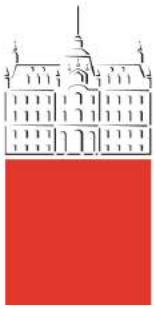
University *of Ljubljana*
Faculty *of Mathematics and Physics*



Master program in
Seminar 2

Essential Dynamics Coarse-Graining

Author: Simon Perovnik
Advisor: prof. dr. Miha Ravnik

Ljubljana, March 6 2023

Abstract

Essential-dynamics coarse graining is a method based on elastic network model used for determining a simplified structure dynamics of a given macromolecule. Such techniques are especially valuable when an all-atom molecular dynamics simulations are not feasible, such as due to large timescales needed to properly study certain biological processes. Main advantage of the method is that it preserves dynamical domains of a molecule, which are often important when studying biological processes. In this seminar we start by introducing the topic of coarse-graining proteins, describe the essential-dynamics method and give an examples of its applicability for modeling monoclonal antibodies, a type of protein. We conclude with a brief discussion about the advantages and disadvantages of the method.

Contents

1 Introduction

Biological processes on subcelluar level are governed by how biomolecules interact with each other. This area of research was first studied with experimental techniques, but with the advancement of computational methods, more and more processes are studied in silico [1]. In case of large systems (such as proteins [2], carbohydrates [3], DNA [4]) strong insight can be gained across diverse time and lenght-scales by using coarse-grained (CG) models of biomolecules which provide a simplified representation of a given macromolecule. CG models are especially useful as they can also be coupled together with each other or with all-atom representations of molecules, forming a multiscale CG model [5]. This broadens the space for different applications significantly as it allows for different levels of molecular resolution being used in the same simulations, so that each type of molecules can be coarse-grained on its own (or preserved as all-atom representation).

One of the first CG models that was developed, was a protein model proposed by Levitt and Warshel [6]. Motivation behind the model was to be able to better understand the process of protein folding [7]. Since then many CG models of different coarseness for proteins were proposed [8–11]. In the recent years protein modeling for monoclonal antibodies (mAbs) has developed [12–15]. As motivated by the need to develop stable mAb molecules that can be used for treating different diseases [16]. One such CG model called essential-dynamics is described in this seminar.

In this seminar, the emphasis is on CG methods used for protein modeling. For that reason we start with a brief review of proteins in chapter 2, which is followed by an overview of CG modeling of proteins in chapter 3. In chapters 4 and 5 we describe a selected model of proteins called essential-dynamics and provide an example of its applicability for mAb molecule. Seminar is concluded with a reflection on advantages and disadvantages of the ED method.

2 Coarse-Grained Models of Proteins

Proteins are linear macromolecules that consist of couple tens and up to hundreds or even thousands amino acids (AAs) which are connected together in a sequence with peptide bonds [19]. This is shown in Fig. 2, where each AA is enclosed with a grey dashed line. Amino acids can be further split into four parts - α-carbon atom (paired with a hydrogen atom), carboxyl group, amino group and a side-chain, which is different for every AA [20]. Even though proteins are linear molecules (they are determined by AA sequence) their structure is much more complex, because bonds between non-adjacent AAs form and protein folds - see figure 1 . All α-carbon atoms (colored with light blue color in figure 1) together constitute the backbone of protein which defines the general shape of a protein.

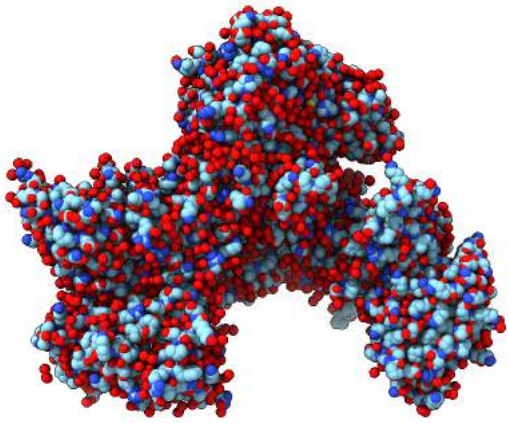


Figure 1: Visualisation of a protein called actin related protein 2/3 complex. Carbon atoms are colored with light blue color, oxygen atoms with red, nitrogen atoms with dark blue and sulfur atoms with yellow. Protein structure is from [17] and visualised with [18].

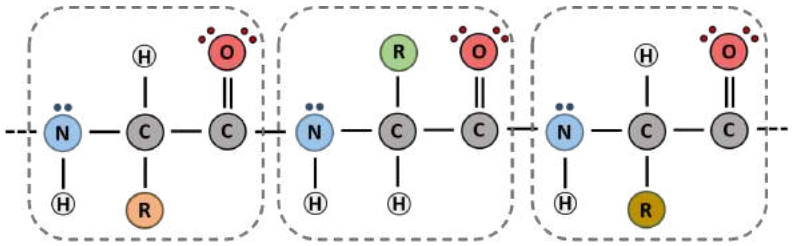


Figure 2: Three amino acids linked together with peptide bonds (between CO and NH). Each consists of amino (NH) and carboxyl (CO) group, side-chain denoted with R and a central α-carbon atom paired with a hydrogen atom. Every side-chain is colored with different color to mirror the fact, that those differ for different AAs. Adapted from [20].

Proteins can be coarse-grained in a variety of ways [21]. The main idea behind all CG models is that the representation of a protein can at times be significantly simplified, as there is no need for a detailed all-atom description of a molecule. Some atoms can be either omitted from simulation all-together or combined to form a pseudo-atom of sorts and CG model would still accurately reflect some physical properties. An example of this would be merging an entire amino acid into a single CG site [14] or to combine the amino acid side-chain into one CG site, and all other atoms into another [6]. Both cases are portrayed in Fig. 3 where multiple amino acids are connected together to form a protein.

With such simplifications in structure we can reduce the degrees of freedom substantially, which makes a system easier to simulate as it requires less computational power, but comes at the expense of losing some information about molecule's properties. Since mechanisms instigating biological processes take place at different scales - some being driven by properties such as net charge of whole macromolecule, while others depend on highly local properties like polarity of binding sites - different degrees of coarseness are appropriate for describing different phenomena [22]. Consequentially various CG models have been proposed throughout the years, each being applicable in certain cases.

Though coarseness is perhaps the most general distinctive property of CG models, they differ also in other aspects. To get a perspective of how broad is the "phase space" of CG modeling, we list some questions which each model can address differently:

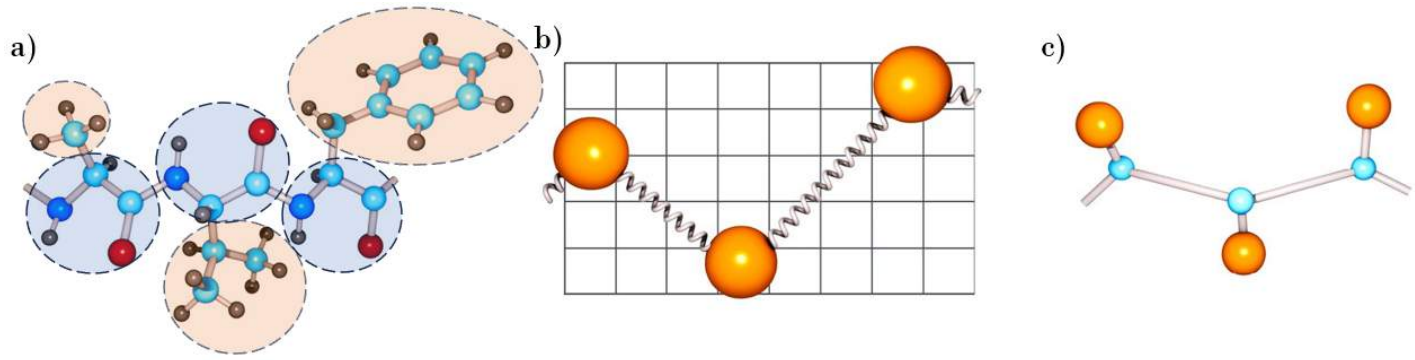


Figure 3: Three representations of different resolutions of a protein: (a) An all-atom representation, each AA begin schematically divided into the side-chain (orange) and the rest of the molecule (blue). (b), (c) CG representation of amino acids that form a protein, where one CG site corresponds to one amino acid (b) and where one CG site describes the side chain and the other one rest of the molecule (c). Adapted from [2].

- How coarse should a model be?
- How to map an all-atom representation to a coarse-grained one?
  - Where to place CG sites and how to determine their sizes?
  - How to properly mirror electric and hydrophobic properties of atoms into a single CG site?
  - Can some particles be omitted from simulation all-together?
- How to formulate interactions between pseudo-atoms (intramolecular potentials) and between macromolecules (intermolecular potentials) to best describe the physical phenomena being modeled?

This makes it apparent that constructing a CG model is far from trivial and requires significant amount of fitting models to experimental data to ensure that they reflect physical reality of the system [14]. Similar also goes for the process of selecting an appropriate model, where it is especially important to be aware in what regime is a certain model valid.

Another important area of CG modeling is defining force fields. Force fields are computational models which include intermolecular and intramolecular potentials that together form a complete force field between atoms and molecules in the simulation. They are analogous to classical force fields but describe interactions on the atomistic scale. They are partially directed by the coarseness of the model, since in a very coarse model for example, some potentials would only offer corrections in the second order so the increase in computational burden would out-weigh the merit of a better approximation [2].

Force fields are typically of two types, i.e. bonded and non-bonded. Former including interactions that bind particles together and model molecular, covalent and other bonds, whilst latter describe repulsive, steric (often described with Lennard-Jones potential) and electrostatic effects. A general force field is given in Eq. (1), where the combined potential U_{total} is split into bonded U_{bonded} and non-bonded $U_{\text{non-bonded}}$ terms, which is described with Lennard-Jones $U_{\text{Lennard-Jones}}$ and electrostatic $U_{\text{electrostatic}}$ term.

$$\begin{aligned} U_{\text{total}} &= U_{\text{bonded}} + U_{\text{non-bonded}} \\ U_{\text{non-bonded}} &= U_{\text{electrostatic}} + U_{\text{Lennard-Jones}} \end{aligned} \tag{1}$$

3 Elastic Network Model - Essential Dynamics

Elastic network model - essential dynamics (ENM-ED) coarse graining method, developed by Voth group [23], provides a way of simplifying molecular structure from hundreds or even thousand atoms into a model constituting of only a few or up to twenty or thirty pseudo-atoms (i.e. CG sites). The method is particularly convenient as it builds the whole model based on the dynamic nature of the protein from a single structure of a molecule and does not require trajectory data from MD simulations, as is the case in other CG models [24] that take into account dynamic nature of a molecule. Since ENM-ED is based upon a method called elastic network model (ENM) which is actually already a CG model, it could be perceived as an extensions of ENM CG model.

3.1 Elastic Network Model

The core idea behind ENM is to approximate atoms with point-like particles and bonds connecting them with Hookean springs [25]. It starts with a reference structure of a protein and only takes into account α-carbon atoms (the backbone of a protein). We then define a distance cut-off and a spring constant. We proceed by checking the distances between all pairs of atoms. If they are shorter then the pre-determined cut-off (typically around 10 Å) we assign that pair a spring in between them, which has an equilibrium length r_{ij}^0 equal to the distance between the atoms from the reference structure. This way we mimic the effect, that atoms which are closer together likely bond together with chemical bonds. A representation of ENM is shown in Fig. 4, where in the first and second tiles cut-off distance is set to 8Å and 15Å, respectively, while the third tile shows all α-carbon atoms of a protein. An extension of this model would be to define several cut-offs and spring constants to account for the different strength's of intramolecular bonds as nearby atoms bind together stronger.

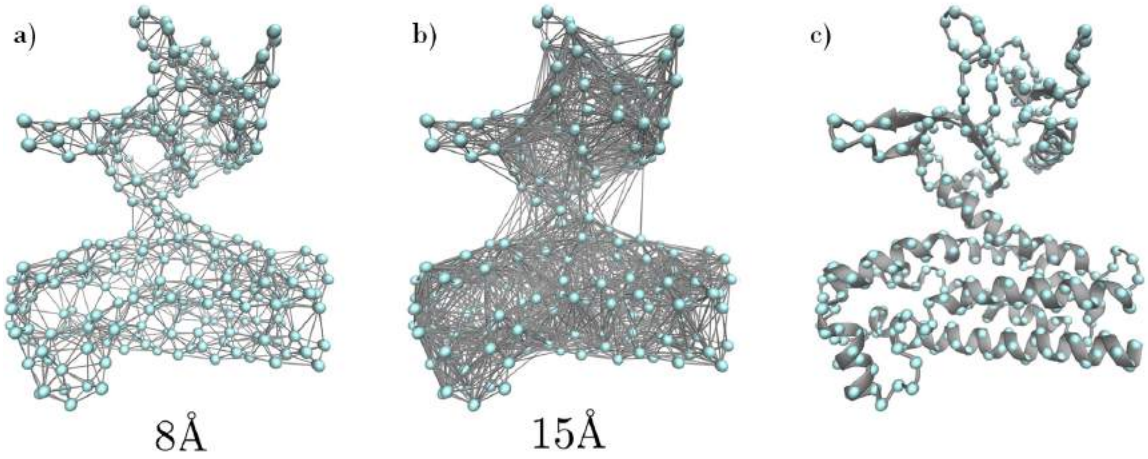


Figure 4: Representation of a protein with beads at locations of α-carbon atoms (c) and two visualisations of ENM with cut-off set to 0.8 nm (a) and 1.5 nm (b). Figure is from [26].

Since we defined interactions between the atoms to mimic linear springs, full potential can be written as in Eq. (2).

$$V = \sum_{i,j>i} k_{ij}(r_{ij} - r_{ij}^0)^2 \quad \text{and} \quad k_{ij} = \begin{cases} c & \text{if } r_{ij} < r_{\text{cs}} \\ c \cdot 10^{-2} & \text{if } r_{\text{cs}} < r_{ij} < r_{\text{cm}} \\ c \cdot 10^{-4} & \text{if } r_{\text{cm}} < r_{ij} < r_{\text{cl}} \\ 0 & \text{if } r_{\text{cl}} < r_{ij} \end{cases} \tag{2}$$

The sum in Eq. (2) goes over all pairs of α-carbon atoms, k_{ij} corresponds to the spring constant between the

atom i and j and r_{ij}^0 is the equilibrium length of the spring determined by the reference structure, so that $r_{i,j} - r_{i,j}^0$ is the stretched or compressed length of a spring between atoms i and j. The spring constant k_{ij} is determined according to the distance between atoms r_{ij} and cut-offs r_{cs}, r_{cm} along with r_{cl}, corresponding to short, middle and long cut-off values, respectively. To keep the model simple, all spring constants depend on a single parameter c. Scalar factors 10^{-2} and 10^{-4} are set approximately, to mirror different magnitudes of covalent and Van der Walls bonds.

This definition indeed assumes that the reference structure corresponds to a minimum energy state, where all springs are neither compressed, nor stretched.

ENM can be analysed in two different ways, as a Gaussian network model and an Anisotropic network model. Former assuming that the fluctuations of each atom are isotropic, so each atom is described only by a magnitude of fluctuations, while the latter is an extension of this model, where the directionality is taken into account which results in a conformational space of $3N$ dimensions, where N is the number of α-carbon atoms. Both models are used to estimate fluctuations of each node in ENM that are obtained through normal mode analysis which is described in the next section.

3.2 Normal Mode Analysis

Normal mode analysis (NMA) is an analytical tool for determining normal modes of a macromolecule such as proteins [25]. Although it identifies a whole spectrum of normal modes, the modes with low frequencies are particularly useful as they correspond to collective motion of a macromolecule [27]. Same modes are also more reliable, because they aren't as influenced as much by structural variability which could be a consequence of a referenced structure that doesn't mirror perfectly natural structure of a macromolecule.

Motion in NMA is approximated to be linear near the equilibrium state. Decomposition to normal modes, which are orthogonal to one another is carried out as a diagonalisation of Hessian matrix, which is defined in Eq. (3). It consists of $N \times N$ terms, one for each pair of α-carbon atoms. The super-elements h_{ij} are 3×3 matrices including second derivatives of potential V along the coordinates r_{ix}, where $\mathbf{r}_i$ is the location of i-th α-carbon atom and subscript x denotes a spatial coordinate (e.g. $x = 1, 2, 3$ represent the first, second and third spatial dimension, respectively). All derivatives can be calculated according to Eq. (4), where the distance between i-th and j-th atom is evaluated at the equilibrium length r_{ij}^0.

$$\mathsf{H} = \begin{bmatrix} \mathsf{h}_{11} & \mathsf{h}_{12} & \dots & \mathsf{h}_{1N} \\ \mathsf{h}_{21} & \mathsf{h}_{22} & \dots & \mathsf{h}_{2N} \\ \vdots & & & \vdots \\ \mathsf{h}_{N1} & \mathsf{h}_{N2} & \dots & \mathsf{h}_{NN} \end{bmatrix} \quad \text{and} \quad \mathsf{h}_{ij} = \begin{bmatrix} \partial^2 V/\partial r_{i_1}\partial r_{j_1} & \partial^2 V/\partial r_{i_1}\partial r_{j_2} & \partial^2 V/\partial r_{i_1}\partial r_{j_3} \\ \partial^2 V/\partial r_{i_2}\partial r_{j_1} & \partial^2 V/\partial r_{i_2}\partial r_{j_2} & \partial^2 V/\partial r_{i_2}\partial r_{j_3} \\ \partial^2 V/\partial r_{i_3}\partial r_{j_1} & \partial^2 V/\partial r_{i_3}\partial r_{j_2} & \partial^2 V/\partial r_{i_3}\partial r_{j_3} \end{bmatrix} \tag{3}$$

$$\partial^2 V/\partial r_{i_x}\partial r_{jy} = -\, k_{ij} \frac{(r_{j_x} - r_{i_x})\,(r_{j_y} - r_{iy})}{r_{ij}^2}\bigg|_{r_{ij}=r_{ij}^0} \tag{4}$$

Right side of Eq. (4) is closely related to the frequency of oscillation - this can be seen easily if we recall the result for frequency of a harmonic oscillator, which is $\omega^2 = k/m$. If we were to divide whole matrix H by $-1/m$, each term in h_{ij} would correspond to a squared frequency for oscillations around the equilibrium state. Since all α-carbon atoms have the same mass, this step can be actually omitted, since it would only scale the whole matrix by a numerical factor. We can define the eigenvalue problem as shown in Eq. (5), where U is the matrix of eigenvectors and Ω the diagonal matrix of eigenvalues. By decomposing the Hessian matrix, we find the eigenspace of the system, where each eigenvalue denotes a frequency of an eigenmode (normal mode), and each eigenvector a corresponding displacements of α-carbon atoms related to the normal mode.It turns out that the first six eigenvalues are all equal to zero, as those modes relate to three translations and three rotations in space.

$$\mathsf{H} = \mathsf{U}\Omega\mathsf{U}^{-1} \tag{5}$$

As stated before, smallest normal modes are related to collective motion of a molecule. It has been shown [28] that these modes agree well with the experimentally measured fluctuations, so those oscillations offer a good description of how a molecule will fluctuate around its equilibrium. Taking that into account we can analyze the structure and determine if some parts often oscillate in the same manner for the lowest couple of modes. This is exactly the idea behind essential dynamics method which is the topic of the next chapter.

3.3 Essential Dynamics

CG modeling strives to simplify the structure of a molecule while preserving most of its properties. One such property that can be preserved is the intrinsic dynamic motion of a molecule (how it bends, which areas are more flexible than the others, ...). Since we are able to calculate normal modes of a protein with NMA, we can then take a look at the lowest couple of modes and start to look for dynamic domains - groups of α-carbon atoms that usually oscillate in the same manner. This concept is realised with the essential dynamics (ED) method, where the idea is to simplify molecules structure whilst preserving its essential dynamical motions (in this case determined with normal mode analysis). It should be mentioned that ED can also be used with different methods that determine what motion is fundamental for a molecule. For example, instead of ENM in combination with NMA, principal component analysis of MD trajectories has been used to figure out what kind of movement is expected given a molecular structure [24]. Regardless of the technique used to analyse dynamic nature of the molecule, the algorithm of ED method stays the same and includes these few step:

1. Select M number of CG sites.
2. Randomly select $M-1$ AAs in the protein sequence. Those designate borders between AAs that will be coarse-grained into one site. We label them as border AAs.
3. Evaluate residual χ^2 as defined in Eq. (6).
4. Variate locations of border AAs, that define CG sites.

   (Repeat steps 3 and 4 until global minimum of χ^2 is reached.)

$$\begin{aligned} \chi^2 &= \frac{1}{3M}\sum_{\mathrm{S}=1}^{M}\sum_{i\in\mathrm{S}}\sum_{j\geq i\in\mathrm{S}}\langle(\Delta\mathbf{r}_i^{\mathrm{ED}})^2 - 2\Delta\mathbf{r}_i^{\mathrm{ED}}\mathbf{r}_j^{\mathrm{ED}} + (\mathbf{r}_j^{\mathrm{ED}})^2\rangle \\ \langle(\Delta\mathbf{r}_i^{\mathrm{ED}})^2\rangle &= k_{\mathrm{B}}T\sum_{x=1}^{3}\sum_{q=7}^{3M}\mathbf{u}_q^{i_x}\omega_q^{-1}\mathbf{u}_q^{i_x} \end{aligned} \tag{6}$$

Residual χ^2 is defined in such a way that it is minimised when groups of atoms have been determined so that atoms in each group oscillate most similarly. To better understand how this is written in Eq. (6), we can inspect it part by part; First sum goes over each dynamic domain (denoted with S) separately, while the second and third summations go over all pairs of atoms i and j that belong to domain S. For each such pair in domain S we calculate $(\Delta\mathbf{r}_i^{\mathrm{ED}})^2$ and $(\Delta\mathbf{r}_j^{\mathrm{ED}})^2$ which are mean squared fluctuations of atoms i and j, respectively, in the subspace of essential modes (first couple of modes). If atoms i and j were to oscillate exactly the same in the essential subspace, then $2\Delta\mathbf{r}_i^{\mathrm{ED}}\mathbf{r}_j^{\mathrm{ED}}$ would equal to $(\Delta\mathbf{r}_i^{\mathrm{ED}})^2$, which would result in this term in summation equaling to zero.

Mean squared fluctuations for a certain atom are calculated as described in the bottom line of Eq. (6). They can be calculated directly from the first $3M$ eigenmodes, while skipping first 6 modes as those correspond to translations and rotations in space, as described before. Eigenvectors are denoted with $\mathbf{u}$ and eigenvalues with ω, while i_x corresponds to spatial coordinate of atom i. The factor $k_{\mathrm{B}}T$ comes from the classical theory of networks [29].

Steps 3 and 4 from the list are carried out using computational methods for finding global minimum of a function like a combination of global simulated annealing and a local steepest descent search. Since the subspace of possible combinations when determining dynamical domains can increase substantially for large molecules and many CG sites, often times finding the global minimum becomes infeasible, which is a downside of ED method.

Once $M-1$ boarder atoms are calculated, CG sites are placed in the center of mass of each domain and the process of coarse-graining is finished.

4 Example: Coarse-Graining Monoclonal Antibodies

Antibodies can be seen as "Y" shaped proteins formed with two pairs of chains (of AAs) referred to as heavy and light chains. Each heavy chain can be again split into four parts, while light chains can be split into to two functional domains, making up a total of 12 separate regions, as shown in tile a) of Fig. 5 [30]. Monoclonal antibodies (mAbs) are structured in the same way, designation monoclonal only refers to the fact that they bind only to single antigen, which makes them important in pharmaceutical industry as they can be used as therapeutics for curing different diseases [16].

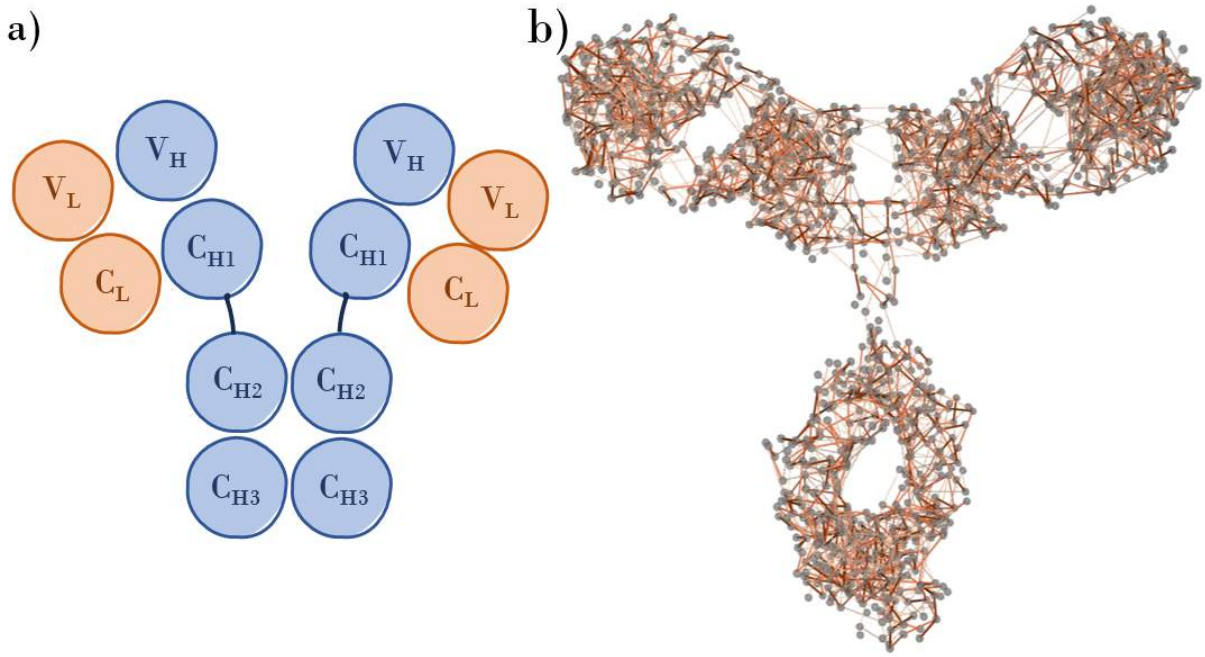


Figure 5: Schematic representation of a monoclonal antibody (a) and a visualisation of elastic network model mapped onto an antibody molecule (b). Springs of three different stiffness are drawn between pairs of atoms; red being the most stiff, orange a bit weaker and grey denoting loose springs.

For illustration of ED method, an example with a monoclonal antibody is presented here. Results have been obtained following the same steps as described before, so starting with a referenced structure, then constructing an elastic network model (see tile b) of Fig. 5), performing normal mode analysis and determining dynamical domains and CG sites using essential dynamics method.

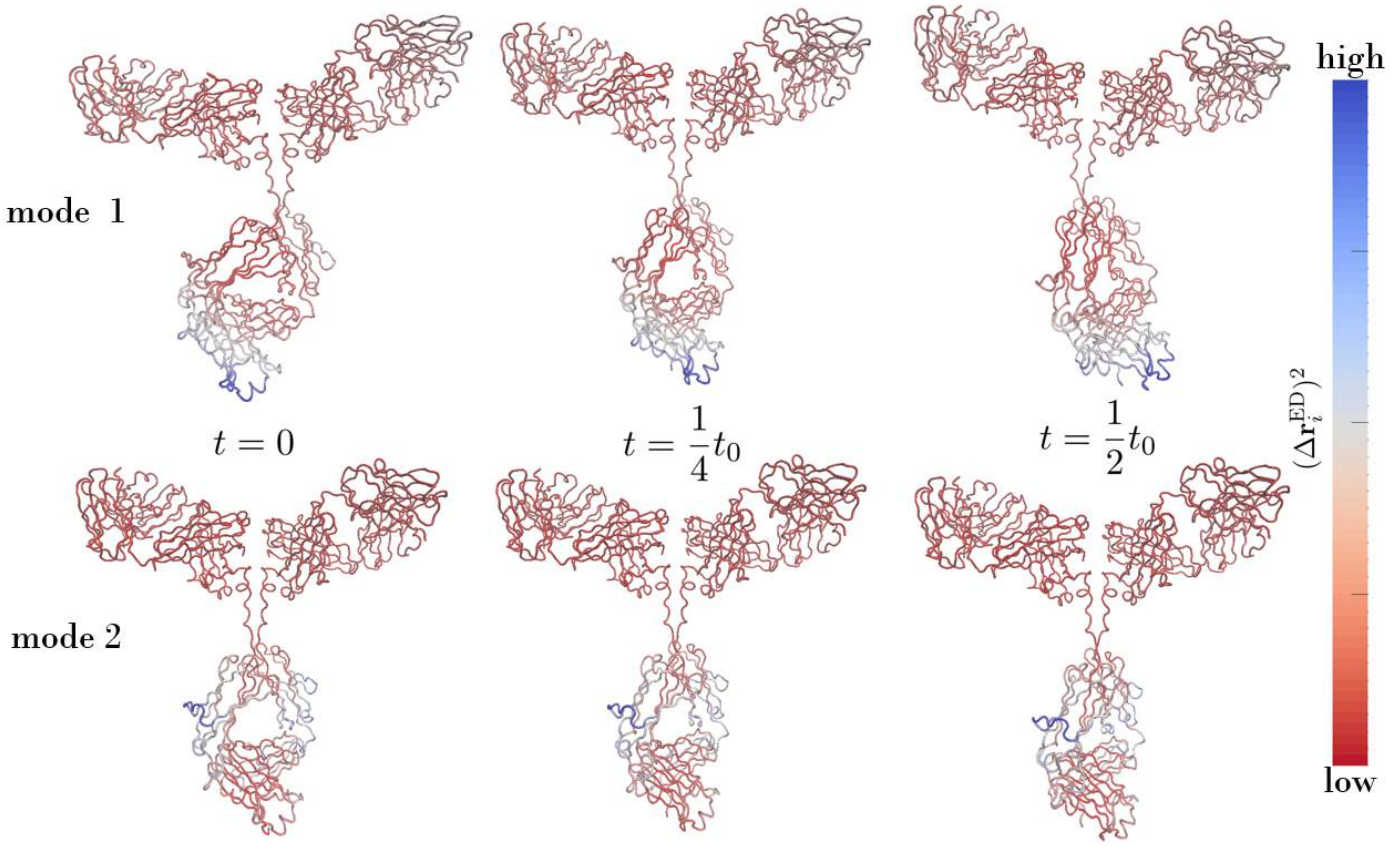


Figure 6: First two normal modes of a Tocilizumab monoclonal antibody. Magnitude of fluctuation $(\Delta \mathbf{r}_i^{\mathrm{ED}})^2$ is mapped onto the backbone, where red color indicates slight fluctuations and blue sizeable fluctuations. Modes are shown with three timeshots, corresponding to $t = 0$, $t = \frac{1}{4}t_0$ and $t = \frac{1}{2}t_0$, where t_0 is the oscillations period of a given mode. All images were prepared using Visual Molecular Dynamics software [31].

Figure 6 shows first two nonzero normal modes at three different timestamps with respect to period of oscillation t_0. Backbone of a proteins is colored on a spectrum from red, where fluctuations are small, to blue, where fluctuations are significant. The first normal mode describes a movement where bottom part of molecule oscillates back and forth between both other branches of a molecule, while second mode corresponds to rotation of the bottom part of heavy chains.

From calculated normal modes, a CG model is determined for different number of target number of CG sites. Results for $3, 6$ and 12 pseudo-atoms are shown in Fig. 7. When only three atoms are mapped, dynamical domains are both heavy chains separately and both light chains coupled together. For six sites the two light chains separate, and so do both heavy chains, which corresponds to 2 functional domains per CG site (see tile a) of Fig. 5). As we model 12 sites, each site represent a functional domain of a mAb, which is a sensible result, as several CG models made specifically for mAbs also divide the molecule in a similar fashion [12, 14, 22, 32].

5 Limitations and Advantages of ED Method

One of the strengths of ED method is the fact that CG model can be determined using only a single structure, which makes it especially useful in situations where we seek to determine a CG model based on dynamical domains but MD simulations of the system are not feasible. Indeed in this case extra effort should be devoted to obtaining a good structure in an energetic minimum state. Moreover ED method is computationally undemanding and

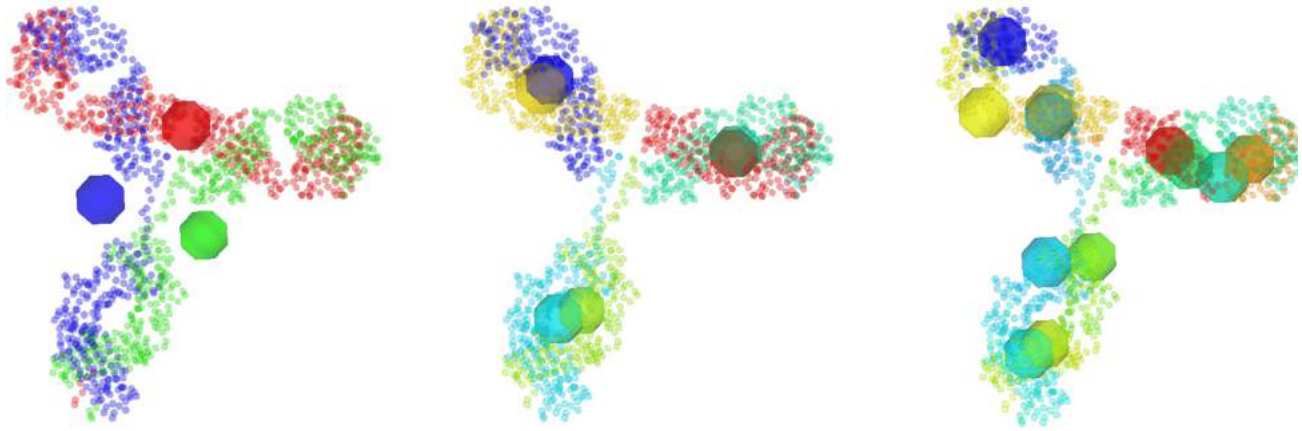


Figure 7: Essential dynamics CG models for $3, 6$ and 12 CG sites. Each domain is represented with a unique color, matching the pseudo-atom that is placed in the center of mass of each dynamic domain.

offers results that are in good agreement with similar methods.

There are also open challenges of ED method, major one being the fact that number of CG-sites needs to be defined by the user. Another challenege is that dynamical domains are determined as sequential parts in AA sequence, which may not be in keeping with the true nature of molecule's motion - artifacts that stem from this property can be spotted in Fig. 7 for 3 CG sites, where a part of dynamical domain colored with red includes a couple of α-carbon atoms at the bottom of molecule. Last challenge is that the conformational space of possible CG sites can become too big and reaching global minimum would not be likely.

6 Conclusion

Essential dynamics method gives promising results, as it can reproduce CG models that were tailored for specific types of protein on its own. As similar CG models, that define CG sites so that dynamical domains are preserved, it offers a good representation of a simplified structure that can still exhibit comparable movement to the all-atom model, but is able to do that without any data from molecular dynamics simulations. Interesting work will be to correlate and compare the calculated protein dynamics with selected protein solution quality attributes.

References

[1] Ashkan Vaziri and Arvind Gopinath. Cell and biomolecular mechanics in silico. *Nature materials*, 7(1):15–23, 2008.

[2] Sebastian Kmiecik, Dominik Gront, Michal Kolinski, Lukasz Wieteska, Aleksandra Elzbieta Dawid, and Andrzej Kolinski. Coarse-Grained Protein Models and Their Applications. *Chemical Reviews*, 116(14):7898–7936, July 2016. Publisher: American Chemical Society.

[3] Cesar A Lopez, Andrzej J Rzepiela, Alex H de Vries, Lubbert Dijkhuizen, Philippe H Hunenberger, and Siewert J Marrink. Martini coarse-grained force field: extension to carbohydrates. *Journal of Chemical Theory and Computation*, 5(12):3195–3210, 2009.

[4] Davit A Potoyan, Alexey Savelyev, and Garegin A Papoian. Recent successes in coarse-grained modeling of dna. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 3(1):69–83, 2013.

[5] Sergei Izvekov and Gregory A Voth. A multiscale coarse-graining method for biomolecular systems. *The Journal of Physical Chemistry B*, 109(7):2469–2473, 2005.

[6] Michael Levitt and Arieh Warshel. Computer simulation of protein folding. *Nature*, 253(5494):694–698, 1975.

[7] Ken A Dill, S Banu Ozkan, M Scott Shell, and Thomas R Weikl. The protein folding problem. *Annu. Rev. Biophys.*, 37:289–316, 2008.

[8] Tristan Bereau and Markus Deserno. Generic coarse-grained model for protein folding and aggregation. *Biophysical Journal*, 96(3):405a, 2009.

[9] Luca Monticelli, Senthil K Kandasamy, Xavier Periole, Ronald G Larson, D Peter Tieleman, and Siewert-Jan Marrink. The martini coarse-grained force field: extension to proteins. *Journal of chemical theory and computation*, 4(5):819–834, 2008.

[10] Carol A Rohl, Charlie EM Strauss, Kira MS Misura, and David Baker. Protein structure prediction using rosetta. In *Methods in enzymology*, volume 383, pages 66–93. Elsevier, 2004.

[11] Julien Maupetit, P Tuffery, and Philippe Derreumaux. A coarse-grained protein force field for folding and structure prediction. *Proteins: Structure, Function, and Bioinformatics*, 69(2):394–408, 2007.

[12] Anuj Chaudhri, Isidro E. Zarraga, Tim J. Kamerzell, J. Paul Brandt, Thomas W. Patapoff, Steven J. Shire, and Gregory A. Voth. Coarse-Grained Modeling of the Self-Association of Therapeutic Monoclonal Antibodies. *The Journal of Physical Chemistry B*, 116(28):8045–8057, July 2012.

[13] Anuj Chaudhri, Isidro E. Zarraga, Sandeep Yadav, Thomas W. Patapoff, Steven J. Shire, and Gregory A. Voth. The Role of Amino Acid Sequence in the Self-Association of Therapeutic Monoclonal Antibodies: Insights from Coarse-Grained Modeling. *The Journal of Physical Chemistry B*, 117(5):1269–1279, February 2013.

[14] Hassan Shahfar, James K. Forder, and Christopher J. Roberts. Toward a Suite of Coarse-Grained Models for Molecular Simulation of Monoclonal Antibodies and Therapeutic Proteins. *The Journal of Physical Chemistry B*, 125(14):3574–3588, April 2021.

[15] Glenn M Ferreira, Cesar Calero-Rubio, Hasige A Sathish, Richard L Remmele Jr, and Christopher J Roberts. Electrostatically mediated protein-protein interactions for monoclonal antibodies: a combined experimental and coarse-grained molecular modeling approach. *Journal of Pharmaceutical Sciences*, 108(1):120–132, 2019.

[16] Dawn M Ecker, Susan Dana Jones, and Howard L Levine. The therapeutic monoclonal antibody market. In *MAbs*, volume 7, pages 9–14. Taylor & Francis, 2015.

[17] Robert C Robinson, Kirsi Turbedsky, Donald A Kaiser, Jean-Baptiste Marchand, Henry N Higgs, Senyon Choe, and Thomas D Pollard. Crystal structure of arp2/3 complex. *science*, 294(5547):1679–1684, 2001.

[18] Elaine C Meng, Thomas D Goddard, Eric F Pettersen, Greg S Couch, Zach J Pearson, John H Morris, and Thomas E Ferrin. Ucsf chimerax: Tools for structure building and analysis. *Protein Science*, 32(11):e4792, 2023.

[19] Carl Ivar Branden and John Tooze. *Introduction to protein structure*. Garland Science, 2012.

[20] P.M. Flatt. Biochemistry: Defining life at the molecular level, 2019. Available at: https://wou.edu/chemistry/courses/online-chemistry-textbooks/ch450-and-ch451-biochemistry-defining-life-at-the-molecular-level/, last accesed on: [6.2.2024].

[21] Helgi I Ingolfsson, Cesar A Lopez, Jaakko J Uusitalo, Djurre H de Jong, Srinivasa M Gopal, Xavier Periole, and Siewert J Marrink. The power of coarse graining in biomolecular simulations. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 4(3):225–248, 2014.

[22] Cesar Calero-Rubio, Atul Saluja, and Christopher J Roberts. Coarse-grained antibody models for "weak" protein–protein interactions from low to high concentrations. *The Journal of Physical Chemistry B*, 120(27):6592–6605, 2016.

[23] Zhiyong Zhang, Jim Pfaendtner, and Gregory A. Voth. Defining Coarse-Grained Representations of Large Biomolecules and Biomolecular Complexes from Elastic Network Models. *Biophysical Journal*, 97(8):2327–2337, October 2009.

[24] Zhiyong Zhang, Lanyuan Lu, Will G. Noid, Vinod Krishna, Jim Pfaendtner, and Gregory A. Voth. A systematic methodology for defining coarse-grained sites in large biomolecules. *Biophysical Journal*, 95(11):5073–5083, December 2008.

[25] Yuichi Togashi and Holger Flechsig. Coarse-Grained Protein Dynamics Studies Using Elastic Network Models. *International Journal of Molecular Sciences*, 19(12):3899, December 2018.

[26] Edvin Fuglebakk, Sandhya P Tiwari, and Nathalie Reuter. Comparing the intrinsic dynamics of multiple protein structures using elastic network models. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1850(5):911–922, 2015.

[27] Qiang Cui Bahar, Ivet, editor. *Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems*. Chapman and Hall/CRC, New York, December 2005.

[28] Herman JC Berendsen and Steven Hayward. Collective protein dynamics in relation to function. *Current Opinion in Structural Biology*, 10(2):165–169, April 2000.

[29] P. J. Flory, M. Gordon, P. J. Flory, and N . G. McCrum. Statistical thermodynamics of random networks. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 351(1666):351–380, January 1997. Publisher: Royal Society.

[30] Eduardo A Padlan. Anatomy of the antibody molecule. *Molecular immunology*, 31(3):169–217, 1994.

[31] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.

[32] Anuj Chaudhri, Isidro E. Zarraga, Tim J. Kamerzell, J. Paul Brandt, Thomas W. Patapoff, Steven J. Shire, and Gregory A. Voth. Coarse-Grained Modeling of the Self-Association of Therapeutic Monoclonal Antibodies. *The Journal of Physical Chemistry B*, 116(28):8045–8057, July 2012.