Challenging level of rigid-body approach involving numerical elements (CHLORAINE) applied to repeated elastin peptides

C. Depenveiller, H. Wong, J.M. Crowet, Laurent Debelle, S. Baud, Manuel Dauchez, N. Belloy

Journal of Structural Biology

Challenging Level Of Rigid-body Approach Involving Numerical Elements (CHLORAINE) applied to repeated elastin peptides.

--Manuscript Draft--

Manuscript Number:	JSB-23-82R1
Article Type:	VSI: Protein with Tandem Repeats
Keywords:	mesoscopic simulations, multiscale approach, repeated elastin-like peptides, elastomeric polymers, extracellular matrix
Corresponding Author:	Nicolas Belloy Université de Reims Champagne Ardenne Reims, Champagne-Ardenne FRANCE
First Author:	Camille Depenveiller, PhD
Order of Authors:	Camille Depenveiller, PhD
	Hua Wong, PhD
	Jean-Marc Crowet, PhD
	Laurent Debelle, PhD
	Stéphanie Baud, PhD
	Manuel Dauchez, PhD
	Nicolas Belloy, PhD
Abstract:	Elastic proteins and derived biomaterials contain numerous tandemly repeated peptides along their sequences, ranging from a few copies to hundreds. These repetitions are responsible for their biochemical, biological and biomechanical properties. These sequences are considered to be intrinsically disordered, and the variations in their behavior are actually mainly due to their high flexibility and lack of stable secondary structures originating from their unique amino acid sequences. Consequently, the simulation of elastic proteins and large elastomeric biomaterials using classical molecular dynamics is an important challenge. Here, we propose a novel approach that allows the application of the DURABIN protocol to repeated elastin-like peptides (r-ELPs) in a simple way. Four large r-ELPs were studied to evaluate our method, which was developed for simulating extracellular matrix proteins at the mesoscopic scale. After structure clustering applied on molecular dynamic trajectories of constitutive peptides (5-mers and 6-mers), the main conformations were used as starting points to define the corresponding primitives, further used as rigid body fragments in our program. Contributions derived from electrostatic and molecular hydrophobicity potentials were tested to evaluate their influence in the interactions during simple mesoscopic simulations. The CHLORAINE approach, despite the thinner granularity due to the size of the patterns used, was included in the DURABIN protocol and emerges as a promising way to simulate elastic macromolecular systems.
Suggested Reviewers:	Sergey Samsonov, PhD Project Leader at Laboratory of Molecular Modeling, University of Gdansk sergey.samsonov@ug.edu.pl specialist of Molecular Modeling in the field of Extracellular Matrix
	Carlos Rodriguez-Cabello, PhD Pr, University of Valladolid cabello@bioforge.uva.es specialist of Elastin and ELPs
	Brigida Bocchichio, PhD Pr, University of Basilicata brigida.bochicchio@unibas.it Specialist of Elastin and ELPs

Challenging Level Of Rigid-body Approach Involving Numerical Elements (CHLORAINE) applied to repeated elastin peptides.

C. Depenveiller[1,3], H. Wong[1,2], J.M. Crowet[1,2], L. Debelle[1], S. Baud[1,2], M. Dauchez[1,2], N. Belloy[1,2]*

Address

[1] Université de Reims Champagne Ardenne, CNRS, MEDyC UMR 7369, 51097 Reims, France

[2] Université de Reims Champagne Ardenne, Plateau de Modélisation Moléculaire Multi-Echelle (P3M), Maison de la simulation de Champagne Ardenne (MaSCA), 51097 Reims, France

[3] current address: Université de Picardie Jules Verne, CNRS, GEC UMR 7025, 80039 Amiens, France

* corresponding author: nicolas.belloy@univ-reims.fr

Abstract

Elastic proteins and derived biomaterials contain numerous tandemly repeated peptides along their sequences, ranging from a few copies to hundreds. These repetitions are responsible for their biochemical, biological and biomechanical properties. These sequences are considered to be intrinsically disordered, and the variations in their behavior are actually mainly due to their high flexibility and lack of stable secondary structures originating from their unique amino acid sequences. Consequently, the simulation of elastic proteins and large elastomeric biomaterials using classical molecular dynamics is an important challenge. Here, we propose a novel approach that allows the application of the DURABIN protocol to repeated elastin-like peptides (r-ELPs) in a simple way. Four large r-ELPs were studied to evaluate our method, which was developed for simulating extracellular matrix proteins at the mesoscopic scale. After structure clustering applied on molecular dynamic trajectories of constitutive peptides (5-mers and 6-mers), the main conformations were used as starting points to define the corresponding primitives, further used as rigid body fragments in our program. Contributions derived from electrostatic and molecular hydrophobicity potentials were tested to evaluate their influence in the interactions during simple mesoscopic simulations. The CHLORAINE approach, despite the thinner granularity due to the size of the patterns used,

was included in the DURABIN protocol and emerges as a promising way to simulate elastic macromolecular systems.

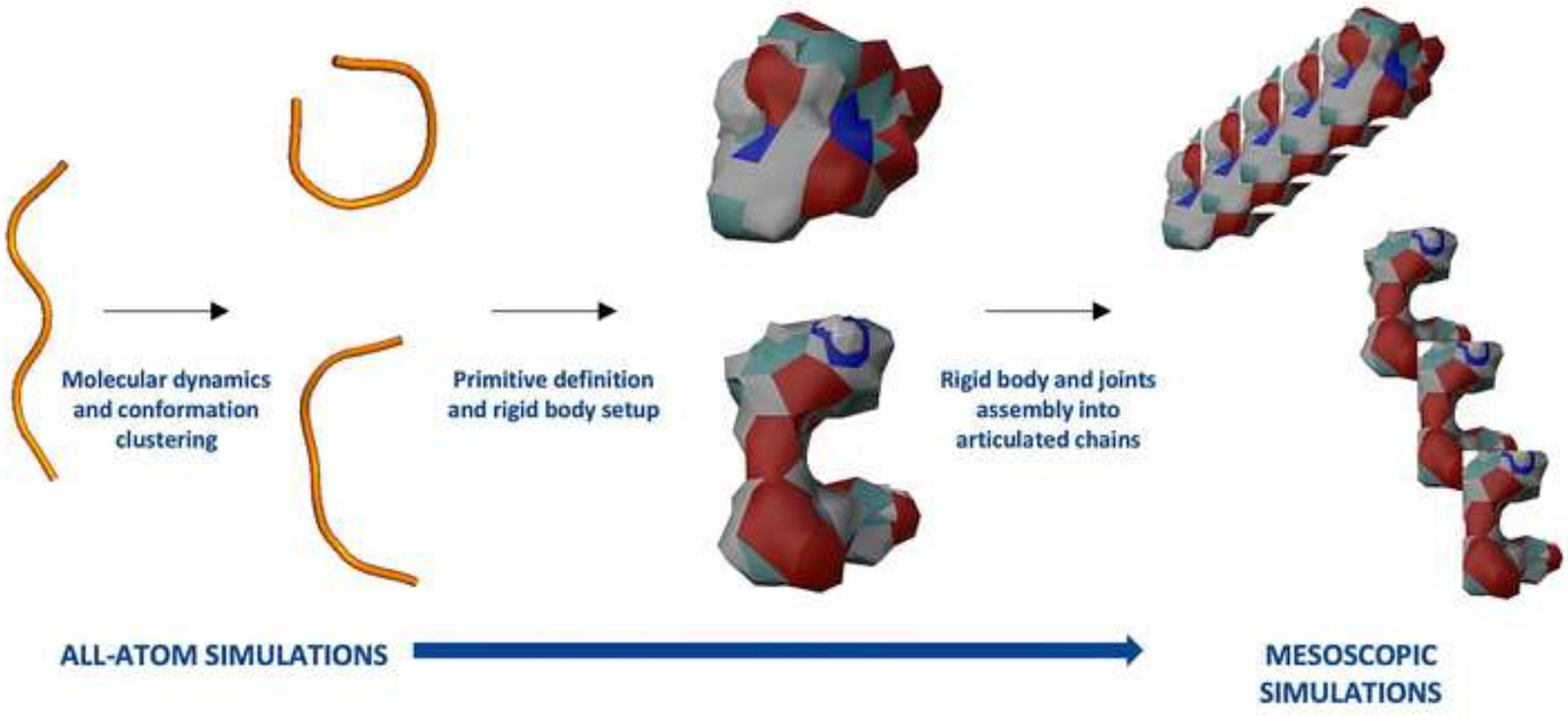
Keywords

mesoscopic simulations, multiscale approach, repeated elastin-like peptides, elastomeric polymers, extracellular matrix

Abbreviations

ECM, extracellular matrix; r-ELPs, repeated elastin-like peptides

Graphical Abstract revised



- Elastin, its monomer tropoelastin and repeated elastin-like peptides (r-ELPs) are of great interest in the field of biomaterials and biocompatibility.
- Classical molecular dynamics methods are not always suitable for the simulation of long sequences with tandem repeats.
- Rigid body dynamics allows the simulation of very large molecular systems of the extracellular matrix.
- This new way to simulate r-ELPs helps in the study of very large macrosystems and/or biomaterials with repeated patterns.

Introduction

Elastomeric proteins are found throughout biology in numerous different species (Muiznieks and Keeley, 2017). They demonstrate a large amount of mechanical properties that allow characteristic mechanisms such as extensibility, stiffness, tensile strength and many other properties. Most of these elastic proteins present tandemly repeated sequences and, at the supramolecular level, they may also propose series of tandemly-repeated proteins (Jorda et al., 2010). The repeated motifs of these elastomers arise from specific sequences. In the case of abductin, a protein from bivalve mollusks, two consensus sequences are found, GGFGGMGGG and GFFMGGGNAG (Cao et al., 1997). In insects, resilin harbors the following repeated motifs, GGRPSDSYGAPGGGN, AQTPSSQYPAG (Lyons et al., 2007). In spider silk fibroins, the fragments GAGAGS and GAGAGY are present in numerous motifs (Zhou, 2000). In mammals, titin, collagens and elastin (Gosline et al., 2002), proteins found in vertebrate muscles and connective tissues respectively, present a significant number of tandemly repeated patterns.

We have focused our work for several decades on the study and role of elastin in the extracellular matrix (ECM). Elastin is an essential component of various tissues that require elasticity. In humans, elastin endows the ability to stretch and recoil to skin, lungs and large arteries, thereby playing an important role in their physiological function (Le Page et al., 2019).

Tropoelastin, the constitutive monomer of the elastin, is a very peculiar structural protein and unfortunately the structure of this protein remains unknown, since X-rays crystallography and/or Nuclear Magnetic Resonance methods failed to unravel this structure at the atomic level. Tarakanova et al. have proposed a 698 residue-long tropoelastin model using SAXS experimental methods in conjunction with molecular modeling simulations (Tarakanova et al., 2018). However, the use of this model is limited, since the data is an envelope (no atomic coordinates) and this envelope only accounts for a single conformation, which is reductive compared to the inherent flexibility of tropoelastin.

Several studies using molecular dynamics with both all-atom and/or coarse-grained force fields have been published on constitutive peptides of tropoelastin (Seo et al., 2012). Although these simulations lead to comprehensive behaviors of elastin-like peptides (ELPs), their structure-function dynamics relationships are still not easily obtained, be it for elastic peptides or larger molecular systems. Most all-atom simulations are performed using force fields derived from amino acid peptides found in globular proteins. CHARMM36m (Huang et al.,

2017) has been developed to better understand (among other systems) intrinsically disordered proteins. However, it is not really suitable for simulations of elastomeric proteins and/or elastic properties, which remain a challenging task.

As discussed above, the very nature of tropoelastin sequence contributes to the limitations of this type of simulation. Full-length human tropoelastin comprises 786 residues and 82% of its sequence consists of only 5 amino acids: glycine (221), alanine (164), valine (98), proline (96), leucine (45), found in numerous repetitive and overlapping patterns. Among these, some 5-mers and 6-mers are found several times.

Due to the lack of dedicated numerical methods, numerical simulations of large elastic peptides or biomaterials are extremely difficult, if not impossible with current methods. Since 2014, we develop a very efficient mesoscopic rigid body approach to simulate very large multi-domains in the ECM (Wong et al., 2018). We have used the Unity3D game and physics engines (along with rigid body dynamics) to propose an application called DURABIN. Using this method, it is possible to study tens to hundreds of proteins in a very simple way. We have validated our approach by studying a peculiar ECM environment, the basement membrane.

In this work, we propose a proof of concept using our mesoscopic tool DURABIN to study elastic systems derived from tandem r-ELPs. This work leads to the Challenging Level Of Rigid-body Approach Involving Numerical Elements - CHLORAINE method applied to r-ELPs. In a multi-scale understanding of biological systems, our study attempts to better decipher the structure/dynamics of homopolymers of small peptides. Indeed, insight into the mesoscopic details of biopolymers derived from elastin is particularly important for understanding the molecular mechanisms underlying their functions.

1. Methods & Materials

1.1 Peptides

Elastin and its constitutive monomer tropoelastin possess numerous repeated fragments that allow the function of these molecules. In this work, the corresponding selected motifs arise from r-ELP penta-peptides VPGXG with X as V, E or K amino acids and hexa-peptides VGZAPG with Z as V or L amino acids. The 4 following homopolymers were investigated:

poly-VPGVG: $(VPGVG)_{27}$

poly-VPGEG: $(VPGEG)_{27}$

poly-VPGKG: $(VPGKG)_{27}$

poly-exon24-like: $[(VGLAPG)(VGVAPG)_2]_{27}$

leading to a respective length of 135 amino acids for the three first peptides, and of 486 amino acids for the last one. n=27 has been identified as a "minimal" length of homopolymers used in biomaterial experimental works for the first 3 sequences (Rodriguez-Cabello et al., 2021), and the last sequence is correlated with the motif found in the exon 24 of human-TE and used as mini-elastin fragment by F. Keeley (Muiznieks and Keeley, 2017). For the $(VPGXG)_{27}$ polypeptides, the role of the X amino acid is evaluated. The comparison with the poly-exon24-like homopolymer allows to study the impact of the primitive length (a building block of 5 amino acids for $(VPGXG)_{27}$ compared to a building block of 18 amino acids for $[(VGLAPG)(VGVAPG)_2]_{27}$).

1.2 Molecular dynamics simulations

Similar modeling and molecular dynamics simulation procedures were used for each individual repeat peptide (5-mers or 6-mers). Each r-ELP was built in an extended conformation using PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC). After 50,000-step minimization using a steepest descent algorithm, 100-ps NVT equilibration and 100-ps NPT equilibration, multiple 500-ns molecular dynamics simulations were performed using GROMACS software (Berendsen et al., 1995) with the OPLS-AA/L force field. Each peptide was solvated in a cubic water box with a generic equilibrated 3-point solvent model (TIP3P), and with a distance of 1 nm between the solute and the box. Na^+ and Cl^- ions were added at a concentration of 0.15 M.

After running the production step of molecular dynamics, the conformations of the peptides along the trajectories were visualized with VMD (Humphrey et al., 1996).

1.3 Clustering leading to the design of primitives

To obtain the major conformations for each of these peptides, we have performed clustering on the conformations sampled during the trajectory, allowing us to distribute the different conformations in different groups. The clustering tool *gmx cluster* with GROMOS method (Daura et al., 1999) available in GROMACS and the TTClust algorithm (Tubiana et al., 2018) have been used to access the major conformations. In the latter method, a distance matrix containing the RMSD values between all pairs of frames of the trajectory is generated.

A second matrix is then generated from this distance matrix using a hierarchical clustering function, so that a dendrogram of this matrix can be generated. For each motif observed in our r-ELPs, this clustering allowed us to identify the main clusters.

The Blender software (https://www.blender.org) is then used to generate representations of the main conformations derived from the clustering step. These representations, called primitives, are simplified shapes describing the conformations of the peptide. The primitives are used as "blocks" to represent r-ELP motifs in the DURABIN mesoscope.

1.4 Electrostatic and molecular hydrophobicity potentials

In order to evaluate the influence of physicochemical properties of our different elastic peptides, we have performed the Adaptative Poisson-Boltzmann Solver (ABPS) to determine electrostatic contributions. APBS is a method for the numerical solution of the Poisson-Boltzmann equation, a continuum model for describing electrostatic interactions (Holst and Saied, 1995). Knowledge of the electrostatic positive and negative contributions is very important to understand the interactions between different molecules. In the case of r-ELPs, due to their intrinsic functional flexibility, these contributions are observed on the whole peptide, but are also present along the chain, around peptide bonds.

In the same way, we used molecular hydrophobicity potential (MHP) (Brasseur, 1991) to get hydrophobic and hydrophilic contributions. The calculation of the MHP along r-ELPS is based on the molecular lipophilicity potential (Furet et al., 1988) and the hydrophobicity potential developed by Fauchère (Fauchère et al., 1988). These parameters are defined by considering a molecule surrounded by organic solvent molecules and by assuming that the overall hydrophobicity is the sum of the hydrophobicity of its different atoms as described by Lins et al. (Lins et al., 2003).

By using the electrostatics and hydrophobicity/hydrophilicity potentials computed as explained above, we could obtain and display, on the peptide surface, the corresponding surfaces of positive/negative electrostatics and hydrophobicity/hydrophilicity. From our simulations of tropoelastin peptides, it appears that electrostatics correspond to smaller surfaces, thus weaker contributions, compared to hydrophilicity/hydrophobicity ones and the hydrophobic areas are much more important than the hydrophilic areas even if this latter

cannot be neglected. Knowing the size of each area is very important to define correctly the corresponding shapes in the DURABIN mesoscope.

1.5 Rigid body simulations

Once the primitives have been created with Blender and validated, they are used in Unity3D (https://unity.com) to apply the rigid body approach. Unity3D is a cross-platform game engine that has many applications beyond video games, such as data visualization of simulation trajectories. Some software packages are chosen for modeling and simulations in the virtual reality model DURABIN and are therefore also used to apply the rigid body approach to the case of tropoelastin at the mesoscopic scale. In Unity3D, we use the primitives set up from Blender to define rigid bodies and model polypeptides as it was previously done for protein domains (Wong et al., 2022). Indeed, the first step is to set up colliders on the primitives. Colliders are components that define the shape and occupancy of objects in space. For a given primitive, the number and spatial arrangement of colliders depends entirely on the conformation of the relative primitive, as well as on the presence (or not) of hydrophilicity/hydrophobicity and/or electrostatic contributions. This step allows to set up the rigid body corresponding to the primitive/conformation under consideration. This process is repeated for each conformation we want to take into account, in order to build a library of rigid bodies representing the different motifs of tropoelastin.

In this protocol applied to elastin polypeptides, the difference with the original DURABIN is the granularity in the representation of the molecules. Indeed, in DURABIN the proteins are modeled by rigid bodies corresponding to domains of 100-200 amino acids. However, in the CHLORAINE method, the granularity is thinner but still larger than for coarse-grained simulations which use designed beads of a few atoms. The CHLORAINE approach defines rigid bodies in elastin polypeptides corresponding rather to tropoelastin motifs of 5 to 18 amino acids.

The next step is to use the rigid bodies also called building blocks from the library, in order to model longer polypeptides with motifs from tropoelastin. A polypeptide can be made up of several rigid bodies (of the same type or of different types) connected by molecular joints. These joints are objects that represent physical constraints and that we specifically define in this work so that they act as articulations between rigid bodies. Indeed, joints are set

up to describe angular constraints through Phi and Psi angles that can be adapted to each motif.

Once the polypeptides are modeled, they can be simulated and the DURABIN simulator allows to compute contacts between rigid bodies during the simulation. In this way, we can get this data and provide matrices of intramolecular (between rigid bodies from the same polypeptide) and intermolecular (between rigid bodies from different polypeptides) interactions.

Thereby, elastomeric polypeptides can be modeled as articulated chains of small rigid bodies and simulated at the mesoscopic scale.

2. Results – Discussion

Simulation of very large molecular systems remains a very complicated task, despite recent progress such as coarse-grained force fields and the optimization of molecular dynamics software running on Graphical Process Unit GPU. Very recently, we have developed a very simple method, called DURABIN, to study the crowding of proteins (Wong et al., 2018). This method uses the rigid body approach to study the ECM in a mesoscopic environment. In the simulated model, each rigid body, also called a module or a building block, corresponds to a specific structural domain of a protein. The proteins already present in the model are type IV collagen, laminin, nidogen, integrins, and perlecan (Wong et al., 2022). For these proteins, the rigid bodies, defined to fully model them, correspond to each constitutive domain/module of the protein studied. So far, in DURABIN, rigid bodies represent domains ranging from 50 to 200-300 amino acids and a primitive volume is used to represent the corresponding domain. The organization of the basement membrane is well known. Further, structures of its constitutive proteins are available. It is therefore simple to obtain simulations with 100 different proteins using this approach. Structural data for elastin and tropoelastin are very scarce. Weiss and co-workers (Tarakanova et al., 2018) have proposed some models and simulations using data from SAXS experiments. Using the SAXS envelope, we have compared the sizes of basement membrane type IV collagen (Fig. 1A) and tropoelastin (Fig. 1B) and the corresponding association of 4 tropoelastin monomers as proposed by Weiss and co-authors (Fig. 1C).

Up to now, the DURABIN methodology is very efficient to simulate large systems with important domains. To be able to model and simulate more complete and complex ECM

elements with this approach, it is imperative to implement additional ECM macromolecules in DURABIN catalogue/database. In this respect, the application of *in silico* approaches for a granularity corresponding to molecules characterized by repetitive motifs can be particularly promising.

As a proof of concept, we have chosen to consider some of the individual characteristic patterns of tropoelastin, which are often used as building blocks in biomaterials. To simplify the discussion, in the text below $(VPGVG)_{27}$ will be referred as VPGVG, $(VPGEG)_{27}$ and $(VPGKG)_{27}$ as VPGEG and VPGKG respectively, and $[VGLAPG(VGVAPG)_2]_{27}$ as Exon24, as it is derived from human tropoelastin exon 24-encoded sequence. In the CHLORAINE approach, one primitive corresponds to 5 or 6 amino acids (namely 5-mer and 6-mer) for single motif, up to 15 and 18 amino acids for di- and tri-motifs. The granularity is thinner than in the initial DURABIN where rigid bodies correspond to whole protein domains. Therefore, considering DURABIN and CHLORAINE method, rigid bodies introduced into the model are either protein domains or peptides of different lengths. However, even if the granularity is different, the process of defining the primitives and the corresponding rigid bodies, and the scaling up to the mesoscopic scale is the same for the domains and the motifs, making this method a multiscale approach.

2.1 Molecular dynamics of peptides and acquisition of future primitives

In a first step, we have performed molecular dynamics simulations of 1500 ns (3 x 500 ns trajectories) for each building block (VPGVG, VPGEG, VPGKG and $VGLAPG(VGVAPG)_2$). From the trajectories, we have performed clustering, using the GROMOS and TTClust methods, to obtain characteristic conformations of each block. Since the number of clusters determined at the end of the all-atom simulations is quite large, it is not possible to make an exhaustive library of conformations. Thus, to constitute our library of blocks, we only consider the major conformations, i.e., the conformations that are representative of the main clusters, corresponding to 2 or 3 conformations at most per motif. In Figure 2, we show the major conformations obtained for Exon24 and used as primitives.

A primitive is thus derived from a conformation substantially adopted by the peptide, and corresponds to a simplified description of that conformation. A primitive comprises the simplified molecular surface of the considered conformation, as well as additional physicochemical envelopes. These envelopes correspond to the representation of the

hydrophobic/hydrophilic and electrostatic properties of the motif. The primitive will influence the arrangement of the colliders, components that define the shape and the occupancy of the objects in space, which will define the shape of the rigid body and best match to its shape (Figure 3). Colliders make it possible to comply the principle of non-overlapping when assembling several rigid bodies.

Then, each primitive (molecular surface + physicochemical envelopes) with its colliders, produced with the CHLORAINE method, can be implemented and used in DURABIN. A library of 38 different building blocks from the different sequences has been proposed and is used to build some selected homopolymers. In this proof of concept, we only use one conformation of a motif to build a homopolymer. However, given the flexibility of elastin and r-ELPs, it will be of great importance to consider several conformations in a polypeptide in our future works.

2.2 Rigid bodies assembly to build polypeptides at the mesoscopic scale

By articulating several basic primitives together, polymer assemblies can be created. Joints are placed between two modules, to link them and articulate the chain in keeping with the local physical constraints (Figure 3). Two types of joints have been implemented. To perform a simulation without any angular constraint at the level of a spherical joint, Freely-Jointed Chains (FJC) joint can be defined as it allows totally free rotations around the bond. Alternatively, it is possible to use a combination of configurable joints and to define angular constraints (Phi/Psi), determined from all-atom simulations. These joints are more suitable to reproduce the degrees of freedom of the local peptide bond, and allow the implementation of structural constraints derived from all-atom molecular dynamics simulations.

2.3 Modeling and simulations of homopolymers

The choice of homopolymers to be constructed and simulated was based on four characteristic peptide motifs representative of tropoelastin.

First, we chose the VGLAPG-$(VGVAPG)_2$ motif, found in the sequence encoded by exon 24 of the human elastin gene. By tandemly repeating this motif, we created a repeating $(VGVAPG)_2$-VGLAPG-$(VGVAPG)_2$ fragment, which mimics the sequence present in domain 24 of tropoelastin.

The three other polymers are inspired by polypeptides derived from bovine tropoelastin sequence that have been widely used in biomaterials. They all correspond to poly-VPGXG sequences, that define classical ELPs. The most common is poly-VPGVG.

The constructed homopolymers are: a polymer mimicking human exon 24 encoded sequence $[VGLAPG\text{-}(VGVAPG)_2]_{27}$ (this material can be defined as a mini-elastin), a poly-VPGVG, $(VPGVG)_{27}$, a poly-VPGEG, $(VPGEG)_{27}$, and a poly-VPGKG, $(VPGKG)_{27}$. The choice of building homopolymers consisting of 27 repeats of 5 or 6 residues, say polymers of sizes varying between 135 and 486 amino acids, matches the minimal sizes of ELPs designed for biomedical approaches (MacEwan and Chilkoti, 2014)(Rodríguez-Cabello et al., 2016)(Rodriguez-Cabello et al., 2021).

For each type of polypeptide, simulations were first carried out on polypeptides where primitives were based on the molecular surface of the corresponding peptide unit. Further, mesoscopic simulations were performed to assess their behavior in terms of hydrophilic/hydrophobic and/or electrostatic interactions. These simulations were performed by adding, in the definition of the primitives, firstly, the hydrophilic/hydrophobic surfaces, then the positively/negatively charged surfaces, and, finally, all properties. In Figure 4, a copy of each homopolymer used in the simulations is given in an extended initial conformation. Then, as a proof of concept, we ran simulations with 10 copies of the considered polypeptide, to be able to study, in addition to the intramolecular interactions, the intermolecular interactions that exist between motifs of different polypeptides. For each type of simulation, the size of the box had to be adapted. Consequently, the box used for the $(VPGXG)_{27}$ simulations could not be kept for $[VGLAPG\text{-}(VGVAPG)_2]_{27}$ simulations, since the size of the molecule is 486 instead of 135 amino acids. The box for $[VGLAPG\text{-}(VGVAPG)_2]_{27}$ was therefore enlarged to maintain stable simulations.

2.4 Polypeptide behavior and interactions

Table 1 presents the number of steps reached after running the simulation for a fixed duration, i.e. 60 h, as a function of the biopolymer and added contributions.

Whatever the considered contribution, the number of steps reached by VPGXG systems are different, although they are very close in terms of size and amino acid contents. This can be explained by the fact that each type of VPGXG have different number of colliders defined for their rigid bodies. Indeed, a single VPGEG has 2 colliders, whereas a single VPGVG

has 3, and a single VPGKG has 4. This means that the VPGEG system (constituted by 27 rigid bodies) has 54 colliders, the VPGVG system has 81 colliders, and the VPGKG one has 108 colliders. In Table 1, the system that reaches the highest number of steps is VPGEG (which has the lowest number of colliders), then VPGVG and finally VPGKG (which has the highest number of colliders). More colliders implies more computational time for resolving contacts and interactions during the simulation. Actually, the number of colliders defined on the peptide surface is based on a geometrical aspect and is adapted depending on the shape/conformation of the peptide as mentioned above. A single "Exon24" has also 2 colliders (Figure 3) but reaches a much lower number of steps than VPGEG. However, it is important to note that one rigid body from the VPGXG systems contains 6 amino acids, whereas one rigid body from the Exon24 system contains 18 amino acids. Thus, the surface of potential contacts is larger for the Exon24 system than for VPGXG systems, resulting in more calculation and a lower number of steps reached. Moreover, for a same system, comparing the number of steps as a function of contributions (Table 1), we observe that they are very different. The highest figures correspond to the neutral condition (no contribution added). With no contribution, computing the interactions consists only in counting the number of contacts between the rigid bodies. The use of additional contributions implies other calculations, which limits the number of steps reached at a given time. In the same way, the simulation of the system with the 2 contributions reaches the lowest number of steps, except for the VPGKG case.

To get insights into this observation, we computed the mean number of intermolecular contacts per simulation step (Table 2).

We observed that the number of contacts per step is generally higher with the combination of the two contributions than for the neutral condition (VPGEG: 0.44 for neutral, 0.9 for 2 contributions; VPGVG: 0.63 for neutral, 2.36 for 2 contributions). This explains the higher contact computation time and, thus, the lower number of steps reached in the simulation. However, the situation is different for VPGKG. This system must be considered differently since the initial conformation used (the main one obtained from molecular dynamics simulations) is highly folded. That way, a more rigid and smaller model constitutes the rigid body seed. Thus, in this situation, the results concerning contacts computing might be biased.

Figure 4 presents the 4 r-ELPs in their elongated forms at the initial step in the simulations. In this figure, the size of the systems can be appreciated and the VPGKG peptide

is shorter. This is because the conformation issued from the molecular dynamics clustering of VPGKG is importantly folded as compared to VPGVG and VPGEG. This prevents the rotations around the joints between the rigid bodies from mimicking the real dynamics. The contacts are distorted as they do not reflect the expected interactions. As a result, our method of representing conformations is less suitable for highly folded peptides with close extremities. In this case, we will have to check several other principal conformations arising from the clustering and rather choose a conformation which will lead to more stable simulations at the mesoscopic scale. In other words, we need a conformation respecting the non-superposition principle when building the articulated chain of rigid bodies.

Furthermore, for all systems that present only one additional contribution, the number of steps is really different if we consider the hydrophobic/hydrophilic contribution or the electrostatics one. For hydrophobicity (Table 2), the number of contacts per step is higher than for electrostatics (VPGEG: 0.74 for hydrophobicity, 0.32 for electrostatics; VPGVG: 1.03 for hydrophobicity, 0.38 for electrostatics). This point explains the higher computation time resulting in the inferior number of steps reached when hydrophobic interactions are considered.

The greater importance of hydrophobic/hydrophilic contacts compared to electrostatics ones can be justified by the fact that, in a general way, hydrophobic/hydrophilic surfaces are much larger than electrostatics ones. Thus hydrophobic/hydrophilic interactions are more likely to occur compared to electrostatics ones. We can also notice that the number of steps reached during the neutral simulation is higher than the one reached by the electrostatic simulation, even if the number of neutral contacts is higher than the number of electrostatic contacts. This is because, as said above, the computing time required is more important for the electrostatic contacts compared to the neutral one. To summarize, if we consider one or two contributions, the contact computation time is longer, and thus the simulation reaches less steps. Moreover, simulations with hydrophobicity perform less steps than with electrostatics because there are more hydrophobic interactions occurring than electrostatics ones. Our experiments show that the simulation progression is totally dependent on the choice of the used contributions.

In order to assess the performance of the CHLORAINE procedure presented in this work, we have simulated an arbitrary number of 10 chains of each polymer. As discussed above, the size of the box had to be adjusted to maintain stable simulations. The longer the

homopolymers simulated, the highest the concentration of molecules in the box for a fixed number of copies. Nevertheless, even in this case, simulations provide nice information about the interaction of the copies and the "local crowding" that may obtained. In Figure 5, the evolution of 10 chains of $[(VGLAPG)(VGVAPG)_2]_{27}$ during the first 6,000 steps of a CHLORAINE procedure using both electrostatic and MHP potentials is presented. In the initial step, the 10 chains are in an extended state in a random position. After few steps of simulation, we observed for all the simulations that the copies bend and interact in a dynamic way. Along the simulations, the different rigid bodies of a copy may interact both in an intramolecular and intermolecular ways. From these simulations, it is possible to derive matrices of interactions and to follow the interactions of the simulated systems. In Figure 6, the matrices of interactions of the polypeptide $(VPGEG)_{27}$ after 38,000 steps (A) and 128,000 steps (B) of simulation are provided. In this figure, each square (27x27) corresponds to one copy of the polymer (of 27 rigid bodies) and thus each axis represents the 10 copies of $(VPGEG)_{27}$ put in the simulation box. An interaction matrix is obtained at each step, with the cumulative number of interactions that occurred since the initial step. Thus, it is possible to follow the evolution of the interactions along the simulation. This allows to distinguish how the simulation has progressed during a period of time and which building blocks of the copies have been involved in the interactions. Here, we chose to represent the matrices of intermolecular interactions at 38,000 steps (6A) and 128,000 steps (6B) of simulation. In the same way, in Figure 7, the matrices of interactions at the same step of $(VPGVG)_{27}$ are provided with dihedral contribution (7A), plus electrostatic contribution (7B), plus hydrophobic/hydrophilic contribution (7C) and all the contributions (7D). Due to the number of colliders used in the simulations, we observe that when the electrostatic contribution is added, we drastically reduce the number of interactions. This is not the case with hydrophobic/hydrophilic parts. When performing the MHP calculations for the isolated peptide, we know that the computed MHP surfaces are much larger than the electrostatic ones. The corresponding primitives defined have thus larger surfaces and, as a consequence, the hydrophilic/hydrophobic contributions are more important in the calculations as seen in Figure 7D, where all the contributions are added.

Thanks to the DURABIN mesoscope, it is possible to confidently simulate elastic repeated peptides associated into a "small" (compared to experiments) homopolypeptide. The CHLORAINE procedure allows us to modulate the size of the studied rigid bodies and even

if the granularity is different, the process of primitives definition and the corresponding rigid bodies. The scaling up to the mesoscopic scale is however the same for domains and motifs, which makes CHLORAINE a multiscale approach.

In the near future, we will focus on controlling the size of the simulation boxes, as this parameter impacts the concentration of r-ELPs. This point is essential to provide reliable mesoscopic simulations that could then be directly compared to experimental data obtained for comparable biomaterials.

Conclusions - Perspectives

In this work, we provide a proof of concept that allows the numerical simulation of r-ELPs in a simple way. r-ELPs are very important for the design of new elastomeric biomaterials (Ibáñez-Fonseca et al., 2019). Up to now, the simulation of elastic peptides associated to biomaterials remains very challenging due to the specificity of r-ELPs sequences, lengths and/or properties. Very recently, some attempts have been made to compute them in all-atom or coarse-grained systems (Li et al., 2021)(Baul et al., 2020) but these simulations remain limited due to the size of the molecular systems. We have adapted our mesoscopic rigid body modeling system (DURABIN (Wong et al., 2018)) to study molecular objects ranging from very large multi-domain proteins of the ECM to potential biomaterials made of tandemly repeated peptides. Here, our Challenging Level Of Rigid-body Approach Involving Numerical Elements (CHLORAINE) method is applied to some chosen r-ELP patterns to evaluate their dynamical behavior, the associated interactions and the influence of the complexity of the simulations.

At the same time, it must be emphasized that the high level of the heterogeneity and complexity of biomaterial simulations represent a possible obstacle to drawing more detailed and particularly quantitative conclusions. The multiscale approach has been implemented with the passage from some conformations of small fragments or peptide motifs of tropoelastin to polypeptides (corresponding to chains of assembled fragments); in other words, from nanoscale to mesoscale levels. Simplified surfaces called primitives were defined from those conformations. These primitives determine the setting up of the colliders, which define the occupancy of the motifs in the space. In this way, the shape of the rigid body corresponding to each represented conformation is established. Besides, hydrophilic/hydrophobic and positive/negative electrostatic contributions can be considered in primitives by adding their corresponding surfaces on the molecular surface of a

conformation. We can thus get the hydrophilic/hydrophobic and electrostatic behavior of r-ELPs, in particular through the study of intra- and interchain interactions.

In this paper, we consider homopolymers of main conformations of four tropoelastin units, but it would be interesting to integrate primitives of other conformations for the same motifs. We have tested our approach on three homopolymers (138 AA in length) composed of pentameric peptides (same motif VPGXG with X as V, E or K) and one homopolymer with 6-mer patterns (486 AA in length). The flexible and fractal nature of tropoelastin implies a high number of conformations for a given sequence motif (conformational entropy). Thus, it would be very interesting to perform simulations of heteropolymers, which would be constructed by assembling different motifs (sequences) and/or different conformations: the consequence would be a dramatical increase of the combinatorics. On top of that, we have seen in this paper the issue of highly folded conformations, which need to be treated differently from the others. Indeed, we should improve the way to model these conformations (for instance defined by a certain threshold of radius of gyration), to avoid potentially biased interactions during the simulation.

Furthermore, in the near future, mesoscopic simulations will have to take into account the sampling problem and therefore the volume of the simulation boxes. In fact, the size of the simulation box should be adaptable according to the length of the polypeptide studied, thus adapting the concentration of polypeptides. It will therefore be important to optimize this parameter for future simulations and to be able to control the volume, concentration or temperature. Indeed, in DURABIN, it is possible to change the temperature to increase thermal agitation thereby modulating the frictions between macromolecules.

Concerning the simulation progression, it will be interesting to study the optimization of the interactions computing. Indeed, by reducing the necessary computational time, our simulation will be able to achieve a more significant number of steps for a given time. It will be necessary to have a calibration between the timestep used in DURABIN and a time duration. Moreover, we have seen that the number of colliders describing the primitive has an influence on computation time. Thus, it is necessary to find a compromise between the number of colliders and the precision of the surface description.

As a conclusion, the CHLORAINE approach allows to apply the rigid body approach to elastin polypeptides and to integrate them in our simulator, DURABIN (Wong et al., 2018).

Thus, we perform simulations of large polypeptides derived from elastin at the mesoscopic scale. This achievement is an essential step towards the characterization of biomaterial systems. As shown in our experiments, highly folded conformations with very short end-to-end distances may bring some difficulties for our simulator and lead to tricky simulations. Indeed, in this case, the assembly of primitives would not respect the non-superposition principle which is inherent to rigid bodies. Instead, we rather consider another major conformational state arising from clustering to get a more stable mesoscopic simulation.

Finally, our objective is to perform simulations of more complex and larger fragments of elastin, and even a full tropoelastin. These precious data would shed new light on the behavior of elastin in the ageing process or as a biomaterial. Indeed, we would then be able to correlate experimental data with parameters of our simulations, which would undoubtedly help us to calibrate and improve our methodological approach. Furthermore, other ECM molecules could be modeled by the CHLORAINE procedure and simulated in DURABIN, allowing us to get a more realistic and comprehensive virtual model of the ECM.

CRediT authorship contribution statement

CD performed molecular dynamics and mesoscopic simulations, analyses, conception of the library of patterns, wrote the paper, HW and JMC performed some molecular dynamics and mesoscopic simulations, LD and SB offered their skills in the field of elastin peptides and simulations, MD and NB wrote the paper and conceptualized the research, provided guidance and supervision. All authors participated in the data interpretation and manuscript preparation, reviewing and editing.

Declaration of Competing Interest

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

Data availability

Data will be available on request.

Acknowledgments

This work was supported by the project DURABIN (Region Champagne-Ardenne) and the chair MAgICS (Grand Reims support). Part of this work was performed on computing resources provided by CRIANN (Normandy, France) and Centre de Calcul Régional ROMEO (Université de Reims Champagne Ardenne, France). It was granted by the Region Grand Est and the Université de Reims Champagne Ardenne.

References

Baul, U., Bley, M., Dzubiella, J., 2020. Thermal Compaction of Disordered and Elastin-like Polypeptides: A Temperature-Dependent, Sequence-Specific Coarse-Grained Simulation Model. Biomacromolecules 21, 3523–3538. https://doi.org/10.1021/acs.biomac.0c00546

Brasseur, R., 1991. Differentiation of lipid-associating helices by use of three-dimensional molecular hydrophobicity potential calculations. J Biol Chem 266, 16120–16127.

Cao, Q., Wang, Y., Bayley, H., 1997. Sequence of abductin, the molluscan 'rubber' protein. Current Biology 7, R677–R678. https://doi.org/10.1016/S0960-9822(06)00353-8

Daura, X., Gademann, K., Jaun, B., Seebach, D., van Gunsteren, W.F., Mark, A.E., 1999. Peptide Folding: When Simulation Meets Experiment. Angewandte Chemie International Edition 38, 236–240.

Fauchère, J.-L., Quarendon, P., Kaetterer, L., 1988. Estimating and representing hydrophobicity potential. Journal of Molecular Graphics 6, 203–206. https://doi.org/10.1016/S0263-7855(98)80004-0

Furet, P., Sele, A., Cohen, N.C., 1988. 3D molecular lipophilicity potential profiles: a new tool in molecular modeling. Journal of Molecular Graphics 6, 182–189. https://doi.org/10.1016/S0263-7855(98)80001-5

Gosline, J., Lillie, M., Carrington, E., Guerette, P., Ortlepp, C., Savage, K., 2002. Elastic proteins: biological roles and mechanical properties. Phil. Trans. R. Soc. Lond. B 357, 121–132. https://doi.org/10.1098/rstb.2001.1022

Holst, M.J., Saied, F., 1995. Numerical solution of the nonlinear Poisson-Boltzmann equation: Developing more robust and efficient methods. J. Comput. Chem. 16, 337–364. https://doi.org/10.1002/jcc.540160308

Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., de Groot, B.L., Grubmüller, H., MacKerell, A.D., 2017. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nat Methods 14, 71–73. https://doi.org/10.1038/nmeth.4067

Humphrey, W., Dalke, A., Schulten, K., 1996. VMD: Visual molecular dynamics. Journal of Molecular Graphics 14, 33–38. https://doi.org/10.1016/0263-7855(96)00018-5

Ibáñez-Fonseca, A., Flora, T., Acosta, S., Rodríguez-Cabello, J.C., 2019. Trends in the design and use of elastin-like recombinamers as biomaterials. Matrix Biology 84, 111–126. https://doi.org/10.1016/j.matbio.2019.07.003

Jorda, J., Xue, B., Uversky, V.N., Kajava, A.V., 2010. Protein tandem repeats - the more perfect, the less structured: Structural state of perfect protein repeats. FEBS Journal 277, 2673–2682. https://doi.org/10.1111/j.1742-4658.2010.07684.x

Le Page, A., Khalil, A., Vermette, P., Frost, E.H., Larbi, A., Witkowski, J.M., Fulop, T., 2019. The role of elastin-derived peptides in human physiology and diseases. Matrix Biology 84, 81–96. https://doi.org/10.1016/j.matbio.2019.07.004

Li, N.K., Xie, Y., Yingling, Y.G., 2021. Insights into Structure and Aggregation Behavior of Elastin-like Polypeptide Coacervates: All-Atom Molecular Dynamics Simulations. J. Phys. Chem. B 125, 8627–8635. https://doi.org/10.1021/acs.jpcb.1c02822

Lins, L., Thomas, A., Brasseur, R., 2003. Analysis of accessible surface of residues in proteins. Protein Sci. 12, 1406–1417. https://doi.org/10.1110/ps.0304803

Lyons, R.E., Lesieur, E., Kim, M., Wong, D.C.C., Huson, M.G., Nairn, K.M., Brownlee, A.G., Pearson, R.D., Elvin, C.M., 2007. Design and facile production of recombinant resilin-like polypeptides: gene construction and a rapid protein purification method. Protein Engineering, Design and Selection 20, 25–32. https://doi.org/10.1093/protein/gzl050

MacEwan, S.R., Chilkoti, A., 2014. Applications of elastin-like polypeptides in drug delivery. Journal of Controlled Release 190, 314–330. https://doi.org/10.1016/j.jconrel.2014.06.028

Muiznieks, L.D., Keeley, F.W., 2017. Biomechanical Design of Elastic Protein Biomaterials: A Balance of Protein Structure and Conformational Disorder. ACS Biomater. Sci. Eng. 3,

661–679. https://doi.org/10.1021/acsbiomaterials.6b00469
Rodríguez-Cabello, J.C., Arias, F.J., Rodrigo, M.A., Girotti, A., 2016. Elastin-like polypeptides in drug delivery. Advanced Drug Delivery Reviews 97, 85–100. https://doi.org/10.1016/j.addr.2015.12.007
Rodriguez-Cabello, J.C., Gonzalez De Torre, I., González-Pérez, M., González-Pérez, F., Montequi, I., 2021. Fibrous Scaffolds From Elastin-Based Materials. Front. Bioeng. Biotechnol. 9, 652384. https://doi.org/10.3389/fbioe.2021.652384
Seo, M., Rauscher, S., Pomès, R., Tieleman, D.P., 2012. Improving Internal Peptide Dynamics in the Coarse-Grained MARTINI Model: Toward Large-Scale Simulations of Amyloid- and Elastin-like Peptides. J. Chem. Theory Comput. 8, 1774–1785. https://doi.org/10.1021/ct200876v
Tarakanova, A., Yeo, G.C., Baldock, C., Weiss, A.S., Buehler, M.J., 2018. Molecular model of human tropoelastin and implications of associated mutations. Proc Natl Acad Sci USA 115, 7338–7343. https://doi.org/10.1073/pnas.1801205115
Tubiana, T., Carvaillo, J.-C., Boulard, Y., Bressanelli, S., 2018. TTClust: A Versatile Molecular Simulation Trajectory Clustering Program with Graphical Summaries. J. Chem. Inf. Model. 58, 2178–2182. https://doi.org/10.1021/acs.jcim.8b00512
Wong, H., Crowet, J.-M., Dauchez, M., Ricard-Blum, S., Baud, S., Belloy, N., 2022. Multiscale modelling of the extracellular matrix. Matrix Biology Plus 13, 100096. https://doi.org/10.1016/j.mbplus.2021.100096
Wong, H., Prévoteau-Jonquet, J., Baud, S., Dauchez, M., Belloy, N., 2018. Mesoscopic Rigid Body Modelling of the Extracellular Matrix Self-Assembly. Journal of Integrative Bioinformatics 15. https://doi.org/10.1515/jib-2018-0009
Zhou, C.-Z., 2000. Fine organization of Bombyx mori fibroin heavy chain gene. Nucleic Acids Research 28, 2413–2419. https://doi.org/10.1093/nar/28.12.2413

Figure captions

Figure 1: Representation and size of a collagen type IV (A), of a monomer of tropoelastin (B), of the assembly of 4 monomers of tropoelastin (C) and of a $[VGLAPG(VGVAPG)_2]_{27}$ r-ELP (D). Scale refers to the four structures.

Figure 2: Molecular surface of one representative conformation for the 4 main clusters of r-ELP $[VGLAPG(VGVAPG)_2]$ represented with Blender and used as primitives. The probability of occurrence of each conformation is provided.

Figure 3: Definition of colliders and joints for an elastic pattern.

Figure 4: View of the 4 r-ELP systems modeled in the mesoscope in an extended form at the beginning of each simulation.

Figure 5: Evolution of 10 chains of $[(VGLAPG)(VGVAPG)_2]_{27}$ during the first 6,000 steps of a CHLORAINE procedure using both electrostatic and MHP potentials.

Figure 6: Matrices of interactions of the polypeptide $(VPGEG)_{27}$ after 38,000 steps (A) and 128,000 steps (B).

Figure 7: $(VPGVG)_{27}$ interactions without contribution (A), with electrostatics (B), with MHP (C) and with both electrostatics and MHP (D).

Table 1: Number of steps reached by the simulation (60 h) for the 4 systems in 4 different conditions: neutral (no added contribution), MHP (hydrophobicity/hydrophilicity contribution), electrostatics (electrostatic contribution) and both (MHP + electrostatic) contributions.

Table 2: Mean number of intermolecular contacts per simulation step, for the 4 systems in the 4 conditions (same conditions as the previous Table 1).

List of Figures

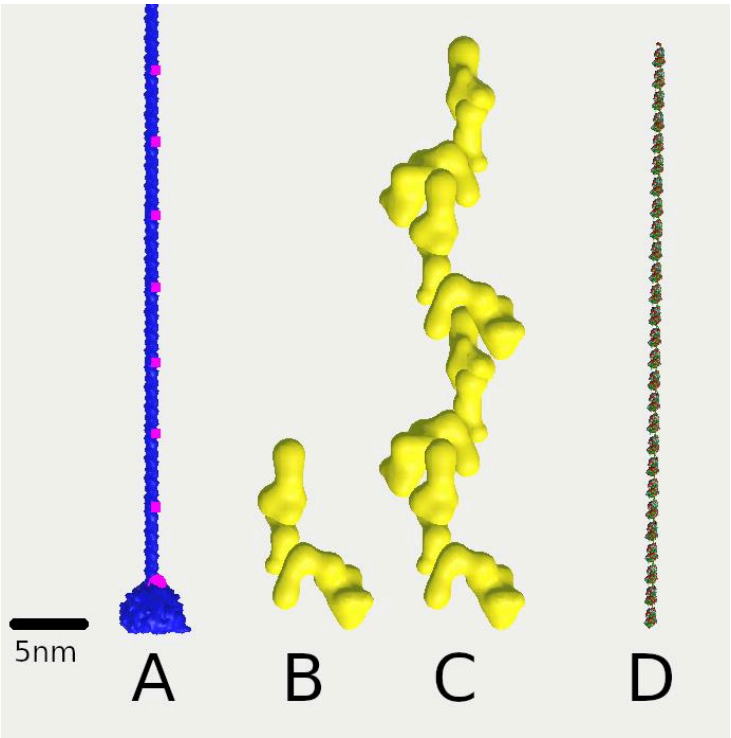


Figure 1: Representation and size of a collagen type IV (A), of a monomer of tropoelastin (B), of the assembly of 4 monomers of tropoelastin (C) and of a $[VGLAPG(VGVAPG)_2]_{27}$ r-ELP (D). Scale refers to the four structures.

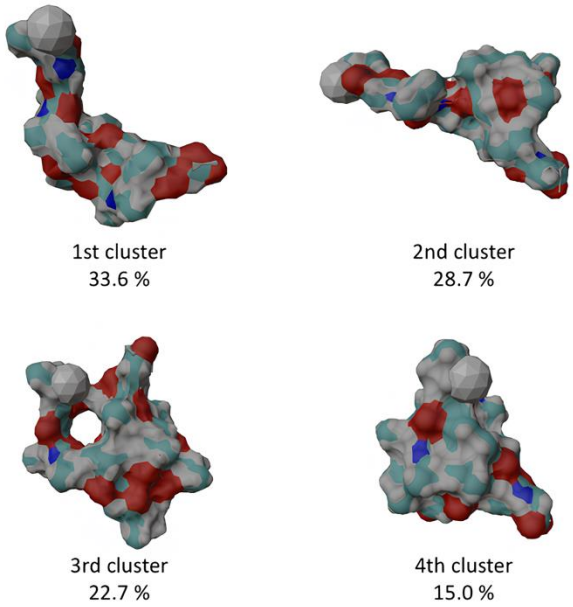


Figure 2: Molecular surface of one representative conformation for the 4 main clusters of r-ELP [VGLAPG(VGVAPG)$_2$] represented with Blender and used as primitives. The probability of occurrence of each conformation is provided.

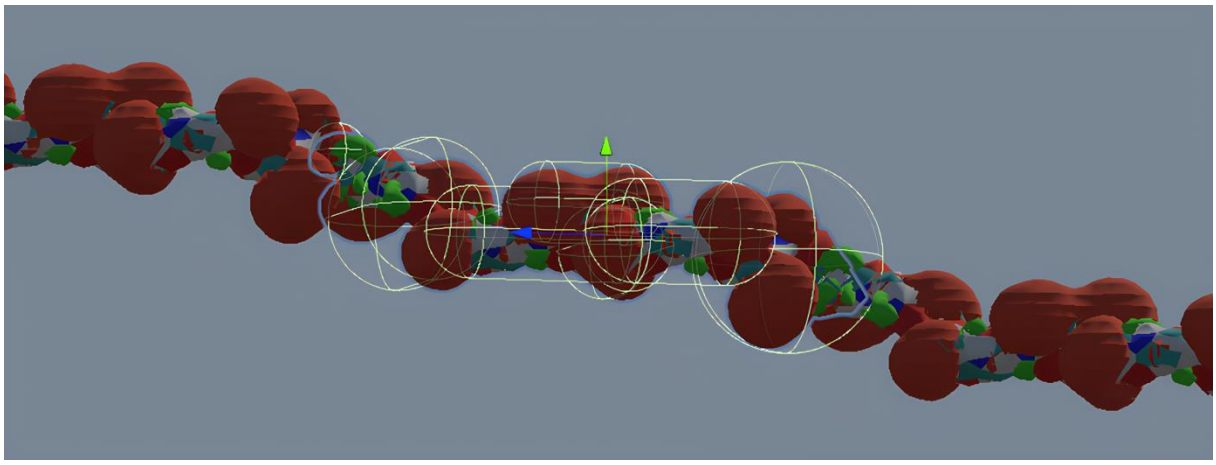


Figure 3: Definition of colliders and joints for an elastic pattern.

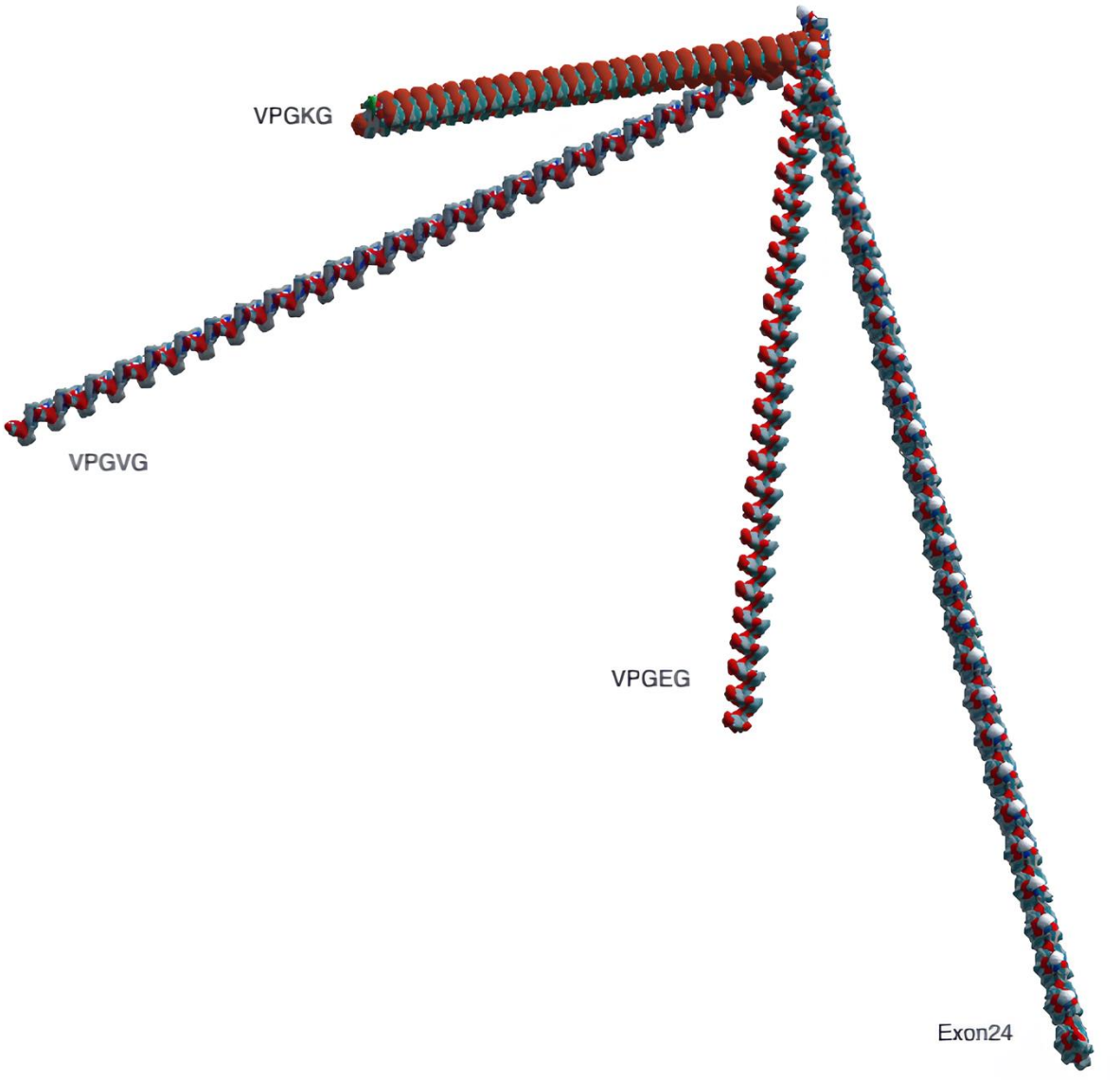


Figure 4: View of the 4 r-ELP systems modeled in the mesoscope in an extended form at the beginning of each simulation.

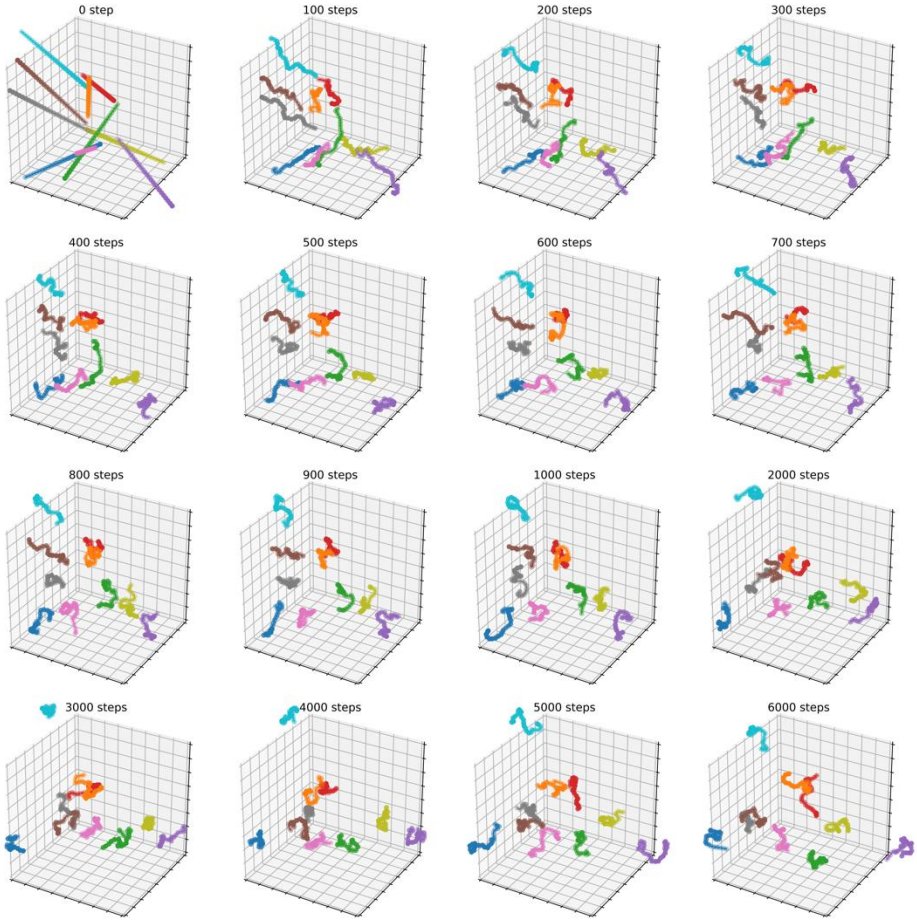


Figure 5: Evolution of 10 chains of $[(VGLAPG)(VGVAPG)_2]_{27}$ during the first 6,000 steps of a CHLORAINE procedure using both electrostatic and MHP potentials.

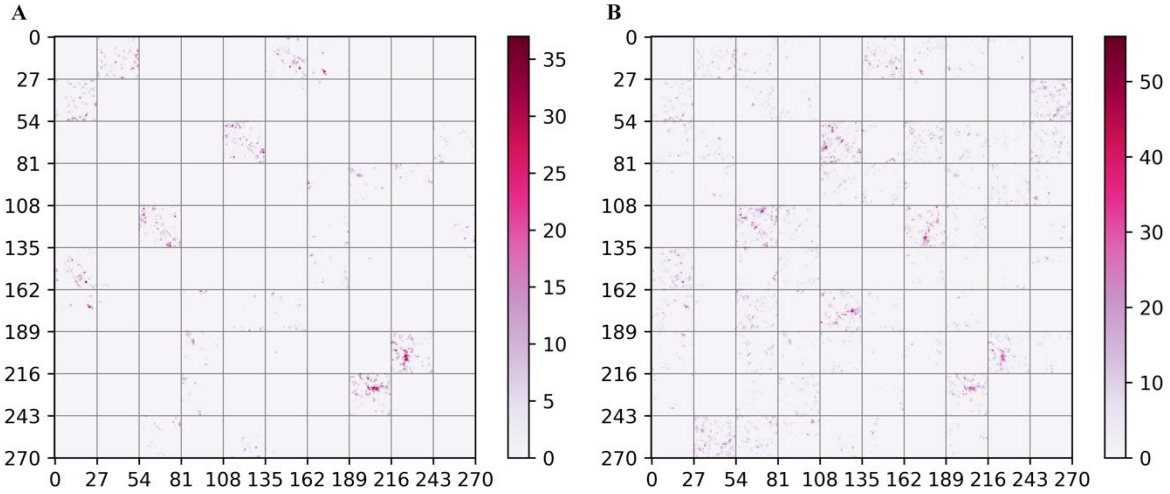


Figure 6: Matrices of interactions of the polypeptide $(VPGEG)_{27}$ after 38,000 steps (A) and 128,000 steps (B).

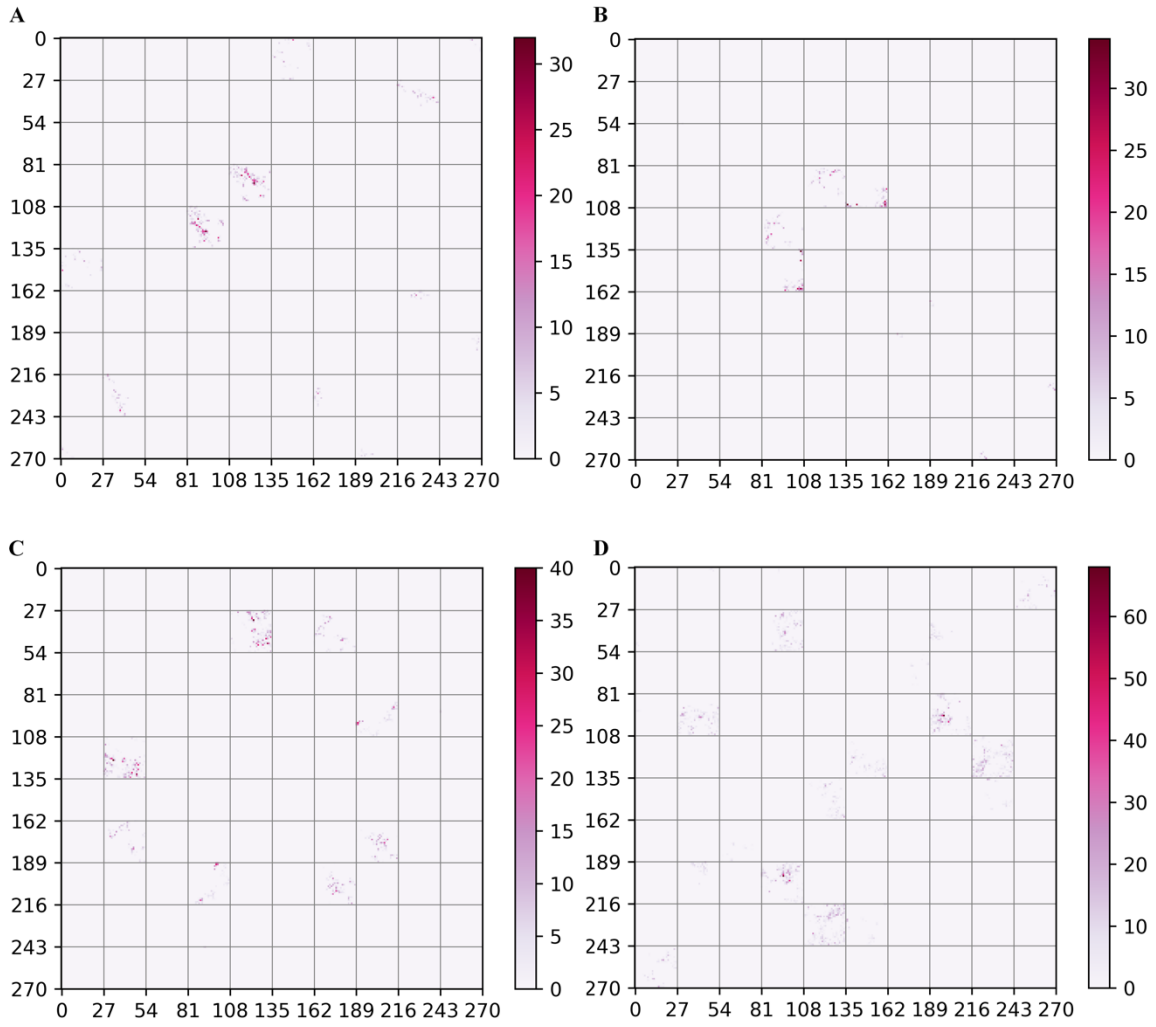


Figure 7: $(VPGVG)_{27}$ interactions without contribution (A), with electrostatics (B), with MHP (C) and with both electrostatics and MHP (D).

List of Tables

	Neutral	MHP	Electrostatics	Both
VPGVG	46,680	4,610	30,460	4,330
VPGEG	128,110	25,550	52,490	11,570
VPGKG	38,220	8,650	31,330	11,230
« Exon24 »	46,260	4,810	33,010	4,410

Table 1: Number of steps reached by the simulation (60 h) for the 4 systems in 4 different conditions: neutral (no added contribution), MHP (hydrophobicity/hydrophilicity contribution), electrostatics (electrostatic contribution) and both (MHP + electrostatic) contributions.

	Neutral	MHP	Electrostatics	Both
VPGVG	0.63	1.03	0.38	2.36
VPGEG	0.44	0.74	0.32	0.99
VPGKG	2.38	0.25	1.40	2.80
« Exon24 »	0.52	1.89	0.34	0.17

Table 2: Mean number of intermolecular contacts per simulation step, for the 4 systems in the 4 conditions (same conditions as the previous Table 1).

CRediT authorship contribution statement

CD performed molecular dynamics and mesoscopic simulations, analyses, conception of the library of patterns, wrote the paper, HW and JMC performed some molecular dynamics and mesoscopic simulations, LD and SB offered their skills in the field of elastin peptides and simulations, MD and NB wrote the paper and conceptualized the research, provided guidance and supervision. All authors participated in the data interpretation and manuscript preparation, reviewing and editing.