to have sampled a large number of conformations, by proposing several closely related models, i.e., with the odd monomer at different chain positions (A, B or C) within the trimer models. Interestingly, MD simulations of the non-physiological RhD$_3$ homotrimer model did not show major

differences with MD simulations of RhD with other trimer compositions. Careful analysis of the global and local conformational variability of the systems considered here demonstrated the stability of the developed models with respect to the trimer composition and to the exact position of the RhAG and RhD chains.

The proposed RhAG/RhD trimer models can be used to delve into the details of the dynamics and interactions of Rh systems and variants, even if we do not model all elements of the physiological system. Firstly, we did not consider the post-translational modifications (PTM) of the proteins. Residues 12 and 186 of RhD proteins are S-palmitoylated, and RhAG proteins, as indicated by their name, are glycosylated (Moore and Green, 1987; Hartel-Schenk and Agre, 1992; Anstee and Tanner, 1993). However, the introduction of these PTMs would significantly increase the complexity of the (already large) system and its analysis. The second limitation of our approach is that human Rh proteins are part of a larger protein complex (Nicolas et al., 2003), and the partner proteins were not modelled in this work. Nonetheless, none of the other proteins of the Rh complex are strictly required for the expression of the Rh antigens, as revealed by the persistence of the Rh proteins in individuals lacking the different proteins of the complex (Colin, 2002), and supported by experimental data (Mouro-Chanteloup et al., 2003).

The proposed models are of crucial importance for the investigation of the impact of genetic variants on the blood group antigens' structure and function. The D antigen, carried by the RhD protein, is crucial in transfusion medicine, due to its high variability with more than 700 alleles reported to-date (Floch et al., 2021a) and to the clinical consequences of anti-D antibodies (Chou and Westhoff, 2010). Assessing the consequences of these genetic variants is a challenge in transfusion medicine. The limited resource of D-negative blood should be preserved for D-negative recipients and carriers of clinically relevant RhD genetic variants (susceptible to producing anti-D if exposed to the standard D antigen). This work is an important step towards a better understanding of the D antigen conformations and the identification of relevant epitopes. Experimental studies and international collaborative efforts have tested many monoclonal anti-D antibodies with RhD genetic variants to try to study the epitopes to the RhD protein (Cartron et al., 1996; Scott, 1996; Scott, 2002). These experimental studies have not considered the 3D conformations of epitopes and the interpretation of the experimental results remain limited.

Using our molecular dynamics simulations to study epitopes, and Rh genetic variants could be particularly interesting and shed light on various experimental findings. The positions of the two biggest extracellular loops (loops 2 and 12) of RhD within the model are consistent with experimental data: mutations in these loops alter the antigens significantly and are responsible for alloimmunisation (Cartron et al., 1996). Other residues are known to play a role in the expression of D antigen epitopes despite their location in transmembrane (TM) domains: residues 169, 170 and 172 in the 6th TM domain (Rouillac et al., 1995), residue 238 is in the 8th TM domain (Noizat-Pirenne et al., 2002), and residue 283 is in the 9th TM domain (von Zabern et al., 2013). Mutations in TM domains could have a long-range effect on

extracellular parts of the protein, indirectly modifying the protein's epitopes. Molecular dynamics simulations of the numerous RhD and RhAG genetic variants (Chou and Westhoff, 2010) will be interesting to analyse for subtle conformational changes. In a close future, it would also be interesting to carry out simulations using coarse-grained approaches, which allow long simulation times to be achieved with relatively good reliability and are particularly well suited to transmembrane systems (Rao et al., 2020; Souza et al., 2021).

Web resources

Name	url
Clustal Omega	http://www.ebi.ac.uk/Tools/msa/clustalo/
ConSurf	https://consurf.tau.ac.il/
HMMTOP	http://www.enzim.hu/hmmtop/
HHpred	http://toolkit.tuebingen.mpg.de/hhpred
HPM	http://www.dsimb.inserm.fr/dsimb_tools/hpmscore/
MAIDEN	http://www.dsimb.inserm.fr/maiden/
MEMSAT	http://bioinf.cs.ucl.ac.uk/psipred/?memsatsvm=1
Modeller 9.12	https://salilab.org/modeller/
MolProbity	http://molprobity.biochem.duke.edu
OPM	http://opm.phar.umich.edu
OREMPRO	http://www.dsimb.inserm.fr/dsimb_tools/OREMPRO/
PBxplore	https://github.com/pierrepo/PBxplore
PDBePISA	https://www.ebi.ac.uk/pdbe/pisa/
Procheck	https://www.ebi.ac.uk/thornton-srv/software/PROCHECK/
ProFit 3.1	http://www.bioinf.org.uk/programs/profit/
Protein DataBank	http://www.rcsb.org/pdb/
Protter	http://wlab.ethz.ch/protter/start/
PSI-BLAST	http://blast.ncbi.nlm.nih.gov/
RCSB PDB	https://www.rcsb.org/
UniProt	http://www.uniprot.org/

Data availability statement

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

Author contributions

AF: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing–original draft, Writing–review and editing. TG: Data curation, Formal Analysis, Visualization, Writing–review and editing. FP: Data curation, Investigation, Resources, Writing–review and editing. CT: Formal Analysis,

Investigation, Visualization, Writing–review and editing. AB: Conceptualization, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing–original draft, Writing–review and editing.

Conflict of interest

The 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.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2024.1360392/full#supplementary-material

References

Abraham, M., Murtola, T., Schulz, R., Páll, S., Smith, J., Hess, B., et al. (2015). GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 2, 19–25. doi:10.1016/j.softx.2015.06.001

Akhila, M. V., Narwani, T. J., Floch, A., Maljković, M., Bisoo, S., Shinada, N. K., et al. (2020). A structural entropy index to analyse local conformations in intrinsically disordered proteins. *J. Struct. Biol.* 210 (1), 107464. doi:10.1016/j.jsb.2020.107464

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids Res.* 25 (17), 3389–3402. doi:10.1093/nar/25.17.3389

Anstee, D. J., and Tanner, M. J. (1993). 4 Biochemical aspects of the blood group Rh (Rhesus) antigens. *Bailliere's Clin. Haematol.* 6 (2), 401–422. doi:10.1016/s0950-3536(05)80152-0

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., et al. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic acids Res.* 44 (W1), W344–W350. doi:10.1093/nar/gkw408

Avent, N. D., Ridgwell, K., Mawby, W. J., Tanner, M. J., Anstee, D. J., and Kumpel, B. (1988). Protein-sequence studies on Rh-related polypeptides suggest the presence of at least two groups of proteins which associate in the human red-cell membrane. *Biochem. J.* 256 (3), 1043–1046. doi:10.1042/bj2561043

Barnoud, J., Santuz, H., Craveur, P., Joseph, A. P., Jallu, V., de Brevern, A. G., et al. (2017). PBxplore: a tool to analyze local protein structure and deformability with Protein Blocks. *PeerJ* 5, e4013. doi:10.7717/peerj.4013

Berendsen, H., Postma, J., van Gunsteren, W., DiNola, A., and Haak, J. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81, 3684–3690. doi:10.1063/1.448118

Berman, H. M., Battistuz, T., Bhat, T. N., Bluhm, W. F., Bourne, P. E., Burkhardt, K., et al. (2002). The protein data bank. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* 58 (1), 899–907. doi:10.1107/s0907444902003451

Bornot, A., Etchebest, C., and de Brevern, A. G. (2011). Predicting protein flexibility through the prediction of local structures. *Proteins* 79 (3), 839–852. doi:10.1002/prot.22922

Burton, N. M., and Daniels, G. (2011). Structural modelling of red cell surface proteins. *Vox Sang.* 100 (1), 129–139. doi:10.1111/j.1423-0410.2010.01424.x

Callebaut, I., Dulin, F., Bertrand, O., Ripoche, P., Mouro, I., Colin, Y., et al. (2006). Hydrophobic cluster analysis and modeling of the human Rh protein three-dimensional structures. *Transfus. clinique Biol.* 13 (1-2), 70–84. doi:10.1016/j.tracli.2006.02.001

Cartron, J. P., Rouillac, C., Le Van Kim, C., Mouro, I., and Colin, Y. (1996). Tentative model for the mapping of D epitopes on the RhD polypeptide. *Transfus. clinique Biol.* 3 (6), 497–503. doi:10.1016/s1246-7820(96)80070-x

Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., et al. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* 66 (1), 12–21. doi:10.1107/s0907444909042073

Chou, S. T., and Westhoff, C. M. (2010). The Rh and RhAG blood group systems. *Immunohematology* 26 (4), 178–186. doi:10.21307/immunohematology-2019-217

Colin, Y. (2002). Rh proteins: a family of structural membrane proteins with putative transport activity. *Vox Sang.* 83 (1), 179–183. doi:10.1111/j.1423-0410.2002.tb05296.x

Conroy, M. J., Bullough, P. A., Merrick, M., and Avent, N. D. (2005). Modelling the human rhesus proteins: implications for structure and function. *Br. J. Haematol.* 131 (4), 543–551. doi:10.1111/j.1365-2141.2005.05786.x

Consortium, U. (2015). UniProt: a hub for protein information. *Nucleic acids Res.* 43, D204–D212. doi:10.1093/nar/gku989

Darden, T., Perera, L., Li, L., and Pedersen, L. (1999). New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations. *Struct. Lond. Engl.* 7 (3), R55–R60. doi:10.1016/s0969-2126(99)80033-1

de Brevern, A. G. (2010). 3D structural models of transmembrane proteins. *Methods Mol. Biol. Clift. NJ* 654, 387–401. doi:10.1007/978-1-60761-762-4_20

de Brevern, A. G., Etchebest, C., and Hazout, S. (2000). Bayesian probabilistic approach for predicting backbone structures in terms of protein blocks. *Proteins* 41 (3), 271–287. doi:10.1002/1097-0134(20001115)41:3<271::aid-prot10>3.0.co;2-z

de Brevern, A. G., Floch, A., Barrault, A., Martret, J., Bodivit, G., Djoudi, R., et al. (2018). Alloimmunization risk associated with amino acid 223 substitution in the RhD protein: analysis in the light of molecular modeling. *Transfusion* 58 (11), 2683–2692. doi:10.1111/trf.14809

de Brevern, A. G., and Hazout, S. (2003). Hybrid protein model' for optimally defining 3D protein structure fragments. *Bioinforma. Oxf. Engl.* 19 (3), 345–353. doi:10.1093/bioinformatics/btf859

de Brevern, A. G., Valadié, H., Hazout, S., and Etchebest, C. (2002). Extension of a local backbone description using a structural alphabet: a new approach to the sequence-structure relationship. *Protein Sci. a Publ. Protein Soc.* 11 (12), 2871–2886. doi:10.1110/ps.0220502

de Haas, M., Thurik, F. F., Koelewijn, J. M., and van der Schoot, C. E. (2015). Haemolytic disease of the fetus and newborn. *Vox Sang.* 109 (2), 99–113. doi:10.1111/vox.12265

Dobson, L., Szekeres, L. I., Gerdán, C., Langó, T., Zeke, A., and Tusnády, G. E. (2023). TmAlphaFold database: membrane localization and evaluation of AlphaFold2 predicted alpha-helical transmembrane protein structures. *Nucleic acids Res.* 51 (1), D517–D522. doi:10.1093/nar/gkac928

Esque, J., Urbain, A., Etchebest, C., and de Brevern, A. G. (2015). Sequence-structure relationship study in all-α transmembrane proteins using an unsupervised learning approach. *Amino acids* 47 (11), 2303–2322. doi:10.1007/s00726-015-2010-5

Flegel, W. A., von Zabern, I., Doescher, A., Wagner, F. F., Vytisková, J., and Písacka, M. (2008). DCS-1, DCS-2, and DFV share amino acid substitutions at the extracellular RhD protein vestibule. *Transfusion* 48 (1), 25–33. doi:10.1111/j.1537-2995.2007.01506.x

Floch, A., Lomas-Francis, C., Vege, S., Brennan, S., Shakarian, G., de Brevern, A. G., et al. (2023a). A novel high-prevalence antigen in the Lutheran system, LUGA (LU24), and an updated, full-length 3D BCAM model. *Transfusion* 63 (4), 798–807. doi:10.1111/trf.17262

Floch, A., Lomas-Francis, C., Vege, S., Burgos, A., Hoffman, R., Cusick, R., et al. (2023b). Two new Scianna variants causing loss of high prevalence antigens: ERMAP model and 3D analysis of the antigens. *Transfusion* 63 (1), 230–238. doi:10.1111/trf.17182

Floch, A., Pirenne, F., Barrault, A., Chami, B., Toly-Ndour, C., Tournamille, C., et al. (2021b). Insights into anti-D formation in carriers of RhD variants through studies of 3D intraprotein interactions. *Transfusion* 61 (4), 1286–1301. doi:10.1111/trf.16301

Floch, A., Téletchéa, S., Tournamille, C., de Brevern, A. G., and Pirenne, F. (2021a). A review of the literature organized into a new database: RHeference. *Transfus. Med. Rev.* 35 (2), 70–77. doi:10.1016/j.tmrv.2021.04.002

Frishman, D., and Argos, P. (1995). Knowledge-based protein secondary structure assignment. *Proteins* 23 (4), 566–579. doi:10.1002/prot.340230412

Gelly, J. C., Joseph, A. P., Srinivasan, N., and de Brevern, A. G. (2011). iPBA: a tool for protein structure comparison using sequence alignment strategies. *Nucleic acids Res.* 39, W18–W23. doi:10.1093/nar/gkr333

Gruswitz, F., Chaudhary, S., Ho, J. D., Schlessinger, A., Pezeshki, B., Ho, C. M., et al. (2010). Function of human Rh based on structure of RhCG at 2.1 Å. *Proc. Natl. Acad. Sci. U. S. A.* 107 (21), 9638–9643. doi:10.1073/pnas.1003587107

Hartel-Schenk, S., and Agre, P. (1992). Mammalian red cell membrane Rh polypeptides are selectively palmitoylated subunits of a macromolecular complex. *J. Biol. Chem.* 267 (8), 5569–5574. doi:10.1016/s0021-9258(18)42803-7

He, X. H., You, C. Z., Jiang, H. L., Jiang, Y., Xu, H. E., and Cheng, X. (2023). AlphaFold2 versus experimental structures: evaluation on G protein-coupled receptors. *Acta Pharmacol. Sin.* 44 (1), 1–7. doi:10.1038/s41401-022-00938-y

Hess, B., Bekker, H., Berendsen, H., and Fraaije, J. (1997). LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* 18 (1463–1472), 1463–1472. doi:10.1002/(sici)1096-987x(199709)18:12<1463::aid-jcc4>3.3.co;2-l

Hildebrand, A., Remmert, M., Biegert, A., and Söding, J. (2009). Fast and accurate automatic structure prediction with HHpred. *Proteins* 77 (9), 128–132. doi:10.1002/prot.22499

Hoover, W. G. (1985). Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A, General Phys.* 31 (3), 1695–1697. doi:10.1103/physreva.31.1695

Huang, J., and Ecker, G. F. (2023). A structure-based view on ABC-transporter linked to multidrug resistance. *Mol. Basel, Switz.* 28 (2), 495. doi:10.3390/molecules28020495

Huang, W. C., Hsu, S. C., Huang, S. J., Chen, Y. J., Hsiao, Y. C., Zhang, W., et al. (2013). Exogenous expression of human SGLT1 exhibits aggregations in sodium dodecyl sulfate polyacrylamide gel electrophoresis. *Am. J. Transl. Res.* 5 (4), 441–449.

Jones, D. T. (2007). Improving the accuracy of transmembrane protein topology prediction using evolutionary information. *Bioinforma. Oxf. Engl.* 23 (5), 538–544. doi:10.1093/bioinformatics/btl677

Jorgensen, W., Chandrasekhar, J., Madura, J., Impey, R., and Klein, M. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935. doi:10.1063/1.445869

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596 (7873), 583–589. doi:10.1038/s41586-021-03819-2

Karelina, M., Noh, J. J., and Dror, R. O. (2023). How accurately can one predict drug binding modes using AlphaFold models? bioRxiv.

Khademi, S., O'Connell, J., 3rd, Remis, J., Robles-Colmenares, Y., Miercke, L. J., and Stroud, R. M. (2004). Mechanism of ammonia transport by Amt/MEP/Rh: structure of AmtB at 1.35 A. *Sci. (New York, NY)* 305 (5690), 1587–1594. doi:10.1126/science.1101952

Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* 372 (3), 774–797. doi:10.1016/j.jmb.2007.05.022

Laskowski, R., MacArthur, M., Moss, D., and Thornton, J. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26, 283–291. doi:10.1107/s0021889892009944

Lee, K. H., Benson, D. R., and Kuczera, K. (2000). Transitions from α to π helix observed in molecular dynamics simulations of synthetic peptides. *Biochemistry* 39 (45), 13737–13747. doi:10.1021/bi001126b

Léonard, S., Joseph, A. P., Srinivasan, N., Gelly, J. C., and de Brevern, A. G. (2014). mulPBA: an efficient multiple protein structure alignment method based on a structural alphabet. *J. Biomol. Struct. Dyn.* 32 (4), 661–668. doi:10.1080/07391102.2013.787026

Li, X., Jayachandran, S., Nguyen, H. H., and Chan, M. K. (2007). Structure of the Nitrosomonas europaea Rh protein. *Proc. Natl. Acad. Sci. U. S. A.* 104 (49), 19279–19284. doi:10.1073/pnas.0709710104

Lomize, A. L., Pogozheva, I. D., and Mosberg, H. I. (2011). Anisotropic solvent model of the lipid bilayer. 2. Energetics of insertion of small molecules, peptides, and proteins in membranes. *J. Chem. Inf. Model.* 51 (4), 930–946. doi:10.1021/ci200200k

Lomize, M. A., Lomize, A. L., Pogozheva, I. D., and Mosberg, H. I. (2006). OPM: orientations of proteins in membranes database. *Bioinforma. Oxf. Engl.* 22 (5), 623–625. doi:10.1093/bioinformatics/btk023

Lupo, D., Li, X. D., Durand, A., Tomizaki, T., Cherif-Zahar, B., Matassi, G., et al. (2007). The 1.3-Å resolution structure of Nitrosomonas europaea Rh50 and mechanistic implications for NH_3 transport by Rhesus family proteins. *Proc. Natl. Acad. Sci. U. S. A.* 104 (49), 19303–19308. doi:10.1073/pnas.0706563104

Martí-Renom, M. A., Stuart, A. C., Fiser, A., Sánchez, R., Melo, F., and Sali, A. (2000). Comparative protein structure modeling of genes and genomes. *Annu. Rev. biophysics Biomol. Struct.* 29, 291–325. doi:10.1146/annurev.biophys.29.1.291

McLachlan, A. (1982). Rapid comparison of protein structures. *Acta Crystallogr. Sect. A* 38, 871–873. doi:10.1107/s0567739482001806

Moore, S., and Green, C. (1987). The identification of specific Rhesus-polypeptide-blood-group-ABH-active-glycoprotein complexes in the human red-cell membrane. *Biochem. J.* 244 (3), 735–741. doi:10.1042/bj2440735

Mouro-Chanteloup, I., D'Ambrosio, A. M., Gane, P., Le Van Kim, C., Raynal, V., Dhermy, D., et al. (2002). Cell-surface expression of RhD blood group polypeptide is posttranscriptionally regulated by the RhAG glycoprotein. *Blood* 100 (3), 1038–1047. doi:10.1182/blood.v100.3.1038

Mouro-Chanteloup, I., Delaunay, J., Gane, P., Nicolas, V., Johansen, M., Brown, E. J., et al. (2003). Evidence that the red cell skeleton protein 4.2 interacts with the Rh membrane complex member CD47. *Blood* 101 (1), 338–344. doi:10.1182/blood-2002-04-1285

Narwani, T. J., Craveur, P., Shinada, N. K., Santuz, H., Rebehmed, J., Etchebest, C., et al. (2018). Dynamics and deformability of α-3₁₀- and π-helices. *Archives Biol. Sci.* 70, 21–31. doi:10.2298/abs170215022n

Nicolas, V., Le Van Kim, C., Gane, P., Birkenmeier, C., Cartron, J. P., Colin, Y., et al. (2003). Rh-RhAG/ankyrin-R, a new interaction site between the membrane bilayer and the red cell skeleton, is impaired by Rh(null)-associated mutation. *J. Biol. Chem.* 278 (28), 25526–25533. doi:10.1074/jbc.m302816200

Noizat-Pirenne, F. (2012). Relevance of alloimmunization in haemolytic transfusion reaction in sickle cell disease. *Transfus. clinique Biol.* 19 (3), 132–138. doi:10.1016/j.tracli.2012.03.004

Noizat-Pirenne, F., Lee, K., Pennec, P. Y., Simon, P., Kazup, P., Bachir, D., et al. (2002). Rare RHCE phenotypes in black individuals of Afro-Caribbean origin: identification and transfusion safety. *Blood* 100 (12), 4223–4231. doi:10.1182/blood-2002-01-0229

Nosé, S. (1984). A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* 81, 511–519. doi:10.1063/1.447334

Parrinello, M., and Rahman, A. (1981). Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys. (United States)* 52, 7182–7190. doi:10.1063/1.328693

Pham, B. N., Roussel, M., Gien, D., Ripaux, M., Carine, C., Le Pennec, P. Y., et al. (2013). Molecular analysis of patients with weak D and serologic analysis of those with anti-D (excluding type 1 and type 2). *Immunohematology* 29 (2), 55–62. doi:10.21307/immunohematology-2019-125

Postic, G., Ghouzam, Y., and Gelly, J. C. (2016b). OREMPRO web server: orientation and assessment of atomistic and coarse-grained structures of membrane proteins. *Bioinforma. Oxf. Engl.* 32 (16), 2548–2550. doi:10.1093/bioinformatics/btw208

Postic, G., Ghouzam, Y., Guiraud, V., and Gelly, J. C. (2016a). Membrane positioning for high- and low-resolution protein structures through a binary classification approach. *Protein Eng. Des. Sel. PEDS* 29 (3), 87–92. doi:10.1093/protein/gzv063

Rao, R., Diharce, J., Dugué, B., Ostuni, M. A., Cadet, F., and Etchebest, C. (2020). Versatile dimerisation process of translocator protein (TSPO) revealed by an extensive sampling based on a coarse-grained dynamics study. *J. Chem. Inf. Model.* 60 (8), 3944–3957. doi:10.1021/acs.jcim.0c00246

Rouillac, C., Colin, Y., Hughes-Jones, N. C., Beolet, M., D'Ambrosio, A. M., Cartron, J. P., et al. (1995). Transcript analysis of D category phenotypes predicts hybrid Rh D-CE-D proteins associated with alteration of D epitopes. *Blood* 85 (10), 2937–2944. doi:10.1182/blood.v85.10.2937.bloodjournal85102937

Sali, A., and Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234 (3), 779–815. doi:10.1006/jmbi.1993.1626

Scott, M. (1996). Rh serology--coordinator's report. *Transfus. clinique Biol.* 3 (6), 333–337. doi:10.1016/s1246-7820(96)80040-1

Scott, M. (2002). Section 1A: Rh serologyCoordinator's report. *J. de Soc. francaise de Transfus. Sang.* 9 (1), 23–29. doi:10.1016/s1246-7820(01)00211-7

Shen, M. Y., and Sali, A. (2006). Statistical potential for assessment and prediction of protein structures. *Protein Sci. a Publ. Protein Soc.* 15 (11), 2507–2524. doi:10.1110/ps.062416606

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

Silva, G. M., Cui, J. Y., Dalgarno, D. C., Lisi, G. P., and Rubenstein, B. M. (2023). Predicting relative populations of protein conformations without a physics engine using AlphaFold2. bioRxiv.

Silvy, M., Chapel-Fernandes, S., Callebaut, I., Beley, S., Durousseau, C., Simon, S., et al. (2012). Characterization of novel RHD alleles: relationship between phenotype, genotype, and trimeric architecture. *Transfusion* 52 (9), 2020–2029. doi:10.1111/j.1537-2995.2011.03544.x

Souza, P. C. T., Alessandri, R., Barnoud, J., Thallmair, S., Faustino, I., Grünewald, F., et al. (2021). Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nat. methods* 18 (4), 382–388. doi:10.1038/s41592-021-01098-3

St-Louis, M., Richard, M., Côté, M., Ethier, C., and Long, A. (2011). Weak D type 42 cases found in individuals of European descent. *Immunohematology* 27 (1), 20–24. doi:10.21307/immunohematology-2019-170

Téletchéa, S., Esque, J., Urbain, A., Etchebest, C., and de Brevern, A. G. (2023). Evaluation of transmembrane protein structural models using HPMScore. *BioMedInformatics* 3 (2), 306–326. doi:10.3390/biomedinformatics3020021

Thornton, N. M., and Grimsley, S. P. (2019). Proceedings from the international society of blood transfusion working party on immunohaematology, workshop on the clinical significance of red blood cell alloantibodies, september 9, 2016, dubai: clinical significance of antibodies to antigens in the ABO, MNS, P1PK, Rh, lutheran, kell, lewis, duffy, kidd, diego, yt, and xg blood group systems. *Immunohematology* 35 (3), 95–101. doi:10.21307/immunohematology-2020-021

Tian, H., Jiang, X., Trozzi, F., Xiao, S., Larson, E. C., and Tao, P. (2021). Explore protein conformational space with variational autoencoder. *Front. Mol. Biosci.* 8, 781635. doi:10.3389/fmolb.2021.781635

Tilley, L., Green, C., Poole, J., Gaskell, A., Ridgwell, K., Burton, N. M., et al. (2010). A new blood group system, RHAG: three antigens resulting from amino acid substitutions in the Rh-associated glycoprotein. *Vox Sang.* 98 (2), 151–159. doi:10.1111/j.1423-0410.2009.01243.x

Tormey, C. A., and Hendrickson, J. E. (2019). Transfusion-related red blood cell alloantibodies: induction and consequences. *Blood* 133 (17), 1821–1830. doi:10.1182/blood-2018-08-833962

Tourlet, S., Radjasandirane, R., Diharce, J., and de Brevern, A. G. (2023). AlphaFold2 update and perspectives. *BioMedInformatics* 3 (2), 378–390. doi:10.3390/biomedinformatics3020025

Trueba-Gómez, R., Rosenfeld-Mann, F., Baptista-González, H. A., Domínguez-López, M. L., and Estrada-Juárez, H. (2023). Use of computational biology to compare the theoretical tertiary structures of the most common forms of RhCE and RhD. *Vox Sang.* 118 (10), 881–890. doi:10.1111/vox.13509

Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Žídek, A., et al. (2021). Highly accurate protein structure prediction for the human proteome. *Nature* 596 (7873), 590–596. doi:10.1038/s41586-021-03828-1

Tusnády, G. E., and Simon, I. (2001). The HMMTOP transmembrane topology prediction server. *Bioinforma. Oxf. Engl.* 17 (9), 849–850. doi:10.1093/bioinformatics/17.9.849

Vallese, F., Kim, K., Yen, L. Y., Johnston, J. D., Noble, A. J., Calì, T., et al. (2022). Architecture of the human erythrocyte ankyrin-1 complex. *Nat. Struct. Mol. Biol.* 29 (7), 706–718. doi:10.1038/s41594-022-00792-w

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. (2005). GROMACS: fast, flexible, and free. *J. Comput. Chem.* 26 (16), 1701–1718. doi:10.1002/jcc.20291

von Zabern, I., Wagner, F. F., Moulds, J. M., Moulds, J. J., and Flegel, W. A. (2013). D category IV: a group of clinically relevant and phylogenetically diverse partial D. *Transfusion* 53 (11), 2960–2973. doi:10.1111/trf.12145

Wang, L., Zhong, H., Xue, Z., and Wang, Y. (2022). Improving the topology prediction of α-helical transmembrane proteins with deep transfer learning. *Comput. Struct. Biotechnol. J.* 20, 1993–2000. doi:10.1016/j.csbj.2022.04.024

Wu, E. L., Cheng, X., Jo, S., Rui, H., Song, K. C., Dávila-Contreras, E. M., et al. (2014). CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. *J. Comput. Chem.* 35 (27), 1997–2004. doi:10.1002/jcc.23702

Zhang, L., Huang, K., Yang, Y. I., and Shi, L. (2022). Editorial: combined artificial intelligence and molecular dynamics (AI-MD) methods. *Front. Mol. Biosci.* 9, 1012785. doi:10.3389/fmolb.2022.1012785

Zhang, Y., and Skolnick, J. (2005). TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic acids Res.* 33 (7), 2302–2309. doi:10.1093/nar/gki524