Eurecon: Equidistant Uniform Rigid-body Ensemble Constructor

Petr Popov, Sergei Grudinin

Eurecon: Equidistant Uniform Rigid-body Ensemble Constructor

P. Popov[a,*], S. Grudinin[b]

[a]*Moscow Institute of Physics and Technology, Dolgoprudny, Russia*
[b]*Inria, Grenoble, France*

Abstract

Conformational ensembles comprise one of the fundamental concepts in statistical bioinformatics and appear in a variety of applications, e.g. molecular docking, virtual screening, searching for pharmacophores, etc. High-throughput applications require billions of conformations to be considered, thus, one often uses the rigid-body representation of molecules or its fragments to cope with the computational cost. Of particular interest is generation of the near-native conformational ensembles, which consist of conformations structurally close to the biologically relevant ones. One possible way to compose such ensembles is to control the root mean square deviation (RMSD) between the original and the generated conformations. To the best of our knowledge there is no computational approach that guarantees that all the generated conformations have the desired RMSD with respect to the reference structure. In this study we presented a fast algorithm for the construction of rigid-body conformational ensembles, which possess two main properties: **i**) each generated conformation has a fixed RMSD with respect to the original conformation, **ii**) generated conformations are distributed uniformly over the sphere of axes corresponding to the rigid-body motions. The algorithm is very efficient, it does not require any standard RMSD computation between the conformations and has the O(N+M) complexity to generate the required rigid-body transforms, where N is the number of atoms in the system, and M is the size of the conformational ensemble. Eurecon is applicable to an arbitrary atomic system, thus, it could be used

*Corresponding author

Email addresses: popov.pa@mipt.ru (P. Popov), sergei.grudinin@inria.fr (S. Grudinin)

for molecular systems of various size and type. We demonstrated Eurecon application by generating near-native conformational ensembles for a ligand placed inside a binding site, a protein dimer embedded into a membrane, and a ribosomal complex. We implemented the developed algorithm in C++ and called it Eurecon, which stands for **E**quidistant **U**niform **R**igid-body **E**nsemble **CON**structor. A user-friendly interface allows to define the desired RMSD value, the relative amplitudes for rotation and translation motions by means of the partition parameter, and the set of axes corresponding to the rigid-body motions. Eurecon is available as the SAMSON Element (`https://samson-connect.net`).

Keywords: conformational ensemble, rigid-body decoys, root mean square deviation, near-native conformations

---

Introduction

Conformational ensembles are one of the fundamental concepts in statistical bioinformatics. In general, a molecular system can adopt an infinite number of conformations, which form the conformational space of the system. In practice, however, one is only interested in some particular set of points in the conformational space, to which we will refer as to the *conformational ensemble*. Conformational ensembles appear in a variety of applications in structural bioinformatics. Even though it is not possible to describe all the applications comprising conformational ensembles, we will provide some references related to the recent advances in different problems in structural bioinformatics involving conformational ensembles. For example, the docking algorithms sample conformational space of a molecular complex and output a conformational ensemble of putative binding poses that comprises candidates for the native conformation, i.e. the biologically relevant conformation [1]. ParticularIy, in protein-ligand docking one often treats protein and ligand flexible during the sampling, while in protein-protein docking one uses the rigid-body sampling as the initial strategy, where the primary candidates are described by means of the rigid-body transforms. Furthermore, state-of-the art docking approaches use flexible conformational ensembles of the receptor as the input for docking, e.g. ensemble docking or 4D-docking [2, 3]. Using of the different receptor conformers helps to cover larger area of the conformational space and, thus, improves the chance to identify the native conformation [4]. To generate input conformational ensembles for

docking and virtual screening one uses known structures of the target receptor, molecular modeling techniques, e.g. molecular dynamics or normal mode analysis [1, 5], or employs both approaches [6]. Another example is the pharmacophore approach, which is often employed in drug discovery and requires a training set of diverse ligands in their bioactive conformations, i.e. that are adopted inside the binding pocket. Therein, the bioactive conformation of the ligand does not necessarily corresponds to the ligand's lowest energy conformation, since it typically undergoes conformational changes upon binding. Thus, to ensure the presence of bioactive conformations, one describes 'the shape' of a ligand as an ensemble of energetically accessible conformations [7, 8]. Derivation of data-driven scoring functions for intermolecular interactions, e.g. protein-ligand [9] or protein-protein [10], often requires a set of decoys, i.e. a conformational ensemble of various orientations of one partner, often a smaller molecule, with respect to the other. Decoy ensembles are used in order to construct a prediction model to discriminate between the native and non-native conformations. Similarly, many scoring functions for protein folding are also based on the pre-generated conformational ensembles of decoys [11]. Finally, conformational ensembles help to study activation/inactivation mechanisms of the target molecules. For example, several methods exist to investigate the activation mechanism of G-protein coupled receptors (GPCRs) based on the energy profiles derived from the modeled conformational ensembles [12, 13].

As one can see from the above-mentioned examples, the construction of an adequate conformational ensemble is a practical and important task. Some applications require billions of conformations to be considered, thus, the computationally expensive approaches become intractable. To cope with a large scale generation of the ensembles one often uses the rigid-body representation, which provides a dramatic reduction of the computational cost. For example, in protein-protein docking one often considers molecules as the rigid bodies during the sampling stage, while taking into account protein flexibility only for the most promising docking candidates. Another approach is to consider structurally ordered parts of a molecule, e.g. alpha-helices, as the rigid-bodies. For example in GPCR modeling, the Liticon [12] and the SuperBiHelix [13] methods consider seven transmembrane helices of the receptor as rigid bodies, sample conformations of each helix in the rigid-body subspace, and then combine the helices together resulting in the conformational ensemble of the complete receptor. Depending on the application, conformational ensembles must possess certain properties, e.g. contain only

structures close to the native structure in the conformational space (so called near-native structures). In particular, near-native conformational ensembles play important role in the development of protein folding force-fields [11] or statistical scoring functions (SSFs) for intermolecular interactions [14, 15]. Root mean square deviation (RMSD) between the two conformations corresponds to the distance between them in the conformational space, thus, it is the most natural and common metric to determine if a conformation is near-native or not. A near-native conformational ensemble consists of conformations that have low RMSD with respect to the native conformation. For example, in rigid-body modeling, to construct low-RMSD structures one typically samples conformations along the axes corresponding to the rigid-body motion (motion axes), using a fixed number of relatively small rotation and translation amplitudes. However, this method has a potential pitfall of producing non-native conformations, because even small rotation angles may result in conformations with large RMSD values with respect to the native conformation. Indeed, let us consider a DNA helix and two rotation axes, one pointing toward the direction of the helix and the other orthogonal to it (see Fig. 1). Rotation about these axes by a fixed angle results in two different conformations. Apparently, RMSD of the conformation produced by the rotation about the axis of the helix direction is much smaller compared to the RMSD of the other conformation. Thus, in order to compose the near-native conformational ensemble, one should take into account characteristics of the rigid body and the motion amplitudes with respect to the motion axes. To the best of our knowledge there is no computational approach that guarantees that all generated conformations have the desired RMSD with respect to the reference structure.

In this study we present an elegant algorithm to construct an *equidistant and uniform rigid-body conformational ensemble*, which possesses two main properties: **i**) each generated conformation has a fixed RMSD with respect to the original conformation, **ii**) generated conformations are distributed uniformly over the sphere of motion axes. The algorithm is very efficient and does not require any standard RMSD calculation between the conformations, which is linear with respect to the number of atoms in the input structure. Instead, it uses the rapid computation of RMSDs corresponding to macromolecular rigid-body motion[16]. Given the number of atoms in the original molecule N and the number of the generated conformations in the ensemble M, the complexity of the algorithm is $O(N \times M)$, if one generates explicit conformations (writes coordinates to the output), or $O(N + M)$, if one gen-

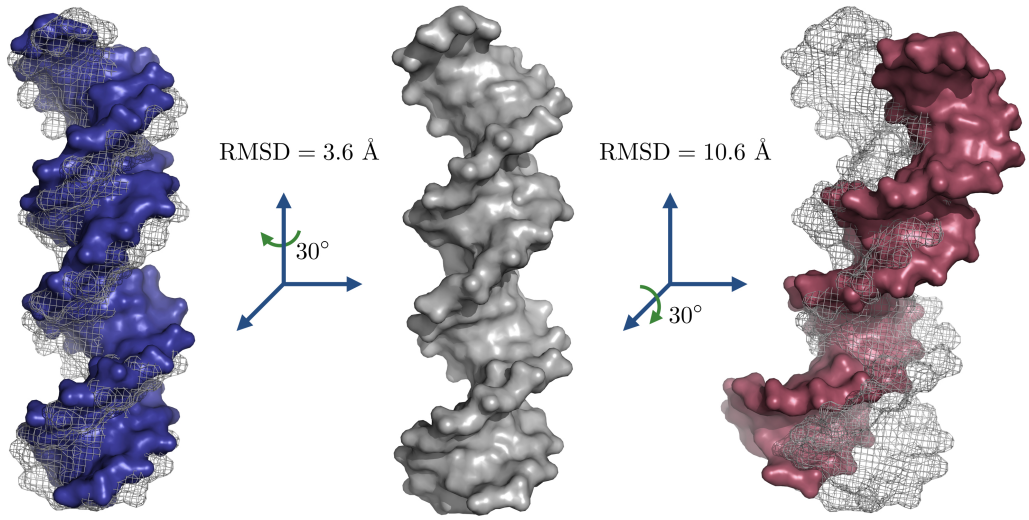

Figure 1: Rigid-body rotation of the DNA molecule (PDB code 5EGB) by fixed angle $\alpha = 30°$ about two different axes. Original DNA conformation is colored in gray, conformation corresponding to the axis pointing towards DNA is colored in blue, and conformation corresponding to the orthogonal axis is colored in red. RMSD values between the original and transformed conformations are shown.

erates implicit conformations (writes rigid-body transforms to the output). We implemented the presented algorithm in C++ and called it Eurecon, which stands for **E**qudistant **U**niform **R**igid-body **E**nsemble **CON**structor. Due to the equidistant property, the algorithm is very useful in applications where one needs to control the distances between the conformations in the conformational space.

Below, firstly we present the methodology and describe the algorithm in detail. Then, we provide computational experiments that demonstrate efficiency and functionality of Eurecon. We show the use of Eurecon by generating near-native conformational ensembles for different molecular systems, including a ligand placed inside the binding site, a protein dimer embedded into a membrane, and a ribosomal complex. Finally, we present implementation of Eurecon in the SAMSON platform (`https://samson-connect.net`), where it can be used for an arbitrary atomic system.

Methods

Generation of an equidistant conformational ensemble assumes the control of the RMSD of conformations with respect to the reference structure. Previously we showed that the mean square deviation (MSD) corresponding to the rigid-body motion could be efficiently calculated with the following equation [16] :

$$\mathrm{MSD} = \frac{4}{W}\sin^2\frac{\alpha}{2}\mathbf{n}^T\mathbf{I}\mathbf{n} + T^2, \tag{1}$$

where $\mathbf{I}$ is the inertia tensor of the rigid-body, α is the rotation angle about the unit axis $\mathbf{n}$, W is the mass of the rigid-body, and T is the translation amplitude. The first term in Eq. 1 corresponds to the rotational part of the MSD, and the second term corresponds to the translational part. Note, that the same MSD value could be achieved in different ways by varying the rotational and translational parts. Let us fix the MSD value and introduce the partition parameter p, which reflects the motion amplitudes. Then one can express the rotation and translation MSD terms as:

$$\begin{aligned} \mathrm{MSD}_{\mathrm{rotation}} &= p \cdot \mathrm{MSD} \\ \mathrm{MSD}_{\mathrm{translation}} &= (1-p) \cdot \mathrm{MSD}, \end{aligned}$$

where

$$\mathrm{MSD}_{\mathrm{translation}} = T^2 \tag{2}$$

$$\mathrm{MSD}_{\mathrm{rotation}} = \frac{4}{W}\sin^2\frac{\alpha}{2}\mathbf{n}^T\mathbf{I}\mathbf{n} \tag{3}$$

Note that the translational MSD depends only on the translation amplitude T, while the rotational MSD depends on the rotation angle, rotation axis and the geometry of the molecule. Thus, given the RMSD value, the partition parameter, and the motion axis, one can compute the rigid-body transform to accomplish the RMSD:

$$\alpha = \pm 2\arcsin\left(\frac{\mathrm{RMSD}}{2}\sqrt{\frac{pW}{\mathbf{n}^T\mathbf{I}\mathbf{n}}}\right) \tag{4}$$

$$T = \mathrm{RMSD}\sqrt{(1-p)}, \tag{5}$$

provided that the rotation angle α exists. Having this, the problem of the ensemble construction is reduced to collecting a sufficient number of the motion

axes corresponding to the rotation and translation movements. We compute these axes using the unit sphere tessellation by icosahedron subdivision [17]. More precisely, starting from an icosahedron with twelve vertices and twenty triangular facets, one connects midpoints of each edge within each facet, thus splitting each triangle into four new triangles. This procedure is repeated for the obtained set of triangles until the desired level of tessellation is achieved. Figure 2 demonstrates the first, second and third tessellation levels, which correspond to 80, 320 and $1,280$ motion axes, respectively. Finally, the set of normalized radius-vectors to the centroids of each triangle is taken as the collection of the motion axes. This set of the motion axes is generated only once and could be used further for any atomic system. Then, for each motion axis $\mathbf{n}$ from the set, one calculates the rigid-body transform corresponding to the given RMSD value according to Eq. 5 The complexity of these steps is $O(N + M)$, where N is the number of atoms in the system and M is the number of motion axes in the set. The $O(N)$ term corresponds to the calculation of the inertia tensor and must be computed only once, while the $O(M)$ term corresponds to the calculation of the rigid-body transform, and it is independent from the atomic coordinates. Writing output coordinates of each generated conformation is linear with respect to N, resulting in $O(N \times M)$ complexity. However, producing rigid-body transforms only may be sufficient for some applications, e.g. rigid-body docking. Thus, we will refer to the case with output coordinates as to the *explicit mode* , and to the case with output rigid-body transforms as to the *implicit mode*.

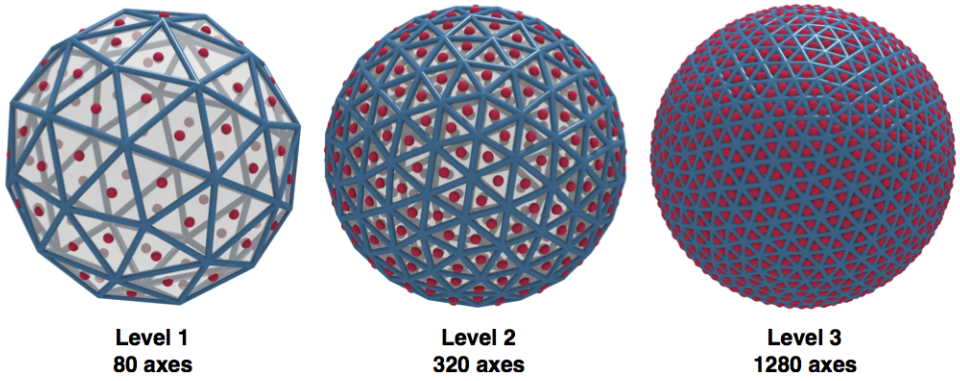

Figure 2: Sphere tessellations by icosahedron subdivision obtained on the first, second, and third levels, which correspond to 80, 320 and $1,280$ triangular facets, respectively. The midpoint of each triangular facet corresponds to the motion axis.

We implemented the presented algorithm in C++ and called it Eurecon, which stands for **E**qudistant **U**niform **R**igid-body **E**nsemble **CON**structor. For the following computational experiments we used 64-bit Linux Fedora operating system with Intel(R) Core(TM) i7-4700HQ CPU 2.4 GHz and g++ compiler version 4.6 with O3 optimization level. All the computations were performed using a single core, however the source code is easy to parallelize over the conformations. For the graphical presentation of the generated conformational ensembles we used PyMOL [18].

Results and Discussion

Timings

To demonstrate the efficiency and to support claims about the computational complexity of Eurecon we performed two tests. For the first test, we fixed the number of atoms in the molecular system and measured the elapsed time to generate ensembles of various number of conformations. More precisely, we used system of 1,000 atoms, generated ensembles of 100, 200, ..., 1,000 conformations, and measured the elapsed time to generate each of the ensembles. For the second test, we fixed the number of conformations in the output ensemble and measured the elapsed time to generate ensembles for molecular systems of various number of atoms. More precisely, we generated ensembles of 500 conformations and used molecular systems of 500, 1,000, 2,000, ..., 5,000 atoms. We performed these tests in the explicit (writing output coordinates) and the implicit (writing output rigid-body transforms) modes. Each test was repeated 200 times to calculate the mean and standard deviation values. Figures 3 **A** and 3 **B** present the obtained results for the first and the second tests, respectively.

As one can see Eurecon is very efficient - it takes seconds to generate $1,000$ conformations for a molecule of $5,000$ atoms in the explicit mode, and only milliseconds in the implicit mode. The most time consuming part is writing the output coordinates to a file (see 3 **B**), which transfers the $O(N + M)$ complexity to $O(N \times M)$. Both modes have linear complexity with respect to the number of output conformations. The explicit mode has linear complexity with respect to the number of atoms, whereas the implicit mode in practice has constant time complexity. This is because the $O(N)$ term in the implicit mode corresponds to the initialization step, which is performed only once, therefore the $O(M)$ term dominates over the $O(N)$ term.

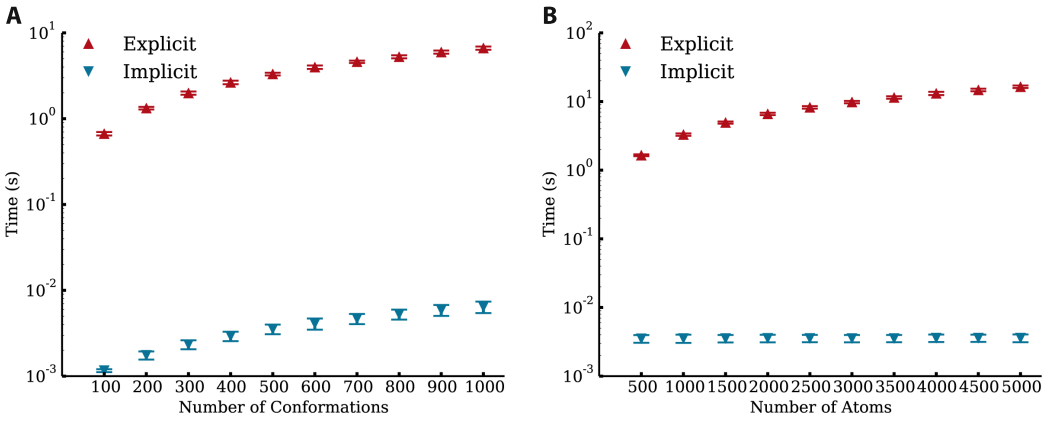

Figure 3: **A.** Elapsed time spent by Eurecon on generating ensembles for the molecular system of 1,000 atoms with respect to the number of conformations in the ensemble.
B. Elapsed time spent by Eurecon on generating ensembles of 500 conformations with respect to the number of atoms in the molecular system.
Explicit (write output coordinates) and implicit (write output rigid-body transforms) modes are represented with red upper and blue down triangles, respectively. Error bars represent standard deviations, which are computed for each point from 200 independent runs.

Functionality

Uniform and equidistant properties are the main advantages of Eurecon over the standard rigid-body sampling strategy. Indeed, fixed-angle sampling over the rotation axes is not uniform with respect to the conformational space. Fig. 1 demonstrates that distance between such conformations could be very large (3.6 Å against 10.6 Å for the rotation angle $\alpha = 30°$). In contrast, Eurecon, having the same computational complexity, allows to control the distance between the conformations in the conformational space. In the next test we demonstrate examples of the conformational ensembles produced by Eurecon with different parameters. More precisely, we used three RMSD values (1.0 Å, 3.0 Å, 5.0 Å) and three values of the partition parameter (1.0, 0.5, 0.0). Note, that the partition values of 1.0 and 0.0 correspond to the pure rotation and pure translation cases, respectively. For clarity of representation we used small protein alpha-helix as the molecular system, and positions of the six octahedron vertices as the motion axes. Figure 4 shows the obtained conformational ensembles.

Please note that a pure rotational ensemble could not be produced for any RMSD value. From Eq. 3 the maximum RMSD value is achieved with

the rotation angle $\alpha = 180^\circ$ and depends on the geometry of the rigid-body as well as on the motion axis. From the other hand, the pure translation case is trivial and does not require RMSD computation. Therefore, Eurecon is practically useful for the construction of conformational ensembles with rotational motions. Finally, we would like to note that different applications could require various size of conformational ensemble to adequately represent the target molecule. We implemented 5 tessellation levels corresponding to the minimum size of 20 and the maximum size of 5120 conformations of the ensemble. Furthermore, the algorithm works with an arbitrary set of motion axes, and we use the tessellation levels only to ensure uniform property of the generated ensemble.

Generation of conformational ensemble for molecular systems

Since the Eurecon algorithm relies on only geometrical properties of the system, it can be used in different applications with molecular systems of various size and type. For example, Eurecon could be applied to study and analyze binding modes of molecular complexes, e.g. protein-protein or protein-ligand complexes, and, particularly, to derive statistical scoring functions (SSFs) for molecular docking. SSFs appear to be more computationally efficient, compared to the molecular mechanics force-fields, and very suitable for the docking and virtual screening [10]. To derive SSFs one typically uses a training set, which comprises few native conformations and a lot of non-native conformations obtained from the docking runs. However, recently we showed that near-native conformations help to derive powerful SSFs both for protein-protein [14] and protein-ligand complexes [15]. Thus, Eurecon could be used to generate conformational ensemble of the near-native binding modes in order to enlarge the training set for the derivation of SSFs. Figure 5 **A** illustrates a conformational ensemble composed for the ZMA ligand in the binding pocket of the adenosine receptor (PDBID: 52KC) [19].

Another example comes from the molecular dynamic simulations, where the atomic system consists of a dimer of the transmembrane alpha-helical protein glycophorin A (GpA) embedded into the membrane environment. GpA in a lipid bilayer is a widely used atomic system to study the free energy of association of the transmembrane alpha-helices as well as the influence of the membrane on the dimerization [20, 21, 22]. One of the approaches to study the GpA association is to set up different starting orientations of the monomers in the dimer, and then to launch molecular dynamics simulations using a pulling force to dissociate the monomers. Then one could determine

and study the relevant association factors by analyzing the output simulation trajectories. For such systems Eurecon allows to generate equidistant starting orientations of the monomers with respect to each other as well as to the membrane environment. Varying RMSD parameter, for example, one can determine the critical RMSD value and motion axes that prevent favorable interactions for the dimer association. We applied Eurecon to the GpA system taken from MemProtMD [23] and the corresponding starting orientations of the monomers are shown in Fig. in 5 **B**.

Finally, some applications require sampling of constituting fragments rather than the entire molecule. Sampling of different parts of the molecule as rigid-bodies is useful to simulate partial flexibility of the system and, thus, to compose flexible conformational ensembles. Particularly, this type of approaches is used for the construction of flexible conformational ensembles of GPCRs [12, 13], where one samples seven alpha-helices of the receptor as rigid-bodies and then combines rigid-body ensembles with the rest of the protein together, resulting in the flexible conformational ensemble of the entire protein. Eurecon could be advantageous to these methods, due to its equidistant property. Indeed, Eurecon allows to control the amplitude of the motion for each part of the molecule, hence, for more (less) structurally conserved parts of the molecule one could use smaller (larger) RMSD values. Figure 5 **C** shows conformational ensembles obtained with Eurecon for different protein chains of the large ribosomal unit (PDBID : 3CC2, 29 unique protein chains, nucleic acids are hidden for clarity of representation) [24].

Availability

We implemented Eurecon as the SAMSON-Element application (Eurecon App) in the SAMSON platform (`https://samson-connect.net`). Due to the graph representation of SAMSON, Eurecon App can be used for an arbitrary structural system, and it works with all file formats supported by SAMSON, e.g. .pdb, .mol, .xyz, etc. Eurecon App has a user-friendly interface, where in order to generate a conformational ensemble, the user selects atomic system as structural nodes and inputs the RMSD value, the partition parameter, and the tessellation level. Then Eurecon App generates the conformational ensemble for the selected structural nodes, stores it on a disk and displays it in SAMSON as a set of conformations (see Fig. 6).

Conclusions

In this study we presented Eurecon - a fast algorithm for the construction of equidistant rigid-body conformational ensembles The algorithm is very efficient, it does not require any standard RMSD computation between the conformations and has the O(N+M) complexity to generate the required rigid-body transforms, where N is the number of atoms in the system, and M is the size of the conformational ensemble. A user-friendly interface allows to define relative amplitudes for the rotation and translation motions by means of the partition parameter as well as a set of axes corresponding to the rigid-body motion. Eurecon is particularly useful to generate near-native conformational ensembles, i.e. conformations structurally close to the reference structure. Eurecon is applicable to an arbitrary geometrical system, thus, it could be used for molecular systems of various size and type. We demonstrated Eurecon application by generating conformational ensembles for a ligand placed inside a binding site, a protein dimer embedded into a membrane, and a ribosomal complex. Eurecon is available as the SAMSON Element (`https://samson-connect.net`).

Acknowledgements

This work was supported by the Russian Science Foundation (project no. 16-14-10273) and by the Agence Nationale de la Recherche (ANR-11-MONU-006-01).

[1] E. Yuriev, J. Holien, P. A. Ramsland, Improvements, trends, and new ideas in molecular docking: 2012–2013 in review, Journal of Molecular Recognition 28 (10) (2015) 581–604.

[2] W. Sinko, S. Lindert, J. A. McCammon, Accounting for receptor flexibility and enhanced sampling methods in computer-aided drug design, Chemical biology & drug design 81 (1) (2013) 41–49.

[3] G. Bottegoni, I. Kufareva, M. Totrov, R. Abagyan, Four-dimensional docking: a fast and accurate account of discrete receptor flexibility in ligand docking, Journal of medicinal chemistry 52 (2) (2009) 397–406.

[4] C. F. Wong, Flexible receptor docking for drug discovery, Expert opinion on drug discovery 10 (11) (2015) 1189–1200.

[5] G. Moroy, O. Sperandio, S. Rielland, S. Khemka, K. Druart, D. Goyal, D. Perahia, M. A. Miteva, Sampling of conformational ensemble for virtual screening using molecular dynamics simulations and normal mode analysis, Future medicinal chemistry 7 (17) (2015) 2317–2331.

[6] K. L. Damm, H. A. Carlson, Exploring experimental sources of multiple protein conformations in structure-based drug design, Journal of the American Chemical Society 129 (26) (2007) 8225–8235.

[7] C. H. Schwab, Conformations and 3d pharmacophore searching, Drug Discovery Today: Technologies 7 (4) (2010) e245–e253.

[8] S. Riniker, G. A. Landrum, Better informed distance geometry: Using what we know to improve conformation generation, Journal of chemical information and modeling 55 (12) (2015) 2562–2574.

[9] M. A. Khamis, W. Gomaa, Comparative assessment of machine-learning scoring functions on pdbbind 2013, Engineering Applications of Artificial Intelligence 45 (2015) 136–151.

[10] I. H. Moal, M. Torchala, P. A. Bates, J. Fernández-Recio, The scoring of poses in protein-protein docking: current capabilities and future directions, BMC bioinformatics 14 (1) (2013) 286.

[11] H. Deng, Y. Jia, Y. Zhang, 3drobot: automated generation of diverse and well-packed protein structure decoys, Bioinformatics (2015) btv601.

[12] S. Bhattacharya, N. Vaidehi, Liticon: a discrete conformational sampling computational method for mapping various functionally selective conformational states of transmembrane helical proteins, Membrane Protein Structure and Dynamics: Methods and Protocols (2012) 167–178.

[13] J. K. Bray, R. Abrol, W. A. Goddard, B. Trzaskowski, C. E. Scott, Superbihelix method for predicting the pleiotropic ensemble of g-protein–coupled receptor conformations, Proceedings of the National Academy of Sciences 111 (1) (2014) E72–E78.

[14] P. Popov, S. Grudinin, Knowledge of native protein–protein interfaces is sufficient to construct predictive models for the selection of binding

candidates, Journal of chemical information and modeling 55 (10) (2015) 2242–2255.

[15] M. Kadukova, S. Grudinin, Convex-pl: a novel knowledge-based potential for protein-ligand interactions deduced from structural databases using convex optimization, Journal of Computer-Aided Molecular Design (2017) 1–16.

[16] P. Popov, S. Grudinin, Rapid determination of rmsds corresponding to macromolecular rigid body motions, Journal of computational chemistry 35 (12) (2014) 950–956.

[17] G. Hoffmann, G. Hoffmann, Sphere tessellation by icosahedron subdivision (2002).

[18] Schrödinger, LLC, The PyMOL molecular graphics system, version 1.8 (November 2015).

[19] A. Batyuk, L. Galli, A. Ishchenko, G. W. Han, C. Gati, P. A. Popov, M.-Y. Lee, B. Stauch, T. A. White, A. Barty, et al., Native phasing of x-ray free-electron laser data for a g protein–coupled receptor, Science advances 2 (9) (2016) e1600292.

[20] H. I. Petrache, A. Grossfield, K. R. MacKenzie, D. M. Engelman, T. B. Woolf, Modulation of glycophorin a transmembrane helix interactions by lipid bilayers: molecular dynamics calculations, Journal of molecular biology 302 (3) (2000) 727–746.

[21] E. Psachoulia, D. P. Marshall, M. S. Sansom, Molecular dynamics simulations of the dimerization of transmembrane α-helices, Accounts of chemical research 43 (3) (2009) 388–396.

[22] J. Koehler Leman, M. B. Ulmschneider, J. J. Gray, Computational modeling of membrane proteins, Proteins: Structure, Function, and Bioinformatics 83 (1) (2015) 1–24.

[23] P. J. Stansfeld, J. E. Goose, M. Caffrey, E. P. Carpenter, J. L. Parker, S. Newstead, M. S. Sansom, Memprotmd: automated insertion of membrane protein structures into explicit lipid membranes, Structure 23 (7) (2015) 1350–1361.

[24] G. Blaha, G. Gürel, S. J. Schroeder, P. B. Moore, T. A. Steitz, Mutations outside the anisomycin-binding site can make ribosomes drug-resistant, Journal of molecular biology 379 (3) (2008) 505–519.

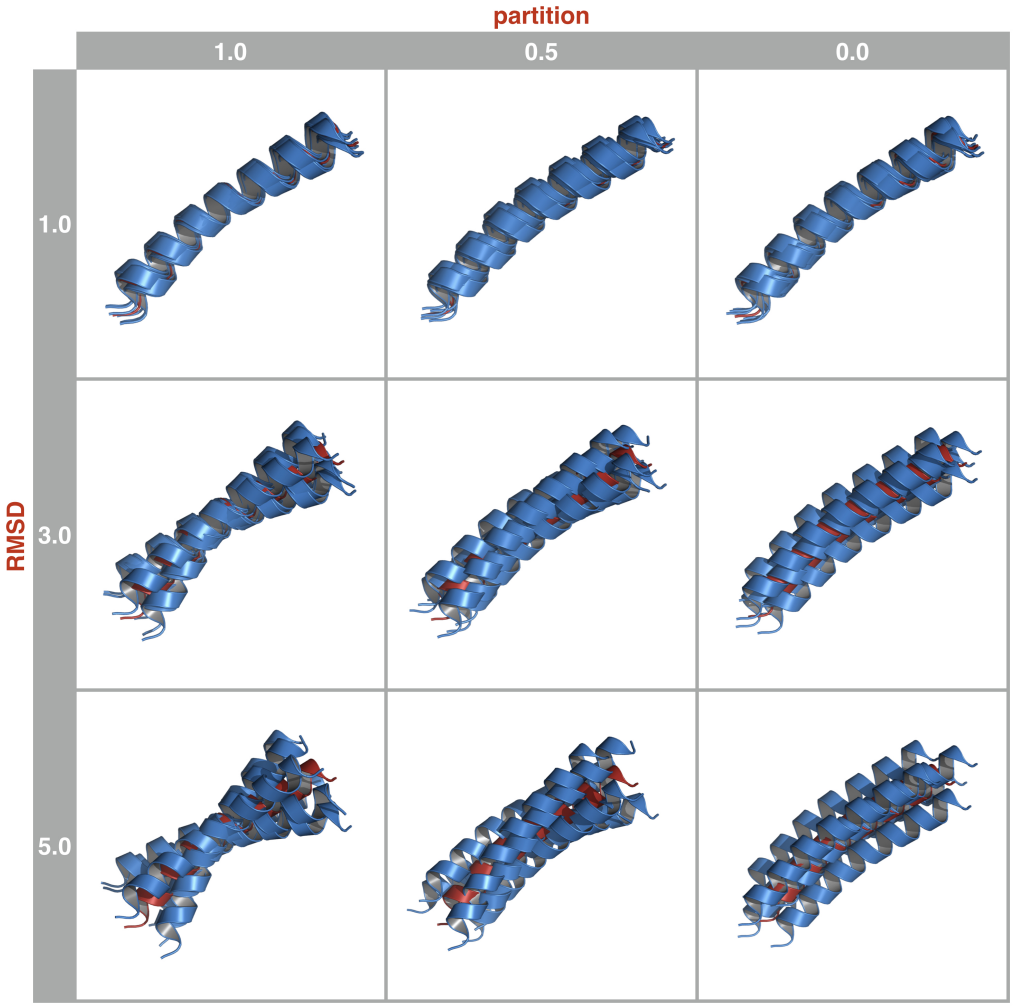

Figure 4: Example of conformational ensembles for alpha-helical protein produced by Eurecon with different parameters. Set of the motion axes correspond to six vertices of the octahedron. Conformational ensembles in one column correspond to the fixed partition value and RMSD values of 1.0 Å, 3.0 Å, and 5.0 Å, respectively. Conformational ensembles in one row correspond to the fixed RMSD value and partition values of 1.0 (pure rotation), 0.5 (rotation and translation), and 0.0 (pure translation), respectively. The original conformation of the alpha-helical protein is colored in red, and the generated conformations are colored in blue.

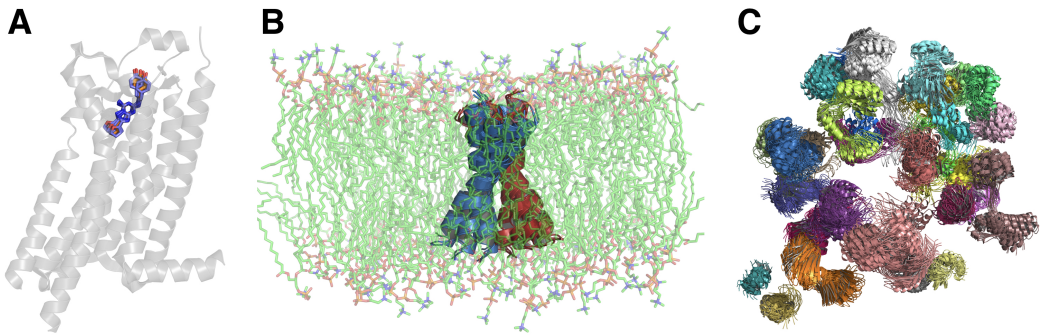

Figure 5: Examples of applying Eurecon for molecular systems of various type and size. **A.** Conformational ensemble of the ZMA ligand (orange sticks) inside the binding pocket of the adenosine receptor (grey cartoon). Generated conformations of the ligand are presented as magenta sticks.
B. Conformational ensemble of the glycophorin A dimer embedded into the lipid bilayer. Eurecon is applied to each monomer, and the corresponding conformational ensembles are represented as red and blue ribbons.
C. Conformational ensemble of the large ribosomal subunit complex. Eurecon is applied to each protein chain of the subunit (RNA chains are hidden for the clarity).

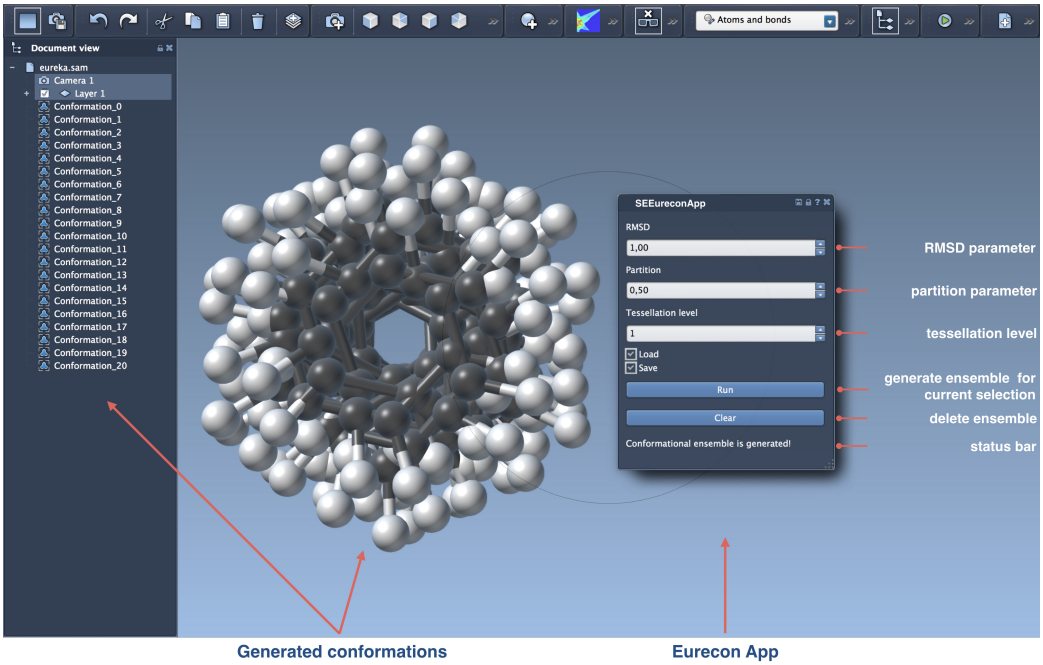

Figure 6: Screenshot of Eurecon App implemented in the SAMSON modeling platform.