# Fast, Accurate, and System-Specific Variable-Resolution Modeling of Proteins

Raffaele Fiorentini, Thomas Tarenzi, and Raffaello Potestio*

Cite This: https://doi.org/10.1021/acs.jcim.2c01311

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

**ABSTRACT:** In recent years, a few multiple-resolution modeling strategies have been proposed, in which functionally relevant parts of a biomolecule are described with atomistic resolution, with the remainder of the system being concurrently treated using a coarse-grained model. In most cases, the parametrization of the latter requires lengthy reference all-atom simulations and/or the usage of off-shelf coarse-grained force fields, whose interactions have to be refined to fit the specific system under examination. Here, we overcome these limitations through a novel multiresolution modeling scheme for proteins, dubbed coarse-grained anisotropic network model for variable resolution simulations, or CANVAS. This scheme enables a user-defined modulation of the resolution level throughout the system structure; a fast parametrization of the potential without the necessity of reference simulations; and the straightforward usage of the model on the most commonly used molecular dynamics platforms. The method is presented and validated with two case studies, the enzyme adenylate kinase and the therapeutic antibody pembrolizumab, by comparing the results obtained with the CANVAS model against fully atomistic simulations. The modeling software, implemented in Python, is made freely available for the community on a collaborative github repository.
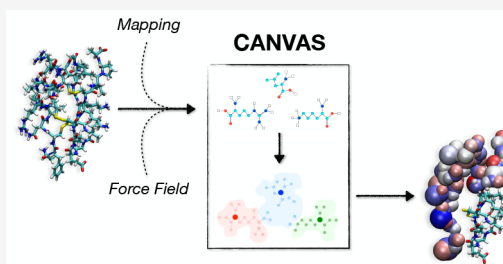
## INTRODUCTION

Steady improvements in high-performance computing hardware and molecular dynamics (MD) simulation software over several decades have ushered in impressive advancements in the computer-aided investigation of soft and biological matter systems, in particular, macromolecules of biological origin such as lipids, proteins, and nucleic acids.[1−3] At the same time, a detailed modeling of molecular systems in which each atom is described as an interaction center often turns out to be inconvenient or even undesirable—on the one hand, because of major computing and data storage requirements and, on the other hand, because of the effort in analyzing the simulation outcome. To overcome both limitations, simplified, *coarse-grained* (CG) models,[4−9] in which several atoms are lumped together in effective interaction sites, are frequently employed. CG models enable the simulation of larger systems over longer time scales, thanks to a smoother (free) energy profile and fewer degrees of freedom with respect to all-atom representations.

Coarse-grained models have been successfully employed for a number of biologically and pharmacologically relevant applications. These include the study of spontaneous protein—ligand binding,[10] where both the macromolecule, the ligand, and the solvent are modeled in a coarse-grained fashion. The approach proved useful for the identification of binding pockets and the estimation of binding free energies on a number of systems; however, the simplified representation of the ligand requires a not-so-obvious new parametrization of the interactions and limits the distinction between similar

molecules (as in the case of enantiomers).[10] In addition, the employment of a coarse-grained solvent model limits the accuracy in the case where single water molecules are actively involved in the stabilization of the ligand in the binding site. Similarly, a number of coarse-grained force fields, including, among others, MARTINI,[11,12] SIRAH,[13,14] AWSEM,[15] and Scorpion[16] force fields, have been used to investigate protein—protein interactions, providing accurate results in terms of binding free energies. However, those models prevent an even coarser representation of the protein interfaces, which might be desirable in the case of very large protein assemblies; furthermore, they do not provide an accurate description of the system if specific atomistic details, possibly crucial for the properties or behavior of interest, are effectively integrated out in the low-resolution model.

Hence, whereas all-atom models provide the necessary accuracy at the expense of substantial computational resources, CG models enable a more efficient and intelligible representation of the system at the cost of losing possibly crucial detail. Although several problems in computational biophysics can be tackled with one of these two methods,

many open questions remain that necessitate an approach at the interface between chemical accuracy and computational efficiency. In this regard, methods have been developed in which molecules described at different resolutions are simultaneously simulated within the same setup. Examples include the following: coupling of MARTINI with atomistic force fields;[17,18] the simulation of atomistic proteins and nucleic acids in a multiresolution solvent with the SWINGER algorithm;[19,20] and the simulation of soluble proteins with the PACE force field, which has been developed with the specific aim of coupling united-atom protein models with a coarse-grained solvent representation.[21,22] Pushing the "resolution mix" even further, in some applications it might be desirable to couple different levels of detail within the same biomolecule to limit the computationally expensive high-resolution modeling to a subset of protein residues or nucleic acid base pairs. This approach was pioneered by the quantum mechanics/molecular mechanics (QM/MM) methods,[23−27] which allow a connection between a small region where *ab initio* models are used, and a classical all-atom description in the remainder of the system. Along the same lines, several methodologies have been developed to couple atomistic and coarse-grained levels of resolution within the same simulation setup, and even within the same molecular structure. For example, in the Molecular Mechanics/Coarse-Grained (MM/CG) scheme developed in 2005 by Neri et al.,[28] the atomistically detailed active site is incorporated into a coarse-grained Gō-like model, which aims at reproducing the correct conformational fluctuations of the full protein.[29] The MM/CG method was later tailored for the simulation of membrane protein/ligand complexes,[30] and in the last version of the method, dubbed open-boundary MM/CG,[31] the dual-resolution description of the protein is coupled with an adaptive multiscale model of the solvent, namely, the Hamiltonian adaptive resolution scheme (H-AdResS);[32,33] in the latter, regions of different resolution are defined in the simulation box, allowing water molecules to change their resolution *on the fly* when diffusing from one region to the other. More recently, a similar method[34,35] employed a high-resolution force field in small regions of a protein, most notably, the active site, while treating the remainder in a coarse-grained fashion, for example, as an elastic network model.[36]

Dual-resolution methods have been successfully applied for the study of several biological systems, including soluble[35] and membrane proteins.[37−39] However, the available approaches share some common shortcomings: first, the standard modeling of the CG region allows little flexibility in the choice of the CG sites; second, the CG region is usually defined *ad hoc*, and new mappings require a completely new reparametrization of the interactions; third, nonbonded interactions (such as electrostatics) are typically not taken into account in the CG model, thus preventing interactions between different structural domains that might come in close contact during the course of the simulation. CG models with an accurate description of electrostatics have been developed;[40−42] however, in such cases, the protein is uniformly coarse-grained at a resolution intermediate between the atomistic and one-bead-per-amino acid one, thus limiting the level of coarse-graining and preventing a straightforward coupling between regions at different resolutions. These limitations hinder the applicability of standard multiple-resolution models, with detrimental consequences for the *in silico* investigation of proteins and their interactions.

In this work we propose a novel approach, dubbed coarse-grained anisotropic network model for variable resolution simulations, or CANVAS, which enables a fast parametrization of multiple-resolution models. The CANVAS strategy leverages the blurred and approximate nature of coarse-grained models to identify effective sites based on a user-provided input and determines the interactions among them based on the molecule's structure and all-atom force field, making it unnecessary to run reference simulations. This strategy makes the parametrization of the model practically instantaneous and allows the modulation of the system's resolution in a quasi-continuous manner across the structure, from all-atom to (very) coarse-grained. Most notably, the interaction between regions of the system at different resolutions (including the solvent) is accounted for and straightforward to set up, allowing the seamless implementation in standard MD software packages (e.g., GROMACS or LAMMPS).

The article is structured as follows: first, we describe in detail the CANVAS model, focusing on the construction of the multiple-resolution representation and on the parametrization of the interactions. A Methods section follows, providing the simulation details. The results of the validation of the CANVAS approach are then presented by comparing results from all-atom and multiscale simulations of two biomolecules, namely, the enzyme adenylate kinase and the IgG4 antibody pembrolizumab, each modeled with three resolution levels. Finally, conclusions and perspectives are discussed.

## ■ THE CANVAS MODEL

In the CANVAS approach to multiresolution protein modeling, a decimation mapping is implemented for the choice of the interactions sites:[9] those atoms included in a user-defined list are retained, whereas the other ones are discarded. If all atoms of a given subregion of the molecule are retained, the high-resolution atomistic description is employed; in contrast, regions where atoms are removed are described at a varying level of detail. In lower-resolution regions, the physical properties of the survived atoms are modified so as to incorporate in effective interactions those atoms that have been integrated out (Figure 1). Specifically, each discarded atom is associated with the closest surviving one, and the properties of the latter are determined from those of the group of discarded atoms it represents.

The CANVAS model enables in principle a quasi-continuous modulation of the resolution of a protein or part of it, in that the detail of the representation can be gradually reduced from the all-atom level to a very coarse one, possibly lower than a few (from one to three) amino acids per bead; between the highest and lowest resolutions, an arbitrary number of intermediate levels are feasible. In the current implementation, we performed the choice of employing three levels of resolution:

- **all-atom (AT)**: the highest level of detail, where all the atoms of a given amino acid are retained;
- **medium-grained (MG)**: intermediate level of detail, where only the backbone atoms of an amino acid are retained, that is, the carbon alpha $CA_{mg}$, the nitrogen $N_{mg}$ of the amino group, the oxygen $O_{mg}$, and the carbon $C_{mg}$ of the carboxyl group;
- **coarse-grained (CG)**: the lowest level of resolution (in the applications presented here, only the $C_\alpha$ atoms of each CG residue are kept, dubbed $CA_{cg}$.
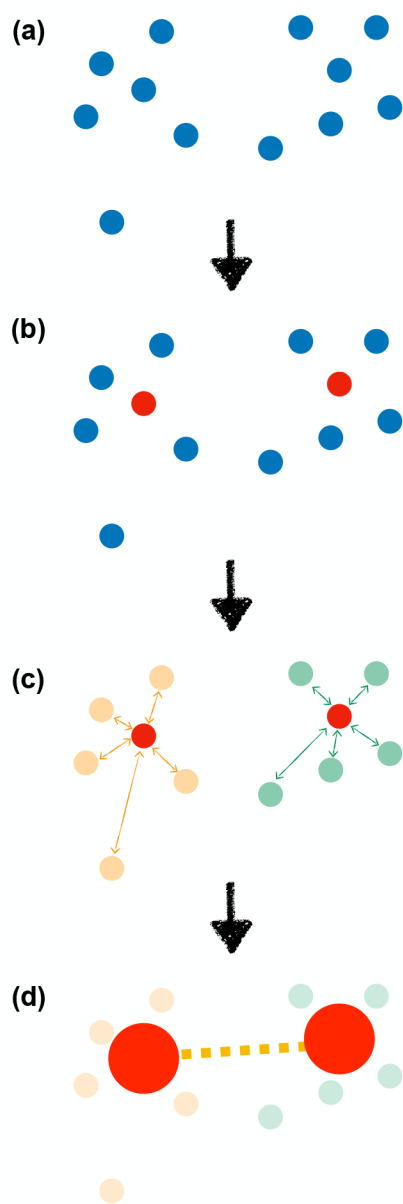
**Figure 1.** Scheme of the decimation process in the low-resolution part of the biomolecule. (a) Blue circles show all the atoms in the low-resolution part. (b) Choice of atoms that survive depicted in red. (c) Decimated atoms (light orange and light green) are mapped onto their closest survived atom in terms of Euclidean distance (orange and green arrows), according to the Voronoi tessellation. (d) Each survived atom, shown with a large red circle, is representative of the closest not-survived atoms mapped by it. A harmonic spring, depicted with a dashed yellow line, connects the neighboring survived atoms.

The sets of protein residues modeled with an AT, MG, or CG detail are specified by the user and do not change during the simulation, that is, the biomolecule has a time-independent triple resolution. Table 1 summarizes the survived atoms in each region and their label.

The first step of the model construction is to identify the region of the system where the chemical details play a crucial role such that no simplification of the atomistic description is desirable. Residues described at MG and CG resolutions can be either specified from the user or directly identified on the

**Table 1. Description of Survived Atoms for Amino Acids (aa) for Each Level of Resolution**

| label | region | survived atoms per aa |
|---|---|---|
| at | fully-AT | all ($CA_{at}$, $N_{at}$, etc.) |
| mg | medium-grained | backbone ($N_{mg}$, $CA_{mg}$, $C_{mg}$, $O_{mg}$) |
| cg | coarse-grained | $C_\alpha$ ($CA_{cg}$) |

basis of the atomistic residues; in the latter case, the MG region is built by including those residues at a distance of 1 nm from the AT region, whereas the rest of the biomolecule is automatically assigned a CG representation.

The AT part is modeled through a standard atomistic force field (in the implementation discussed here, these are Amber99SB-ildn[43] or CHARMM36m[44]), where the classical functional form and parametrization of the bonded and nonbonded interactions between atoms are employed. In the MG and CG domains, the potential energy is given by

$$E = E_{AA} + E_{harmonic} + E_{VdW} + E_{coulomb} \tag{1}$$

The first term, $E_{AA}$, corresponds to bonded interactions from the atomistic force field, namely, chemical bonds, angles, and proper/improper dihedrals:

$$E_{AA} = E_{bonds}h(r) + E_{angles}h(\theta) + E_{dihedrals}h(\phi)$$
$$+ E_{improper}h(\omega) \tag{2}$$

Here, $h(r)$, $h(\theta)$, $h(\phi)$, and $h(\omega)$ are Heaviside functions taking the value 1 if a bond, angle, dihedral, or improper dihedral exists in the atomistic force field for a couple, triplet, or quadruplet of survived atoms. Therefore, stretching, bending, and torsion potentials with their original equilibrium values are possible only if, respectively, the pair, triplet, and quadruplet of atoms (where at least a CG bead is involved) from the all-atom representation of reference are maintained in the MG and CG regions. The second term in eq 1, $E_{harmonic}$, describes the bonded interactions between and within the low-resolution domains. The bonded connectivity and its para-metrization are strictly dependent on the resolution levels employed and on the chemical nature of the retained sites, namely, on their *atom type*. In the current implementation, beads are connected by harmonic springs as schematically depicted in Figure 1d and described in detail in Figure 2. Specifically, the reference bond length corresponds to the distance between the two atoms/beads in the starting structure, whereas the value of the elastic constant depends on the nature of the bonded particles and their position along the sequence: (1) A stiff spring ($k_b$) is employed for consecutive beads (red line of Figure 2); its value is $5 \times 10^4$ kJ·mol$^{-1}$·nm$^{-2}$. (2) A weaker spring $k_{nb}$ is used for non-consecutive $C_\alpha$ beads ($CA_{cg} - CA_{mg}$, $CA_{cg} - CA_{cg}$, $CA_{mg} - CA_{mg}$) whose distance in the reference (native) conformation lies below a fixed cutoff equal to 1.4 nm (orange line of Figure 2). Critically, the magnitude of $k_{nb}$ depends on the distance $d$ between the two $C_\alpha$ beads, farther CG units interacting through looser springs. The profile of $k_{nb}(d)$ was obtained through a statistical analysis performed over an ensemble of effective pair potentials acting among nonconsecutive $C_\alpha$ atoms in the pembrolizumab antibody; such potentials were extracted via direct Boltzmann inversion. See Section S1 in the Supporting Information for additional technical details. (3) A second weaker spring $k_{if}$ is employed between an atomistic $C_\alpha$ and a CA bead ($CA_{at} - CA_{mg}$ or $CA_{at} - CA_{cg}$) if they do not
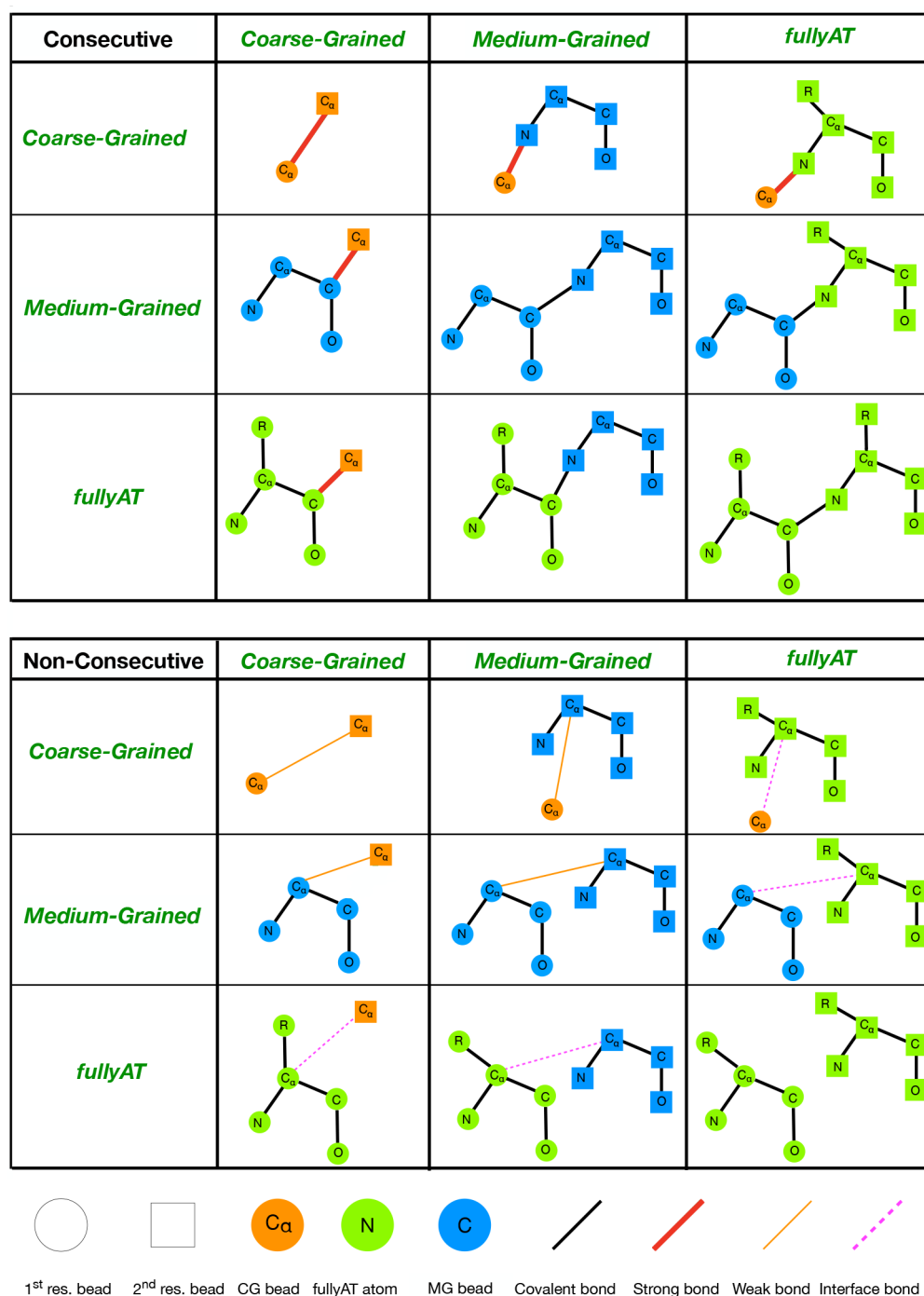
**Figure 2.** Schematic representation of the bonded interactions in the three regions at different resolutions. On the top of the figure, only consecutive residues are considered, and on the bottom, nonconsecutive ones. The atoms/beads that belong to the first residue are traced with a circle, whereas those that belong to the second residue are sketched with a square. R stands for the side chain. In the figure, hydrogen atoms are ignored for clarity, while being explicitly accounted for in the model. Bonded interactions are represented with different colors and thicknesses according to the spring constant.

belong to consecutive residues, and their distance in the reference conformation is less than a fixed cutoff equal to 1.4 nm (magenta dashed line of Figure 2). The recommended value of $k_{if}$ is 50 kJ·mol$^{-1}$·nm$^{-2}$ to guarantee the appropriate degree of flexibility.

We stress that if the two survived atoms taken into account are connected by a covalent bond in the fully atomistic representation, the latter replaces $E_{harmonic}$ (black line of Figure 2). Similarly, bending and torsion potentials with their original

atomistic parametrization are maintained if the triplet and quadruplet of atoms (where at least a CG bead is involved) in the all-atom representation of reference are retained in the coarse regions. Rescaled nonbonded 1−4 interactions are introduced only in the AT region. In addition, to guarantee the correct degree of flexibility in multidomain proteins, no bond is introduced between those beads that are close in space in the starting configuration but belong to distinct structural domains; the latter can be defined either on the basis of the

knowledge of the system or through appropriate algorithms developed to decompose protein structures in rigid sub-units.[45,46] The indices of the residues belonging to each domain are specified by the user in an optional input file.

Finally, $E_{VdW}$ and $E_{coulomb}$ in eq 1 are the van der Waals and Coulomb nonbonded contributions to the potential energy between nodes. For the AT region, standard force-field parameters are taken, whereas in the MG and CG regions, the charge and Lennard-Jones parameters of each bead are computed from the average properties of the neighboring atoms, as defined through a procedure akin to a Voronoi tessellation.[47−49] First, a Voronoi cell is defined by associating the decimated atoms (blue circles of Figure 1b) to the closest survived atom (in terms of Euclidean distance $l$), which is now treated as a CG bead (Figure 1c and Figure 3b,c). We emphasize that because a geometric criterion is employed to group atoms, the resulting bead is representative of atoms that
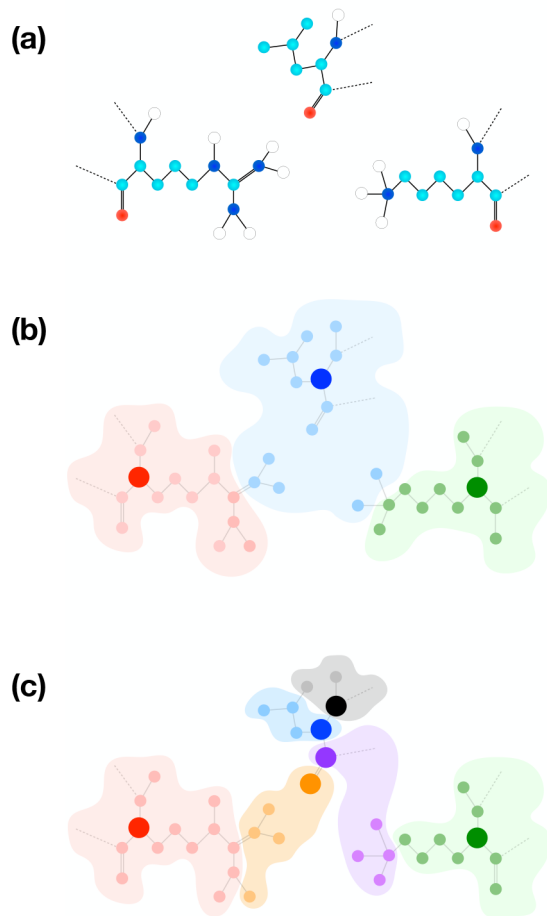


**Figure 3.** (a) Schematic representation of three amino acids: arginine, leucine, and lysine (from left to right). Dashed black lines represent the peptide bonds between two residues. The aliphatic hydrogens are not displayed for simplicity. (b) All three amino acids are modeled as CG, where only the $C_\alpha$ atom $CA_{cg}$ (red, blue, and green bigger circles) are retained. The other atoms are decimated and mapped onto the closest survived atom (shown in pink, light blue, and light green). A bead is not necessarily representative of atoms belonging to the same residue since the grouping criterion is merely based on euclidean distance. (c) Arginine and lysine are modeled in CG (red and green bigger circles), whereas the leucine is described in MG ($CA_{mg}$ in blue, $N_{mg}$ in black, $C_{mg}$ in violet, and $O_{mg}$ in orange).

could also belong to separate residues, as schematically shown in Figure 3. For this reason, the protein's starting structure plays a relevant role in the Voronoi tessellation because the relative orientation of side chains might influence the construction of the cells. Therefore, it is important that the structure employed as a reference for coarse graining is minimized and equilibrated. For the same reason, the Voronoi tessellation-based coarse-graining procedure is strongly dependent on the starting structure, and we can expect the relative arrangements of secondary elements to be preserved during the simulation. If conformational changes are desirable, a careful distribution of the different degrees of resolution along the structure is required, and a more informed partition of the system should be done with explicit input from the user.

After the definition of the Voronoi cells, nonbonded potential parameters are computed for each CG bead. Specifically, for a mapping that retains $N$ atoms out of $n$: (1) The charge $Q_I$ is defined as the algebraic sum of the charges $q_i$ of the atoms it represents:

$$Q_I \equiv \sum_{i \in I} q_i \tag{3}$$

(2) The diameter $\sigma_I$ is twice the gyration radius $R_g$:

$$\sigma_I \equiv 2R_g \tag{4}$$

where

$$R_g{}^2 = \frac{1}{N} \sum_{i=1}^{N} |\mathbf{r}_i - \mathbf{r}_{cog}|^2$$

$$\mathbf{r}_{cog} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{r}_i$$

Here, $\mathbf{r}_i$ represents the coordinates of each atom, whereas $\mathbf{r}_{cog}$ corresponds to the coordinates of the center of geometry of the group. (3) $\epsilon_I$ is the geometric average of the $\epsilon_i$ values of the atoms it represents:

$$\epsilon_I \equiv \prod_{i \in I} \epsilon_i^{1/N} \tag{5}$$

As opposed to the network of bonded interactions, where a predefined set of parameters is employed, the nonbonded part is automatically constructed on the basis of the properties of the retained sites, independently of the level of resolution and the bonded connectivity between them. The combination rule used to determine, from these parameters, Lennard-Jones interactions between CG beads is the same as the one employed by the atomistic force field in the high-resolution region; namely, it is based on the Lorentz−Berthelot rules for both the Amber and CHARMM force fields. In addition, in the case of interactions between nonconsecutive coarse-grained sites, nonbonded interactions are fully accounted for, whereas nonbonded interactions are switched off in the case of bonds involving atoms in the high-resolution region, as in the standard atomistic description.

We stress that because CG beads in the CANVAS representation may not be representative of a single residue, a direct residue-based analysis can not be performed. This is a specific feature of the CANVAS approach: the latter, in fact, was conceived to be easily generalized to very coarse mappings, where one bead is representative of more than one residue, or for inhomogeneous mappings, where the retained low-

resolution sites are distributed throughout the protein independently from the residue at which they originally belong (so that some residues might be represented by one or more beads, whereas others might be discarded completely). In such a case one can use mappings that are different from the intuitive, chemistry-based ones, but that are the most efficient in preserving the information contained in the all-atom protein representation.[9,50]

The code and examples of input files for simulating a system with the CANVAS model are freely available at https://github.com/potestiolab/canvas. The code consists of two Python scripts: the first one (*block.py*) has the purpose of creating the list of survived atoms with their relative labels (AT, MG, or CG); the second script (*CANVAS.py*) returns the input files needed for simulating a solvated biomolecule in LAMMPS or GROMACS, according to the choice made by the user. The mandatory arguments for the successful execution of the code are the list of survived atoms, the coordinate file (.gro) and the topology file (.top) of fully atomistic representation. A detailed description of the other parameters (mandatory and optional) and a tutorial for the construction and simulation of a CANVAS model, starting from the atomistic representation, are available on the same repository.

## ■ MATERIALS AND METHODS

The two systems employed in the present work as a test bed for the CANVAS model are the enzyme adenylate kinase[51−53] and the antibody pembrolizumab.[54]

Adenylate kinase (ADK) plays a critical role in maintaining the energetic balance in the cell, interconverting adenosine diphosphate (ADP) molecules into adenosine monophosphate (AMP) and adenosine triphosphate (ATP).[55] The structure of ADK can be partitioned in three domains, called the CORE, LID, and NMP, and two distinct nucleotide binding sites, as shown in Figure 4a,b.

The second system used here as a test case, pembrolizumab, is a humanized IgG4 antibody consisting of four chains, covalently bound by disulfide bonds (Figure 5a). Pembrolizumab—which is the generic name for the trade drug name Keytruda—is currently used in immunotherapy as an anticancer drug.[56] Its antigen is the programmed cell death protein 1 (PD-1), expressed on the membrane of T cells, B cells, and natural killer cells; the formation of the high-affinity complex between the antibody and its antigen prevents the binding of PD-1 with the programmed cell death receptor ligands PD-L1 and PD-L2, which would lead to a suppression of the antitumor activity of T cells.[57]

The reference structures employed for the construction of the multiscale models were obtained from equilibrated all-atom simulations. Specifically, the crystallographic ADK structure (PDB ID: 4AKE) was solvated in an aqueous solution at 0.15 M NaCl concentration; after energy minimization, the system was equilibrated for 125 ns in the NPT ensemble, using the Parrinello−Rahman barostat[58] with a time constant of 2.0 ps at 1 bar and the Langevin thermostat[59] to keep the temperature at 300 K. The all-atom simulation was extended for 500 ns, on which the analyses were performed. For the CANVAS simulation, the equilibrated structure was placed in a cubic simulation box of 9.1 nm per side and solvated in an aqueous solution at 0.15 M NaCl concentration.

The reference structures of pembrolizumab are given by the representative conformations sampled from four all-atom 500 ns long simulations of the antibody in the apo form, after
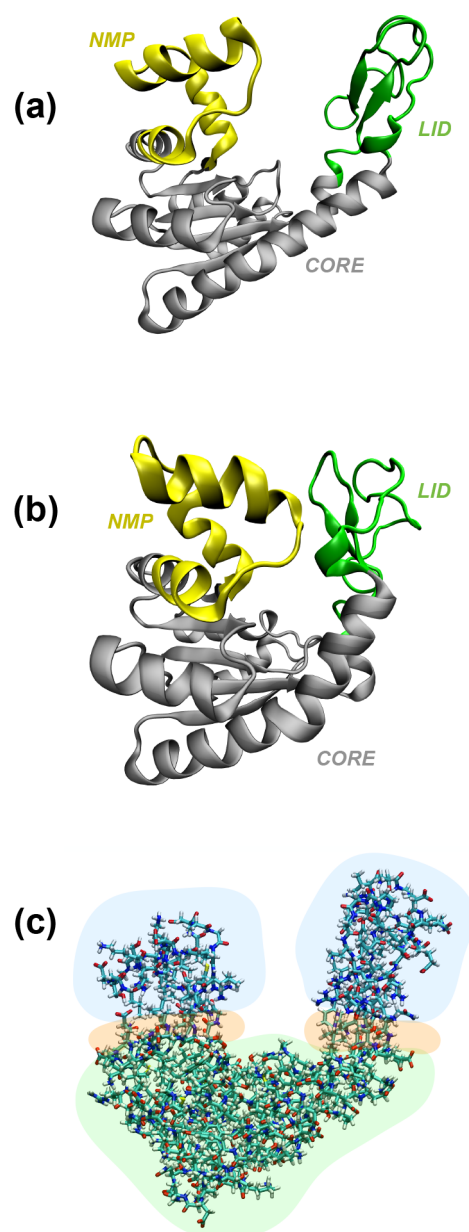


**Figure 4.** Fully atomistic representation of ADK. In particular, (a) and (b) show the open and compact conformation of the protein, respectively, in terms of secondary structure. LID, NMP, and CORE domains are depicted in green, yellow, and gray. (c) displays a schematic representation of the reference structure of ADK before conversion from the all-atom representation to the CANVAS one. Specifically, the CORE of the protein, modeled atomistically, is depicted in green; the part that is described in MG is shown in orange; the remainder, which is going to be coarse-grained, is shown in blue.

clustering the frames on the basis of their structural similarity. Each of these atomistic simulations was started from the PDB crystallographic structure of the deglycosilated antibody (PDB ID: 5DK3) after modeling of the missing residues; for more details on the all-atom simulation protocol, the reader is referred to ref 60. A CANVAS simulation is started from each representative conformation of the antibody, for a total of six different runs; this choice is dictated by the large conformational variability of the molecule, and the peculiar properties of each conformational basin. The CANVAS models of the
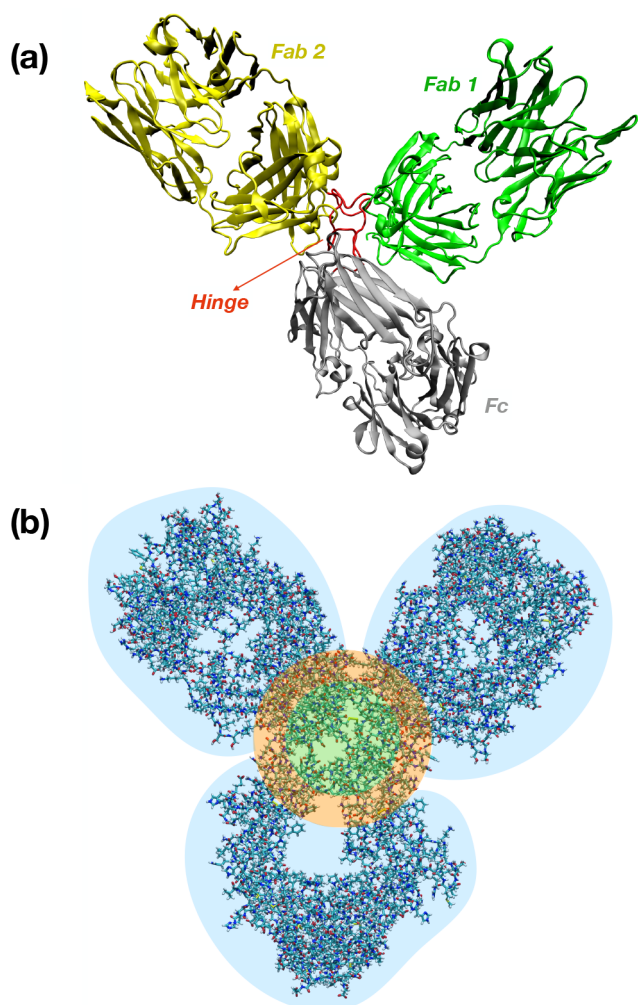
**Figure 5.** (a) Graphical representation of the crystallographic conformation of pembrolizumab in terms of secondary structure. Fab 1, Fab 2, Fc, and the hinge are depicted in green, yellow, gray, and red, respectively. (b) Schematic representation of the $4_A$ reference conformation of pembrolizumab before conversion from the all-atom to the multiresolution. In particular, in green is depicted the hinge of the biomolecule, modeled atomistically; in orange is shown the part that is going to be described as MG; the remainder, shown in blue, is going to be represented as CG.

integration step is 2 fs. The calculation of electrostatic interactions is performed in all cases by using the reaction-field method[62,63] with a dielectric constant of $\epsilon = 80$ and a cutoff of $2.5\sigma_{\max}$; here, $\sigma_{\max}$ is the maximum value of $\sigma$ among all the beads of the system. To validate the choice of the AMBER force field in combination with the reaction-field method making use of the previous set of parameters, we also performed an atomistic MD simulation using PME for the description of electrostatic interactions, with a dielectric constant $\epsilon = 1$ and a cutoff of 1.0 nm. The comparison is performed in terms of RMSF between the two all-atom trajectories of ADK, as shown in Figure S3. We observed that the trends of fluctuations are consistent with each other, providing comforting evidence that the AMBER model can be safely employed with a reaction field. The SETTLE[64] and RATTLE[65] algorithms for rigid water and rigid bonds containing hydrogen were used. The length of the CANVAS simulations is 500 ns for ADK and 200 ns for each antibody system. All simulations were carried out with GROMACS 2018.[66] We stress here that the usage of an explicit solvent, while guaranteeing the highest level of accuracy of the model in the atomistic region, makes the computational cost of the simulation essentially identical to that of a fully atomistic model. In Table 2 we provide a quantitative comparison of the

**Table 2. Comparison of the Time Performance for ADK Simulations Run on 48 Cores, Single Node, for Different Electrostatic Methods, Interaction Cutoffs, and Resolution[a]**

| method | resolution | cutoff [nm] | performance |
|---|---|---|---|
| reaction field | all-atom | 1.000 | 87.90 ns/day |
| reaction field | all-atom | 1.698 | 32.14 ns/day |
| reaction field | CANVAS | 1.698 | 33.75 ns/day |
| PME | all-atom | 1.000 | 43.04 ns/day |
| PME | all-atom | 1.698 | 19.17 ns/day |

[a]As expected, those employing the PME are about twice as slow as those employing the reaction-field (RF) method for all-atom simulation and cutoff of 1.0 nm. The CANVAS simulation is slightly faster than the all-atom one when using the RF method and a cutoff of 1.698 nm. The latter value corresponds to $2.5\sigma_{\max}$ for the ADK starting configuration when constructing the CANVAS model.

representative structures are placed in a cubic simulation box of 17.9 nm per side and are solvated in a 0.15 M NaCl aqueous solution.

For both ADK and pembrolizumab, the force field employed was Amber99SB-ildn,[43] and the water model was TIP3P.[61] Furthermore, for the sake of assessing the validity of the approach independently of the specific all-atom force field employed, 10 ns long CANVAS simulations of ADK were performed with the Charmm36m force field, using MD software programs GROMACS and LAMMPS; the results of these tests are provided as Supporting Information in Figure S2. CANVAS systems were prepared starting from the representative structures obtained from the atomistic simulations, after energy minimization with the steepest descent algorithm and 100 ps of NVT equilibration. The temperature is kept constant at 300 K by means of the Langevin thermostat.[59] In the NPT production run, the Parrinello–Rahman barostat is employed, as described above. The

performance of 500 ns long ADK simulations run on a 48-cores single node. These show how the CANVAS simulation is slightly faster (about 1.05 times) than the atomistic one when using the reaction-field electrostatic method and same cutoff. Moreover, as expected, the all-atom simulation employing the reaction field is faster—about twice as fast—than the corresponding one when using PME. One of the long-term targets in the development of variable-resolution models is the boost of computational efficiency through the reduction of the number of model particles; here, however, we apply the multiscale representation for the biomolecule alone because the combined usage of multiple-resolution models of the protein *and* of the solvent would lead to ambiguities in the validation and in interpretation of the outcomes. The usage of CANVAS in combination with computationally efficient models of the solvent (e.g., implicit solvent[67,68] or adaptive resolution simulation schemes[32,33,69]) will be the object of future work.

The analysis of fluctuations was performed with the VMD molecular visualization program.[70] In particular, the root-mean-square deviation (RMSD) was computed through the
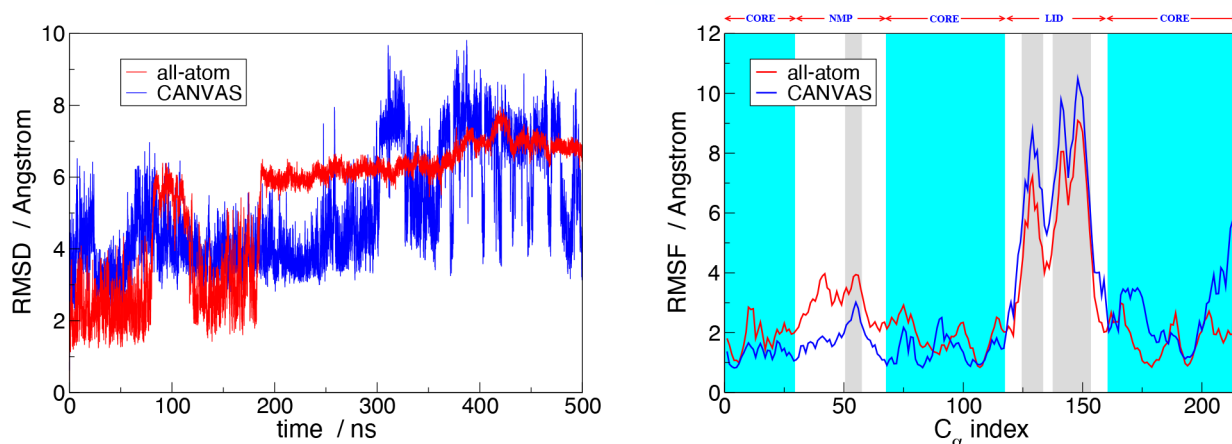
**Figure 6.** (a) RMSD of all ADK $C_\alpha$ comparing the all-atom simulation (red line) and the CANVAS one (blue line). The presence of two different states, one corresponding at about 3 Å and the second one close to 6 Å, are indicative of open and compact conformations, respectively. (b) RMSF for each $C_\alpha$ of ADK from the all-atom simulation (red line) and the CANVAS one (blue line). The cyan area corresponds to the CORE domain, which is described atomistically, whereas the gray and white regions correspond to the parts of the system (LID and NMP domains) modeled in CG and MG, respectively. Videos of the atomistic and CANVAS trajectories are provided on the Zenodo repository 10.5281/zenodo.7225086.

*RMSD Trajectory Tool* considering the sole $C_\alpha$ atoms. The root-mean-square fluctuations (RMSF) were computed by means of an in-house tkl script. The radii of gyration were computed with *gmx gyrate*, whereas the solvent-accessible surface area was computed with *gmx sasa*. The principal component analysis and the calculation of the root-mean-square inner product (RMSIP)[71] between the essential subspaces from atomistic and CANVAS simulations were performed with the Python module MDAnalysis. The calculation of the electrostatic potential was performed with the online adaptive Poisson−Boltzmann solver (APBS)[72] after the creation of an input PQR file that, in the case of the multiscale model, includes the radii and charges as computed with the CANVAS protocol. Protein visualization and rendering was performed with VMD,[70] whereas the plots were created with Xmgrace and Python libraries.

## ■ RESULTS AND DISCUSSION

In this section we compare results from the atomistic and CANVAS simulations for both ADK and pembrolizumb to assess the validity of the proposed multiscale model. In the case of pembrolizumab, the comparison is performed between the six CANVAS simulations and the corresponding ensembles of structurally homogeneous configurations obtained through a clustering of all-atom simulation frames; see Tarenzi et al.[60]

**Adenylate Kinase.** The ADK protein exists in two main conformations, required for the catalytic activity of the enzyme: a fully open one, where the LID and the NMP domains are separated from each other, thus exposing the binding site; and a closed one, which is stabilized by the presence of the substrate and allows for the enzymatic reaction to take place.[73] In the all-atom simulation, ADK samples both the open conformation, which corresponds to the starting structure (Figure 4a), and a more compact one (Figure 4b), where the distance between the LID and NMP arms is substantially reduced. This partially closed conformation of ADK in the apo state was already observed experimentally[74] and in previous MD simulation studies.[75,76] However, we do not observe a complete transition between the open and fully closed states, as expected from the absence of the substrate and from the long time scale of the process (on the order of μs to

ms[77]); indeed, the computed distance between the $C_\alpha$ atoms of residues A55 and V169, previously used to discriminate the two conformational states both in experiments and simulations,[78] is consistent with the open state for the whole duration of the trajectory (Figure S4 in the Supporting Information).

The evolution of the protein between the two aforementioned conformations can be quantified during the simulation in terms of the RMSD of all $C_\alpha$ atoms with respect to the initial frame, which corresponds to the equilibrated structure of ADK in the NPT ensemble (Figure 4a). Because the latter is in the open conformation, higher RMSD values are indicative of closer structures. The resulting plot is shown with a red line in Figure 6a. As expected, two states are clearly visible: one corresponding to 3 Å and the second one around 6 Å. The compact conformation (higher RMSD values) is attained for a few nanoseconds after 80 ns, it reappears subsequently after 200 ns and remains there until the end of the simulation.

Consistent with the previous analysis, the red line of Figure 6b shows the RMSF for each $C_\alpha$, computed with respect to the average structure: we notice that the atoms constituting the protein arms, that is, the LID and NMP domains (indices 118−160 and 30−67) have wider fluctuations with respect to the CORE.

Because the open/closed transition is determined by the relative orientation of LID and NMP with respect to the CORE, the latter is modeled at high resolution in the CANVAS simulation, with the aim of retaining a realistic degree of flexibility of the hinge. In contrast, the LID and the NMP domains are described using two levels of resolution, that is, MG and CG. We recall that all residues whose distance is less than 1 nm with respect to the closest all-atom residues are described as MG to guarantee a *smooth* transition between the highest and lowest levels of resolution. A schematic representation is shown in Figure 4c.

The CANVAS simulation shows two main protein conformations, analogous to the all-atom simulation: the open one, as depicted in Figure 7a, and the compact one as displayed in Figure 7b. The interconversion between the two conformations is monitored, analogous to the fully atomistic simulation, by calculating the RMSD of the $C_\alpha$ atoms (CA$_{at}$,
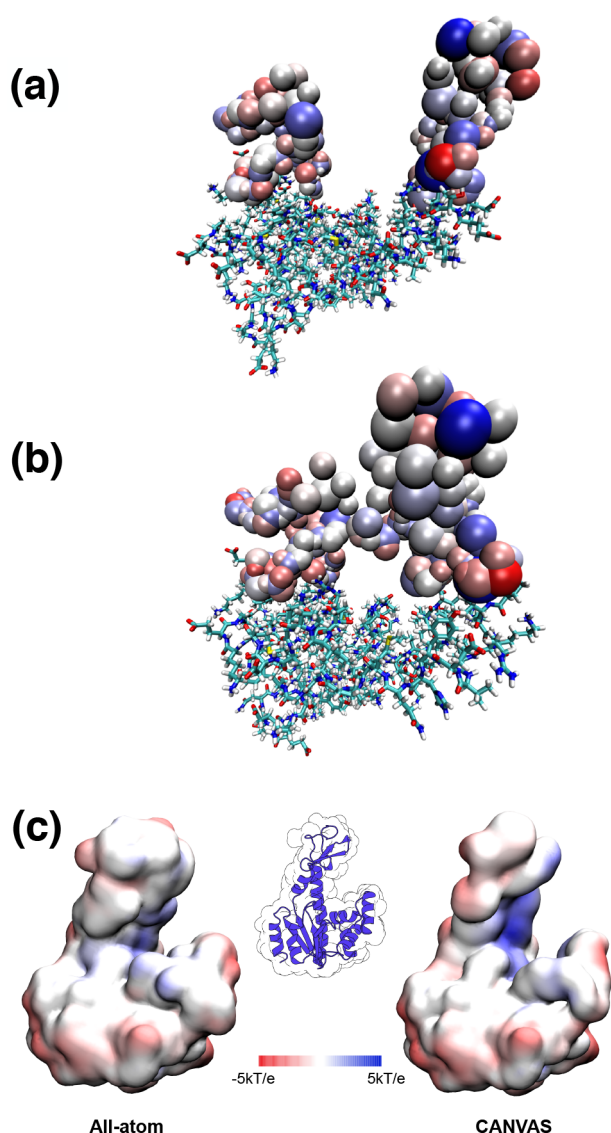
$CA_{mg}$, $CA_{cg}$) with respect to the reference frame. The resulting curve is shown in blue in Figure 6a. The comparison between the all-atom and multiresolution RMSD shows that the CANVAS model reproduces well the conformational changes observed in the fully atomistic system, allowing the protein to transition between the two basins more frequently than the all-atom reference. To assess whether the two sampled states are structurally similar in both simulations, we performed a clustering analysis on the all-atom and CANVAS trajectories, using the RMSD with respect to the starting structure as a distance measure. From the two clusters obtained (corresponding to the fully open and to the compact conformations), the central structures are extracted; representative conformations belonging to the same state are then compared between the atomistic and multiscale cases (Figure S5), and the RMSD value between them was calculated. The resulting RMSD values are 3.7 Å for the open conformations and 5.5 Å for the compact ones; a visual inspection of the representative structures reveals that these deviations are mostly limited to the flexible and disordered regions of the protein, whereas the overall conformational state is the same in the atomistic and multiscale case. Conversely, the comparison of closed and open structures shows larger deviations: the RMSD value between the open atomistic and CANVAS compact representative conformations is 7.4 Å, whereas the RMSD value is 5.8 Å when comparing the compact atomistic and CANVAS open representative conformations. Next, we looked into the fluctuation of each $C_\alpha$ in the all-atom part and each CA bead ($CA_{mg}$, $CA_{cg}$) in the MG and CG ones (whose position is the same for the corresponding $C_\alpha$ atoms in the all-atom representation), as depicted by the blue line in Figure 6b. Also in this case, for both all-atom and lower-resolution regions the fluctuations of $C_\alpha$ atoms are comparable to those from the atomistic simulation. The comparison of fluctuations has also been performed independently on the sets of frames extracted from the atomistic and CANVAS trajectories after the RMSD-based clustering; the resulting RMSF is plotted in Figure S6 of the Supporting Information, and it shows a similar trend in the two cases.

As explained in the description of the model, the values of $Q$, $\sigma$, and $\epsilon$ for each low-resolution bead are different depending on the number and type of atoms that are mapped onto it. Figure 7 shows the two conformations where each CG bead is colored according to its charge, and whose size is based on the $\sigma$ values. The partial charges assigned to each MG and CG bead, in addition to those assigned to each atom by the atomistic force field, were used to compute the electrostatic potential with the APBS.[72] The protein surface, colored



**Figure 7.** CANVAS representations of ADK, where the all-atom region is described in licorice and the MG/CG beads as VdW spheres. The diameter of each bead is given by the value of $\sigma$, whereas its color is dependent on the value of the charge: white spheres are indicative of neutral charge, whereas blue and red beads correspond to positive and negative charges, respectively. (a) shows the open conformation of the protein, whereas (b) shows the more compact one. (c) displays the electrostatic potential calculated with the APBS for the all-atom and CANVAS representations of the starting ADK structure, mapped on the protein surface.
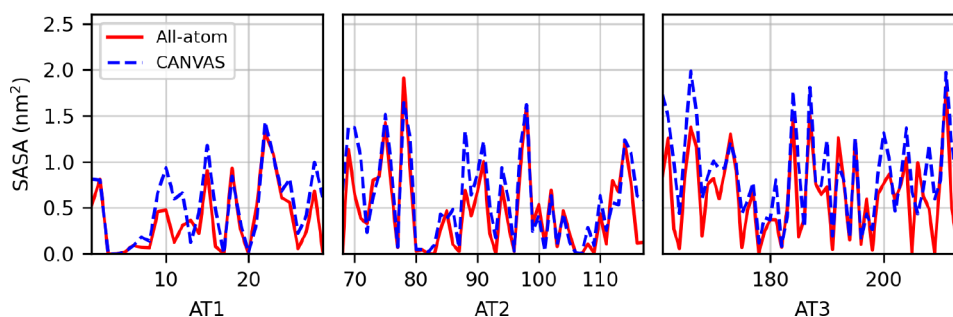


**Figure 8.** Per residue values of SASA, computed for the atomistic region of the ADK. The AT region is composed of three segments of consecutive residues, denoted AT1, AT2, and AT3.
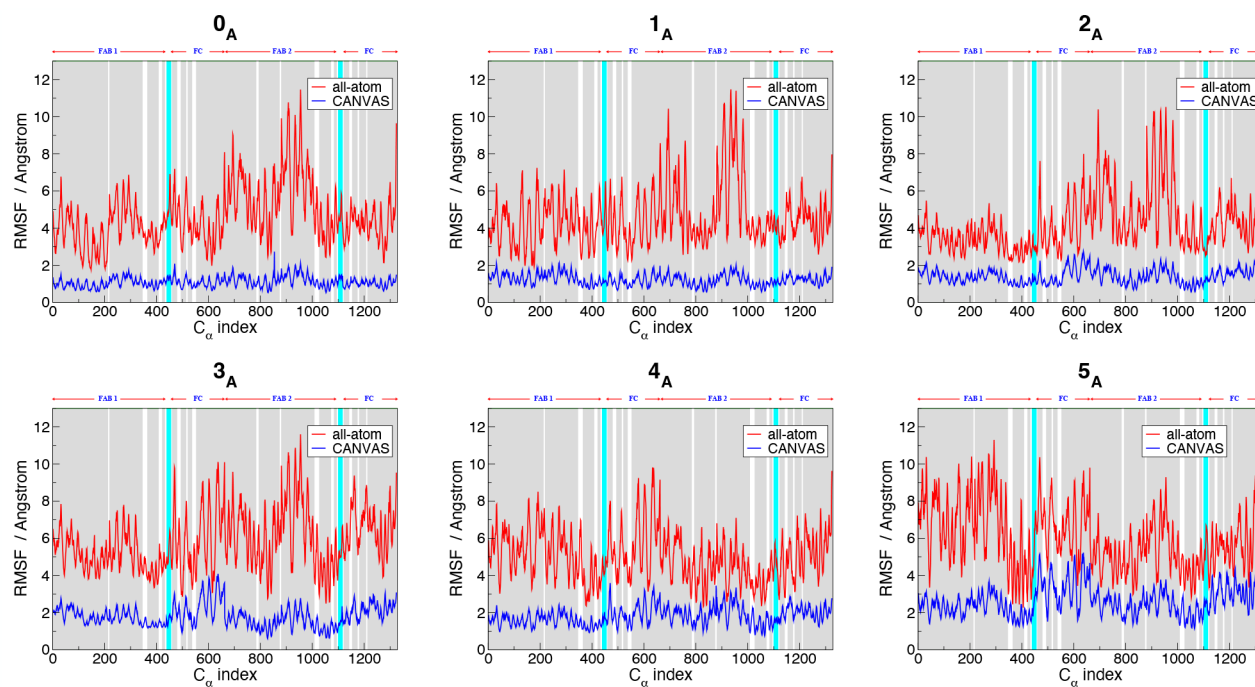
**Figure 9.** RMSF computed on $C_\alpha$ atoms of each apo form of pembrolizumab, for all-atom (red lines) and CANVAS simulations (blue lines). The cyan slabs correspond to the hinge region described atomistically in the CANVAS model, whereas the gray and white regions correspond to the parts of system modeled as CG and MG, respectively. The x-axis corresponds to the $C_\alpha$ indices.

according to the mapped potential, is represented in Figure 7c for both the fully atomistic case and the CANVAS case; the comparison shows that the electrostatic patches are conserved in the multiresolution representation.

To check the accuracy in the description of the AT region in the CANVAS model, we computed the average solvent-accessible surface area (SASA) for each atomistic residue, comparing the results to the values obtained from the atomistic simulation (Figure 8). The results are in good agreement; the slightly larger SASA values for some residues in the CANVAS simulation might be ascribed to the fact that in the fully atomistic case the protein spends a larger portion of the trajectory in the compact state, where the solvent accessibility of a number of residues is reduced.

**Pembrolizumab.** In Tarenzi et al.,[60] four all-atom simulations have been performed starting from the PDB crystallographic structure of the full-length antibody, for a total simulation time of 2 $\mu$s. The antibody conformations sampled from the MD simulations are grouped into six clusters on the basis of their structural similarity. The representative structures of the different clusters are shown in Figure S7 and labeled $0_A$, $1_A$, $2_A$, $3_A$, $4_A$, and $5_A$ according to the increasing value of the protein's average radius of gyration. The conformations differ mainly in the relative orientation of the Fab and Fc domains, which can get in close contact thanks to the flexibility of the hinge region; the latter includes two 18 residue long disordered segments, bridged by two disulfide bonds.

The six representative pembrolizumab structures are taken as starting conformations for six CANVAS simulations. Because the variation in the relative arrangement of Fab1, Fab2, and Fc domains is made possible by the disordered hinge region, the latter is described atomistically, whereas the three large domains are modeled with lower levels of detail. In particular, those residues whose distances are less than 1 nm with respect to the closest fully atomistic ones are described as

MG, whereas the rest is represented as CG. A schematic representation is given in Figure 5b.

Deviations from the starting structure along the simulations are plotted in Figure S8 in terms of RMSD and compared to the average RMSD of the atomistic frames falling within the same conformational cluster. Both atomistic and CANVAS deviations were computed with respect to the same reference structure. For the majority of the conformational clusters, the RMSD values from the CANVAS simulation fall within the error bar of the atomistic reference. However, Figure S8 suggests also that the CANVAS representation of pembrolizumab is slightly more rigid than the fully atomistic case; conversely, the atomistic conformations falling within the same cluster appear more heterogeneous, hence their largest values of RMSD with respect to the representative structures.

The average residue fluctuations were evaluated by computing the RMSF of each $C_\alpha$ in the all-atom part and each CA bead ($CA_{mg}$, $CA_{cg}$) in the MG and CG ones (whose position is the same of the corresponding $C_\alpha$ atoms in the all-atom representation). The analysis of the RMSF plots (Figure 9) shows that, for each cluster, the fluctuations follow the same trend for both all-atom and CANVAS simulations; nonetheless, the RMSD and RMSF values of pembrolizumab in the CANVAS case appear rather low when compared to those of the ADK simulations, where the multiresolution model quantitatively reproduces the atomistic fluctuations. This may be ascribed to the interconnections between distant antibody regions, which take place through an extended network of interdomain correlations within and around the hinge.[60] Indeed, we expect that the differences between atomistic and multiscale fluctuations are due to the particularly small high-resolution region chosen for the pembrolizumab with respect to ADK; in the latter case, ~62% of the residues is described atomistically, whereas in the former case only ~3% of the residues is at high resolution. To test this hypothesis, we

performed an additional 50 ns long CANVAS simulation of pembrolizumab with a larger size of the atomistic region; here, the number of atomistic residues is about 16% of the total. The resulting RMSF (Figure S9) shows that including in the high-resolution region also those Fab and Fc residues that are in contact with the hinge region leads indeed to a better agreement between the all-atom and CANVAS simulations, with respect to the case where the hinge region alone is treated atomistically. In this regard, we stress that the choice of the optimal level and distribution of coarsening to be employed in the construction of a multiple-resolution protein model is a complex and difficult task *per se*;[9,50] nonetheless, CANVAS would represent a powerful instrument to investigate precisely this aspect, in that it allows a simple parametrization of the model and the subsequent study of the optimal resolution modulation required to correctly and quantitatively reproduce specific system features.

Residue fluctuations are further investigated by computing the linear correlation between the RMSF of $C_\alpha$ atoms of fully atomistic simulation and the CANVAS one for each case. The latter is given by the calculation of the Pearson coefficient[79] $\rho$, as reported in the scatter plots of Figure S10. All clusters show satisfactory results, with good RMSF correlations ($\rho \sim 0.7$); moreover, an excellent correlation is found in cluster $0_A$ ($\rho \sim 0.87$). To gain additional information about the latter result, we also calculated the *cross* Pearson coefficient $\rho_{XY}$ between states and models as summarized in Table 3. $X$ and $Y$ take

**Table 3. *Cross* Pearson Coefficients $\rho_{XY}$ between States and Models[a]**

| $\rho_{XY}$ | | CANVAS | | | | | |
|---|---|---|---|---|---|---|---|
| | | $0_A$ | $1_A$ | $2_A$ | $3_A$ | $4_A$ | $5_A$ |
| all-atom | $0_A$ | **0.87** | 0.58 | 0.50 | 0.27 | 0.63 | 0.35 |
| | $1_A$ | 0.81 | **0.72** | 0.63 | 0.42 | 0.72 | 0.47 |
| | $2_A$ | 0.73 | 0.60 | **0.71** | 0.50 | 0.76 | 0.52 |
| | $3_A$ | 0.78 | 0.74 | 0.84 | **0.74** | 0.91 | 0.74 |
| | $4_A$ | 0.49 | 0.76 | 0.77 | 0.80 | **0.69** | 0.76 |
| | $5_A$ | 0.55 | 0.83 | 0.64 | 0.66 | 0.61 | **0.68** |

[a]$X$ and $Y$ refer to the all-atom and CANVAS model, respectively; both indices correspond to the conformation from which the simulations start ($0_A$, $1_A$, $2_A$, $3_A$, $4_A$, $5_A$). On the diagonal, the higher the index $XX$, the less compact the antibody conformation and the lower the value of $\rho_{XX}$.

values in [0, 5], corresponding to the various clusters, whereas each variable is associated with the all-atom ($X$) and CANVAS ($Y$) model. For instance, the value of $\rho_{25}$ corresponds to the Pearson correlation coefficient between the RMSF of $C_\alpha$ atoms for the $2_A$ cluster at fully atomistic resolution versus the RMSF of the $5_A$ state simulated with CANVAS. Diagonal elements $\rho_{XX}$ measure the correlation between $C_\alpha$ atoms of a fully atomistic simulation and corresponding CANVAS one for the same cluster. Such values, already displayed in the scatter plot of Figure S5, are highlighted in bold in the table. One can notice that the higher the cluster index, the lower the value of the Pearson correlation coefficient ($\rho_{00} = 0.87$, $\rho_{55} = 0.68$). Because the clusters are ranked by increasing radius of gyration (or equivalently decreasing compactness), the reason for this

trend can be ascribed to the fact that the CANVAS model of a more open structure has more freedom to explore conformations further and further away from the reference.

Furthermore, Table 3 shows that the Pearson coefficient is not systematically higher when comparing simulations starting in the same conformational basin. This is not a fully unexpected result; indeed, CANVAS simulations were started from the given initial conformations that in this case are also representative of specific groups of structures sampled in an all-atom MD trajectory, but this gives no guarantee that the whole run will explore the same cluster. This is true in general, even in the case of a fully atomistic model: a new all-atom simulation starting from a representative frame of one conformational cluster might, because of its stochastic nature, diffuse toward another cluster and hence show a fluctuation pattern closer to what is observed in a different set of frames. In the case under examination, additionally, the CANVAS model consists of a distinct structural representation and interaction force field with respect to the all-atom reference; hence, even if the simulation starts from a representative frame of the all-atom cluster, this frame will not be an equilibrium, representative configuration of the conformational space that would be sampled by the CANVAS model. What we observe in our analysis is that, in spite of this *caveat*, the CANVAS simulations show a remarkable structural overlap between the conformations sampled starting from a given frame and the all-atom cluster they represent, as can be seen from the CANVAS simulation trajectories provided as Supporting Information; as for the pattern of fluctuations, the strong intracluster consistency is paired by a non-negligible, and sometimes higher, correlation with different reference clusters, whose appearance is thus not unexpected nor surprising. Hence, although further work is certainly needed to perfect the agreement between the all-atom model and its multiple-resolution counterpart, the strong structural consistency and the highly correlated RMSF patterns of CANVAS run against their corresponding references support the idea that the model can already capture rather fine details of the molecule's dynamics.

Further analyses were performed to differentiate the dynamics of all-atom and CANVAS simulations for different clusters. Specifically, we have examined the fluctuation correlations distinguishing residues by their level of resolution (AT, MG, CG) and the domain they belong to (FAB1, FAB2, FC). This analysis highlights other salient properties of the fluctuations of the antibody: (1) *Scatter plot with points colored based on resolution* (AT, MG, CG) in Figure S11. The all-atom part is very small ($\sim$3%); hence, the corresponding value of the Pearson coefficient is not indicative. Conversely, the medium- and coarse-grained parts make up for most of the antibody ($\sim$97%); hence, the value of $\rho_{MG}$ and $\rho_{CG}$ is closer to the one of the full system (dash black line). (2) *Scatter plot with points colored based on the domain partition* (FAB1, FAB2, FC) in Figure S12. Each domain produces a linear pattern in the plot, and the values of the corresponding Pearson coefficients is close to unity. It is worth noticing that, in some of the clusters, the RMSF of the two Fab domains indicates differences in flexibility between the all-atom and the CANVAS models. Specifically, although the overall correlation degree is rather high, the slope of this correlation is different between the two domains. A close inspection reveals indeed that the two heavy chains present a different arrangement of the hinge and the $CH_2$ domain, as already noted elsewhere,[54,60] thus returning a

model whose Fab domains have different interactions and, therefore, different flexibilities.

These analyses provide an additional confirmation that the RMSF correlation between all-atom simulation and the CANVAS one is rather high, although more sophisticated and less straightforward than expected; this, in hindsight, is a reasonable behavior for a system whose structural and dynamical modules are represented, modeled, and simulated with distinct levels of resolution.

The conformational dynamics of the system was further inspected by computing the RMSIP between the essential subspaces given by the first $n$ normal modes of the covariance from the atomistic and CANVAS simulations, with $n$ ranging from 1 to the first 10 modes. A value of 0 indicates that the two mode subspaces are orthogonal, whereas 1 indicates that they are identical.[71] Figure S13 shows that less than 5 modes are enough to attain a very good overlap (RMSIP > 0.8) for all clusters.

We compared the similarity of the structures sampled in the atomistic and CANVAS simulations through the calculation of the radius of gyration (Figure S14). The values present small deviations, with the largest discrepancy of 1.3 Å observed in cluster $3_A$; however, in all cases, the radius of gyration from the all-atom simulations is slightly larger than that from the multiscale case, arguably because the steric effects of the side chains cannot be perfectly matched in the very coarse representation employed here, where only the $C_\alpha$ or backbone atoms are retained for more than 97% of the residues.

As previously done for ADK, the electrostatic potential of the Fab1 domain at MG/CG resolution has been computed for the antibody Fab, on the basis of the partial charges assigned to each bead in the CANVAS model (Figure 10). The comparison between the all-atom and low-resolution case shows a good similarity.
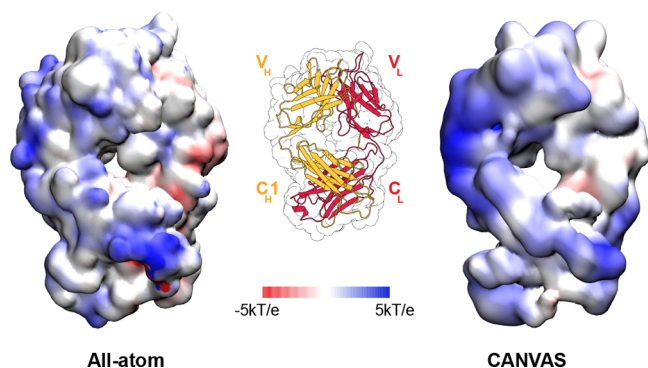


**Figure 10.** Electrostatic potential calculated with the APBS for the all-atom and CANVAS representations of pembrolizumab Fab1, mapped on the protein surface.

The average SASA was computed along the trajectory for each residue of the atomistic region, namely, the two hinge segments (Figure 11). The comparison between the SASA values computed from the all-atom and multiscale simulations, performed for each conformational cluster, shows a very good agreement. The CANVAS model proves able to accurately reproduce the solvent exposure of the atomistic residues in relation to the conformational properties of the fully atomistic system.

## CONCLUSIONS

In this work we introduced the CANVAS model for the multiscale molecular dynamics simulation of proteins. The model couples different levels of detail within the same protein representation, ranging from a fully atomistic description to a coarse one, for example, one bead per amino acid (as in the case studies discussed here) or even lower levels of resolution. CANVAS allows a smooth transition between these resolutions by including regions at intermediate levels of detail. Importantly, the nonbonded components of the interaction potential are taken into account at all resolution levels by assigning to each CG bead the average properties—including charge, size, and dispersion energy—of the atoms that are mapped onto it. This property enables, in principle, the application of CANVAS for the simulation of large, multimeric protein complexes, where also the CG resolution can be used to model realistic molecular interfaces. This application will be explored in future works.

Here, we have tested the CANVAS model on two systems of very different size and conformational dynamics, namely, the enzyme adenylate kinase and the therapeutic antibody pembrolizumab. To validate the model, we performed a comparison among the properties extracted from the fully atomistic and multiscale simulations, in terms of residue fluctuations, large-scale dynamics, solvent exposure, and electrostatic properties; in all cases, the CANVAS model results are in good agreement with the all-atom reference.

The variable-resolution modeling approach presented here achieves two key goals: first, it demonstrates that a sensible modulation of the resolution can be employed to construct models of large molecules whose behavior is the same of, or quantitatively consistent with, that of a reference all-atom model of the system; second, it enables the rapid, practical construction of tailored low-resolution models of such molecules with minimal information and no reference simulations. The possible applications of these models cover a broad spectrum; we here stress those that appear most promising to us, namely, the exploration of the conformational space of molecules whose structure is known with low resolution only, or the characterization of the structure—dynamics—function relation by means of the systematic modulation of the resolution throughout the structure. An additional future application is the efficient calculation of binding free energies, employing an atomistic accuracy only in the active and/or allosteric sites. The relevance of taking into account distant protein domains within the simulation setup has been proven in various cases, as for example specifically observed in the case of antigen—antibody binding affinities;[60,80] the possibility of keeping a simplified description of the vast majority of the molecule thus represents an advantage with respect to the alternative approach of simulating only the protein domain involved in the binding. In addition, we can expect that the impact on entropy due to the reduction of the number of degrees of freedom is similar, and therefore does not affect the result, if the aim is to compute relative binding free energy among a set of similar ligands, where the mapping of the protein is kept the same. All the above-mentioned applications, which involve the usage of the CANVAS model in combination with efficient simulation methods for the solvent (e.g., multiple-time-step,[81,82] implicit solvent,[67,68] or adaptive resolution simulation methods[32,33,69]) are currently under
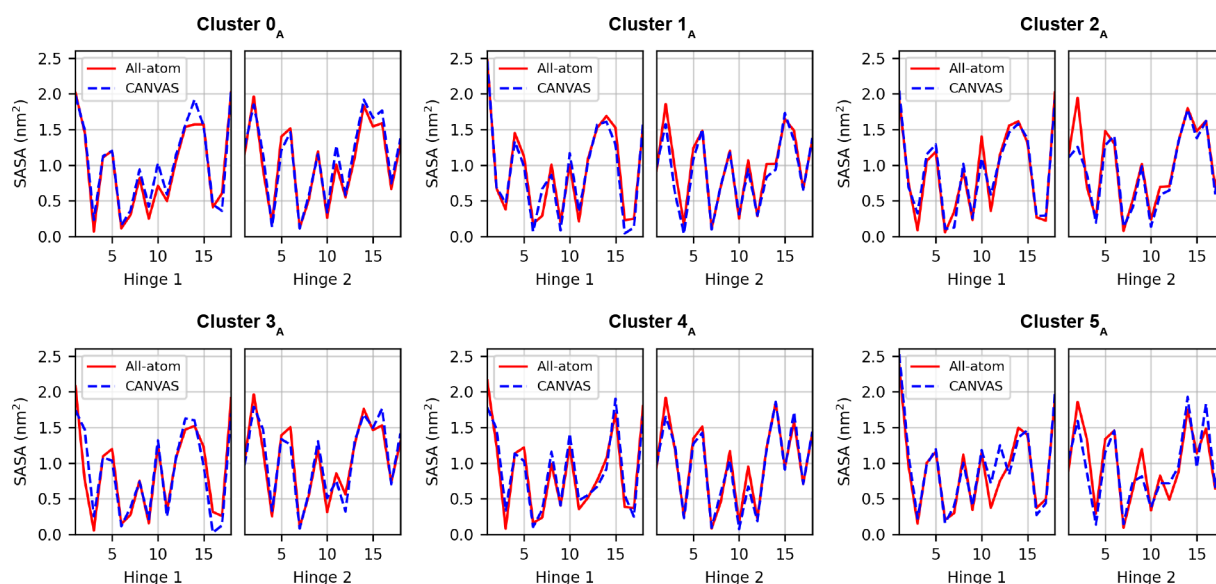
**Figure 11.** Per residue values of SASA, computed for the atomistic region of the antibody in each conformational basin. Hinge 1 belongs to chain B, whereas hinge 2 belongs to chain D.

development and pave the way to a novel approach to computer-aided molecular biochemistry.

## ■ ASSOCIATED CONTENT

### Data Availability Statement

The CANVAS software is available for download at https://github.com/potestiolab/canvas, including the manual and tutorials. The raw data produced and analyzed in this work are freely available on the Zenodo repository 10.5281/zenodo.7225086.

### Ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.2c01311.

> Derivation of the elastic constants $k_{nb}$; RMSD and RMSF of ADK using GROMACS/LAMMPS and Charmm/Amber force fields; RMSF from atomistic simulations of ADK using reaction field or PME; distance between $C_\alpha$ atoms of residues A55 and V169 in the all-atom simulation of ADK; alignment of atomistic and CANVAS representative structures, in open and closed conformations; RMSF of ADK from atomistic and CANVAS simulations, in open and closed conformations; representative structures of pembrolizumab antibody; RMSD from pembrolizumab simulations; RMSF from pembrolizumab simulations with different sizes of the atomistic region; scatter plots of RMSF for each pembrolizumab cluster, and calculation of Pearson coefficient; scatter plots colored on the basis of the resolution and of the structural domain; RMSIP between the essential subspaces computed from the atomistic and CANVAS simulations; average radius of gyration for each conformational cluster of the antibody (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Author

**Raffaello Potestio** − *Department of Physics, University of Trento, I-38123 Trento, Italy; INFN-TIFPA, Trento Institute for Fundamental Physics and Applications, I-38123 Trento, Italy;* ⓞ orcid.org/0000-0001-6408-9380;

Phone: +39 0461 282912; Email: raffaello.potestio@unitn.it

### Authors

**Raffaele Fiorentini** − *Department of Physics, University of Trento, I-38123 Trento, Italy; INFN-TIFPA, Trento Institute for Fundamental Physics and Applications, I-38123 Trento, Italy;* ⓞ orcid.org/0000-0002-7304-7913

**Thomas Tarenzi** − *Department of Physics, University of Trento, I-38123 Trento, Italy; INFN-TIFPA, Trento Institute for Fundamental Physics and Applications, I-38123 Trento, Italy;* ⓞ orcid.org/0000-0002-5575-7452

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jcim.2c01311

### Author Contributions

R.P. designed the project; R.F. developed the software and ran the simulations; R.F. and T.T. performed the analyses. All authors contributed to the interpretation of the results and to the writing of the manuscript.

### Notes

The authors declare no competing financial interest.

## ■ REFERENCES

(1) Stone, J. E.; Hardy, D. J.; Ufimtsev, I. S.; Schulten, K. GPU-accelerated molecular modeling coming of age. *Journal of Molecular Graphics and Modelling* **2010**, *29*, 116−125.

(2) Lazim, R.; Suh, D.; Choi, S. Advances in molecular dynamics simulations and enhanced sampling methods for the study of protein systems. *International journal of molecular sciences* **2020**, *21*, 6339.

(3) Shaw, D. E.; Adams, P. J.; Azaria, A.; Bank, J. A.; Batson, B.; Bell, A.; Bergdorf, M.; Bhatt, J.; Butts, J. A.; Correia, T.; Dirks, R. M.; Dror, R. O.; Eastwood, M. P.; Edwards, B.; Even, A.; Feldmann, P.; Fenn, M.; Fenton, C. H.; Forte, A.; Gagliardo, J.; Gill, G.; Gorlatova, M.; Greskamp, B.; Grossman, J.; Gullingsrud, J.; Harper, A.; Hasenplaugh, W.; Heily, M.; Heshmat, B. C.; Hunt, J.; Ierardi, D. J.; Iserovich, L.; Jackson, B. L.; Johnson, N. P.; Kirk, M. M.; Klepeis, J. L.; Kuskin, J. S.; Mackenzie, K. M.; Mader, R. J.; McGowen, R.; McLaughlin, A.; Moraes, M. A.; Nasr, M. H.; Nociolo, L. J.; O'Donnell, L.; Parker, A.; Peticolas, J. L.; Pocina, G.; Predescu, C.; Quan, T.; Salmon, J. K.; Schwink, C.; Shim, K. S.; Siddique, N.; Spengler, J.; Szalay, T.; Tabladillo, R.; Tartler, R.; Taube, A. G.; Theobald, M.; Towles, B.; Vick, W.; Wang, S. C.; Wazlowski, M.; Weingarten, M. J.; Williams, J. M.; Yuh, K. A. Anton 3: Twenty Microseconds of Molecular Dynamics Simulation before Lunch. *SC '21: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*; Association for Computing Machinery: New York, 2021.

(4) Peter, C.; Kremer, K. Multiscale simulation of soft matter systems: from the atomistic to the coarse-grained level and back. *Soft Matter* **2009**, *5*, 4357−4366.

(5) Kamerlin, S. C.; Vicatos, S.; Dryga, A.; Warshel, A. Coarse-Grained (Multiscale) Simulations in Studies of Biophysical and Chemical Systems. *Annu. Rev. Phys. Chem.* **2011**, *62*, 41−64. PMID: 21034218.

(6) Takada, S. Coarse-grained molecular simulations of large biomolecules. *Curr. Opin. Struct. Biol.* **2012**, *22*, 130−137. Theory and simulation/macromolecular assemblages.

(7) Riniker, S.; Allison, J. R.; van Gunsteren, W. F. On developing coarse-grained models for biomolecular simulation: a review. *Phys. Chem. Chem. Phys.* **2012**, *14*, 12423−12430.

(8) Noid, W. G. Perspective: Coarse-grained models for biomolecular systems. *J. Chem. Phys.* **2013**, *139*, 090901.

(9) Giulini, M.; Rigoli, M.; Mattiotti, G.; Menichetti, R.; Tarenzi, T.; Fiorentini, R.; Potestio, R. From system modeling to system analysis: The impact of resolution level and resolution distribution in the computer-aided investigation of biomolecules. *Frontiers in Molecular Biosciences* **2021**, *8*. DOI: 10.3389/fmolb.2021.676976

(10) Souza, P. C.; Thallmair, S.; Conflitti, P.; Ramírez-Palacios, C.; Alessandri, R.; Raniolo, S.; Limongelli, V.; Marrink, S. J. Protein-ligand binding with the coarse-grained Martini model. *Nat. Commun.* **2020**, *11*, 3714.

(11) Marrink, S. J.; Corradi, V.; Souza, P. C.; Ingolfsson, H. I.; Tieleman, D. P.; Sansom, M. S. Computational modeling of realistic cell membranes. *Chem. Rev.* **2019**, *119*, 6184−6226.

(12) Marrink, S. J.; Monticelli, L.; Melo, M. N.; Alessandri, R.; Tieleman, D. P.; Souza, P. C. Two decades of Martini: Better beads, broader scope. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2022**, *13*, e1620.

(13) Machado, M. R.; Barrera, E. E.; Klein, F.; Sóñora, M.; Silva, S.; Pantano, S. The SIRAH 2.0 force field: altius, fortius, citius. *J. Chem. Theory Comput.* **2019**, *15*, 2719−2733.

(14) Patel, J. S.; Ytreberg, F. M. Fast calculation of protein-protein binding free energies using umbrella sampling with a coarse-grained model. *J. Chem. Theory Comput.* **2018**, *14*, 991−997.

(15) Davtyan, A.; Schafer, N. P.; Zheng, W.; Clementi, C.; Wolynes, P. G.; Papoian, G. A. AWSEM-MD: protein structure prediction using coarse-grained physical potentials and bioinformatically based local structure biasing. *J. Phys. Chem. B* **2012**, *116*, 8494−8503.

(16) Basdevant, N.; Borgis, D.; Ha-Duong, T. A coarse-grained protein- protein potential derived from an all-atom force field. *J. Phys. Chem. B* **2007**, *111*, 9390−9399.

(17) Rzepiela, A. J.; Louhivuori, M.; Peter, C.; Marrink, S. J. Hybrid simulations: combining atomistic and coarse-grained force fields using virtual sites. *Phys. Chem. Chem. Phys.* **2011**, *13*, 10437−10448.

(18) Wassenaar, T. A.; Ingolfsson, H. I.; Prieß, M.; Marrink, S. J.; Schäffer, L. V. Mixing MARTINI: electrostatic coupling in hybrid atomistic-coarse-grained biomolecular simulations. *J. Phys. Chem. B* **2013**, *117*, 3516−3530.

(19) Zavadlav, J.; Sablić, J.; Podgornik, R.; Praprotnik, M. Open-boundary molecular dynamics of a DNA molecule in a hybrid explicit/implicit salt solution. *Biophys. J.* **2018**, *114*, 2352−2362.

(20) Zavadlav, J.; Marrink, S. J.; Praprotnik, M. Multiscale simulation of protein hydration using the SWINGER dynamical clustering algorithm. *J. Chem. Theory Comput.* **2018**, *14*, 1754−1761.

(21) Han, W.; Wan, C.-K.; Jiang, F.; Wu, Y.-D. PACE force field for protein simulations. 1. Full parameterization of version 1 and verification. *J. Chem. Theory Comput.* **2010**, *6*, 3373−3389.

(22) Han, W.; Schulten, K. Further optimization of a hybrid united-atom and coarse-grained force field for folding simulations: improved backbone hydration and interactions between charged side chains. *J. Chem. Theory Comput.* **2012**, *8*, 4413−4424.

(23) Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198−1229.

(24) Wang, B.; Truhlar, D. G. Combined Quantum Mechanical and Molecular Mechanical Methods for Calculating Potential Energy Surfaces: Tuned and Balanced Redistributed-Charge Algorithm. *J. Chem. Theory Comput.* **2010**, *6*, 359−369. PMID: 26617295.

(25) Wallrapp, F. H.; Guallar, V. Mixed quantum mechanics and molecular mechanics methods: Looking inside proteins. *WIREs Computational Molecular Science* **2011**, *1*, 315−322.

(26) Cui, Q.; Pal, T.; Xie, L. Biomolecular QM/MM Simulations: What Are Some of the Burning Issues? *J. Phys. Chem. B* **2021**, *125*, 689−702.

(27) Vennelakanti, V.; Nazemi, A.; Mehmood, R.; Steeves, A. H.; Kulik, H. J. Harder, better, faster, stronger: large-scale QM and QM/MM for predictive modeling in enzymes and proteins. *Curr. Opin. Struct. Biol.* **2022**, *72*, 9−17.

(28) Neri, M.; Anselmi, C.; Cascella, M.; Maritan, A.; Carloni, P. Coarse-grained model of proteins incorporating atomistic detail of the active site. *Physical review letters* **2005**, *95*, 218102.

(29) Schneider, J.; Korshunova, K.; Musiani, F.; Alfonso-Prieto, M.; Giorgetti, A.; Carloni, P. Predicting ligand binding poses for low-resolution membrane protein models: Perspectives from multiscale simulations. *Biochemical and biophysical research communications* **2018**, *498*, 366−374.

(30) Leguèbe, M.; Nguyen, C.; Capece, L.; Hoang, Z.; Giorgetti, A.; Carloni, P. Hybrid Molecular Mechanics/Coarse-Grained Simulations for Structural Prediction of G-Protein Coupled Receptor/Ligand Complexes. *PLoS One* **2012**, *7*, e47332.

(31) Tarenzi, T.; Calandrini, V.; Potestio, R.; Carloni, P. Open-boundary molecular mechanics/coarse-grained framework for simulations of low-resolution G-protein-coupled receptor−ligand complexes. *J. Chem. Theory Comput.* **2019**, *15*, 2101−2109.

(32) Potestio, R.; Fritsch, S.; Español, P.; Delgado-Buscalioni, R.; Kremer, K.; Everaers, R.; Donadio, D. Hamiltonian Adaptive Resolution Simulation for Molecular Liquids. *Phys. Rev. Lett.* **2013**, *110*, 108301.

(33) Tarenzi, T.; Calandrini, V.; Potestio, R.; Giorgetti, A.; Carloni, P. Open boundary simulations of proteins and their hydration shells by Hamiltonian adaptive resolution scheme. *J. Chem. Theory Comput.* **2017**, *13*, 5647−5657.

(34) Fogarty, A. C.; Potestio, R.; Kremer, K. A multi-resolution model to capture both global fluctuations of an enzyme and molecular recognition in the ligand-binding site. *Proteins: Struct., Funct., and Bioinf.* **2016**, *84*, 1902−1913.

(35) Fiorentini, R.; Kremer, K.; Potestio, R. Ligand-protein interactions in lysozyme investigated through a dual-resolution model. *Proteins: Struct., Funct., Bioinf.* **2020**, *88*, 1351−1360.

(36) Tirion, M. M. Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Phys. Rev. Lett.* **1996**, *77*, 1905−1908.

(37) Alfonso-Prieto, M.; Giorgetti, A.; Carloni, P. Multiscale simulations on human Frizzled and Taste2 GPCRs. *Curr. Opin. Struct. Biol.* **2019**, *55*, 8−16.

(38) Alfonso-Prieto, M.; Navarini, L.; Carloni, P. Understanding ligand binding to G-protein coupled receptors using multiscale simulations. *Frontiers in molecular biosciences* **2019**, *6*, 29.

(39) Fierro, F.; Giorgetti, A.; Carloni, P.; Meyerhof, W.; Alfonso-Prieto, M. Dual binding mode of "bitter sugars" to their human bitter taste receptor target. *Sci. Rep.* **2019**, *9*, 8437.

(40) Spiga, E.; Alemani, D.; Degiacomi, M. T.; Cascella, M.; Dal Peraro, M. Electrostatic-consistent coarse-grained potentials for molecular simulations of proteins. *J. Chem. Theory Comput.* **2013**, *9*, 3515−3526.

(41) Darré, L.; Machado, M. R.; Brandner, A. F.; González, H. C.; Ferreira, S.; Pantano, S. SIRAH: a structurally unbiased coarse-grained force field for proteins with aqueous solvation and long-range electrostatics. *J. Chem. Theory Comput.* **2015**, *11*, 723−739.

(42) Souza, P. C. T.; Alessandri, R.; Barnoud, J.; Thallmair, S.; Faustino, I.; Grünewald, F.; Patmanidis, I.; Abdizadeh, H.; Bruininks, B. M. H.; Wassenaar, T. A. A.; Kroon, P. C.; Melcr, J.; Nieto, V.; Corradi, V.; Khan, H. M.; Domanski, J.; Javanainen, M.; Martinez-Seara, H.; Reuter, N.; Best, R. B.; Vattulainen, I.; Monticelli, L.; Periole, X.; Tieleman, D. P.; de Vries, A. H.; Marrink, S. J. Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nat. Methods* **2021**, *18*, 382−388.

(43) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1950−1958.

(44) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; De Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, *14*, 71−73.

(45) Potestio, R.; Pontiggia, F.; Micheletti, C. Coarse-grained description of protein internal dynamics: an optimal strategy for decomposing proteins in rigid subunits. *Biophysical journal* **2009**, *96*, 4993−5002.

(46) Bernhard, S.; Noe, F. Optimal identification of semi-rigid domains in macromolecules from molecular dynamics simulation. *PloS one* **2010**, *5*, e10491.

(47) Burrough, P. A.; McDonnell, R. A.; Lloyd, C. D. *Principles of Geographical Information Systems*, 3rd ed.; Oxford, 2015.

(48) Longley, P. A.; Goodchild, M.; Maguire, D. J.; Rhind, D. W. *Geographic Information Systems and Science*, 3rd ed.; Wiley Publishing, 2010.

(49) Sen, Z. *Spatial Modeling Principles in Earth Sciences*, 2nd ed.; Springer, 2016.

(50) Diggins, P., IV; Liu, C.; Deserno, M.; Potestio, R. Optimal coarse-grained site selection in elastic network models of biomolecules. *J. Chem. Theory Comput.* **2019**, *15*, 648−664.

(51) Dzeja, P.; Terzic, A. Adenylate Kinase and AMP Signaling Networks: Metabolic Monitoring, Signal Communication and Body Energy Sensing. *International Journal of Molecular Sciences* **2009**, *10*, 1729−1772.

(52) Formoso, E.; Limongelli, V.; Parrinello, M. Energetics and Structural Characterization of the large-scale Functional Motion of Adenylate Kinase. *Sci. Rep.* **2015**, *5*, 8425.

(53) Müller, C.; Schlauderer, G.; Reinstein, J.; Schulz, G. Adenylate kinase motions during catalysis: an energetic counterweight balancing substrate binding. *Structure* **1996**, *4*, 147−156.

(54) Scapin, G.; Yang, X.; Prosise, W. W.; McCoy, M.; Reichert, P.; Johnston, J. M.; Kashi, R. S.; Strickland, C. Structure of full-length human anti-PD1 therapeutic IgG4 antibody pembrolizumab. *Nature structural & molecular biology* **2015**, *22*, 953−958.

(55) Ionescu, M. I. Adenylate kinase: a ubiquitous enzyme correlated with medical conditions. *Protein Journal* **2019**, *38*, 120−133.

(56) Ivashko, I. N.; Kolesar, J. M. Pembrolizumab and nivolumab: PD-1 inhibitors for advanced melanoma. *American journal of health-system pharmacy* **2016**, *73*, 193−201.

(57) Chen, Y.; Pei, Y.; Luo, J.; Huang, Z.; Yu, J.; Meng, X. Looking for the optimal PD-1/PD-L1 inhibitor in cancer treatment: a comparison in basic structure, function, and clinical practice. *Frontiers in Immunology* **2020**, *11*, 1088.

(58) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.

(59) Goga, N.; Rzepiela, A. J.; De Vries, A. H.; Marrink, S. J.; Berendsen, H. J. C. Efficient algorithms for Langevin and DPD dynamics. *Chem. Theory Comput.* **2012**, *8* (10), 3637−3649.

(60) Tarenzi, T.; Rigoli, M.; Potestio, R. Communication pathways bridge local and global conformations in an IgG4 antibody. *Sci. Rep.* **2021**, *11*, 23197.

(61) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.

(62) Barker, J.; Watts, R. Monte Carlo studies of the dielectric properties of water-like models. *Mol. Phys.* **1973**, *26*, 789−792.

(63) Watts, R. Monte Carlo studies of liquid water. *Mol. Phys.* **1974**, *28*, 1069−1083.

(64) Miyamoto, S.; Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **1992**, *13*, 952−962.

(65) Andersen, H. C. Rattle: A "velocity" version of the shake algorithm for molecular dynamics calculations. *J. Comput. Phys.* **1983**, *52*, 24−34.

(66) Abraham, M.; van der Spoel, D.; Lindahl, E.; Hess, B.; the GROMACS development team. GROMACS User Manual version 2018.2018; http://www.gromacs.org (accessed Feb 3, 2023).

(67) Onufriev, A. V.; Case, D. A. Generalized Born implicit solvent models for biomolecules. *Annual review of biophysics* **2019**, *48*, 275.

(68) Chen, Y.; Krämer, A.; Charron, N. E.; Husic, B. E.; Clementi, C.; Noé, F. Machine learning implicit solvation for molecular dynamics. *J. Chem. Phys.* **2021**, *155*, 084101.

(69) Praprotnik, M.; Matysiak, S.; Delle Site, L.; Kremer, K.; Clementi, C. Adaptive resolution simulation of liquid water. *J. Phys.: Condens. Matter* **2007**, *19*, 292201.

(70) Humphrey, W.; Dalke, A.; Schulten, K. VMD − Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.

(71) Amadei, A.; Ceruso, M. A.; Di Nola, A. On the convergence of the conformational coordinates basis set obtained by the essential dynamics analysis of proteins' molecular dynamics simulations. *Proteins: Struct., Funct., Bioinf.* **1999**, *36*, 419−424.

(72) Jurrus, E.; Engel, D.; Star, K.; Monson, K.; Brandi, J.; Felberg, L. E.; Brookes, D. H.; Wilson, L.; Chen, J.; Liles, K.; Chun, M.; Li, P.; Gohara, D. W.; Dolinsky, T.; Konecny, R.; Koes, D. R.; Nielsen, J. E.; Head-Gordon, T.; Geng, W.; Krasny, R.; Wei, G.-W.; Holst, M. J.; McCammon, J. A.; Baker, N. A. Improvements to the APBS biomolecular solvation software suite. *Protein Sci.* **2018**, *27*, 112−128.

(73) Onuk, E.; Badger, J.; Wang, Y. J.; Bardhan, J.; Chishti, Y.; Akcakaya, M.; Brooks, D. H.; Erdogmus, D.; Minh, D. D.; Makowski, L. Effects of catalytic action and ligand binding on conformational ensembles of adenylate kinase. *Biochemistry* **2017**, *56*, 4559−4567.

(74) Henzler-Wildman, K. A.; Thai, V.; Lei, M.; Ott, M.; Wolf-Watz, M.; Fenn, T.; Pozharski, E.; Wilson, M. A.; Petsko, G. A.; Karplus, M.; Huebner, C. G.; Kern, D. Intrinsic motions along an enzymatic reaction trajectory. *Nature* **2007**, *450*, 838−U13.

(75) Li, D.; Liu, M. S.; Ji, B. Mapping the dynamics landscape of conformational transitions in enzyme: the adenylate kinase case. *Biophysical journal* **2015**, *109*, 647−660.

(76) Wang, J.; Peng, C.; Yu, Y.; Chen, Z.; Xu, Z.; Cai, T.; Shao, Q.; Shi, J.; Zhu, W. Exploring conformational change of adenylate kinase by replica exchange molecular dynamic simulation. *Biophysical journal* **2020**, *118*, 1009−1018.

O

(77) Zheng, Y.; Cui, Q. Multiple pathways and time scales for conformational transitions in apo-adenylate kinase. *J. Chem. Theory Comput.* **2018**, *14*, 1716−1726.

(78) Whitford, P. C.; Miyashita, O.; Levy, Y.; Onuchic, J. N. Conformational transitions of adenylate kinase: switching by cracking. *Journal of molecular biology* **2007**, *366*, 1661−1671.

(79) Benesty, J.; Chen, J.; Huang, Y.; Cohen, I. *Noise reduction in speech processing*; Springer, 2009; pp 1−4.

(80) Knapp, B.; Dunbar, J.; Alcala, M.; Deane, C. M. Variable regions of antibodies and T-cell Receptors may not be sufficient in molecular simulations investigating binding. *J. Chem. Theory Comput.* **2017**, *13*, 3097−3105.

(81) Humphreys, D. D.; Friesner, R. A.; Berne, B. J. A multiple-time-step molecular dynamics algorithm for macromolecules. *J. Phys. Chem.* **1994**, *98*, 6885−6892.

(82) Krieger, E.; Vriend, G. New ways to boost molecular dynamics simulations. *Journal of computational chemistry* **2015**, *36*, 996−1007.