



MoMA-LoopSampler: A web server to exhaustively sample protein loop conformations

Amélie Barozet, Kevin Molloy, Marc Vaisset, Christophe Zanon, Pierre Fauret, Thierry Siméon, Juan Cortés

► To cite this version:

Amélie Barozet, Kevin Molloy, Marc Vaisset, Christophe Zanon, Pierre Fauret, et al.. MoMA-LoopSampler: A web server to exhaustively sample protein loop conformations. *Bioinformatics*, 2022, 38 (2), pp.552-553. 10.1093/bioinformatics/btab584 . hal-03326493

HAL Id: hal-03326493

<https://laas.hal.science/hal-03326493>

Submitted on 26 Aug 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Structural bioinformatics

MoMA-LoopSampler: A web server to exhaustively sample protein loop conformations

Amélie Barozet¹, Kevin Molloy^{2,3}, Marc Vaisset¹, Christophe Zanon¹,
Pierre Fauret¹, Thierry Siméon¹, and Juan Cortés^{1,*}

¹LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France

²Computer Science, James Madison University, Harrisonburg, VA. 22181, USA

³School of Biology, James Madison University, Harrisonburg, VA. 22181, USA

*To whom correspondence should be addressed.

Abstract

Summary: MoMA-LoopSampler is a sampling method that globally explores the conformational space of flexible protein loops. It combines a large structural library of three-residue fragments and a novel reinforcement-learning-based approach to accelerate the sampling process while maintaining diversity. The method generates a set of statistically-likely loop states satisfying geometric constraints, and its ability to sample experimentally observed conformations has been demonstrated. This paper presents a web user interface to MoMA-LoopSampler through the illustration of a typical use-case.

Availability: MoMA-LoopSampler is freely available at: <https://moma.laas.fr/applications/LoopSampler/>. We recommend users to create an account, but anonymous access is possible. In most cases, jobs are completed within a few minutes. The waiting time may increase depending on the server load, but it very rarely exceeds an hour. For users requiring more intensive use, binaries can be provided upon request.

Contact: juan.cortes@laas.fr

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Loops are irregular components of protein structures, which display significant flexibility in many cases. They play important roles in processes such as molecular recognition, enzyme catalysis, allosteric regulation or signaling. Their structural investigation is challenging due to the limitations of experimental methods to accurately characterize flexible bio-molecules. Therefore, experimental data needs to be complemented with models obtained from computational methods. Numerous loop modeling approaches have been proposed over more than two decades (Shehu and Kavraki, 2012; Papaleo *et al.*, 2016; Kundert and Kortemme, 2019). Despite such active research, loop modeling remains an open problem, in particular for long flexible loops (Marks *et al.*, 2018; Barozet *et al.*, 2021). We recently proposed a loop sampling method combining ideas of *ab initio* and knowledge-based approaches, and applying an artificial-intelligence-based technique for enhanced performance (Barozet *et al.*, 2020). Compared with other state-of-the-art loop sampling methods, MoMA-LoopSampler provides higher-quality samples with comparable computational cost. Here, we present a web interface for this method and we illustrate its application to generate conformational ensemble models of missing loops in

protein structures determined by X-ray crystallography. The application to build missing loops in protein structures obtained from homology modeling is another very interesting scenario that remains to be studied.

2 Summary of the method

MoMA-LoopSampler is based on a subdivision of the loop into a sequence of three-residue fragments, called *tripeptides*. From this subdivision, it performs an incremental process to sample loop states. At each iteration, a tripeptide t_i is randomly selected. Then, the configurations of the remaining tripeptides in the loop are built: from the N-terminal anchor residue to t_{i-1} and from the C-terminal anchor residue to t_{i+1} . The backbone dihedral angles are sampled from an extensive database of tripeptides extracted from coil regions in experimentally determined protein structures. Collision detection and forward reference checking are performed to prune the search space, and the algorithm backtracks if constraints are violated. A reinforcement-learning-based approach is applied to sample more frequently regions in the database that lead to valid loop states. Once the two halves of the loop have been sampled, the configuration of tripeptide t_i is obtained using a semi-analytical inverse kinematics (IK) method (Cortés *et al.*, 2004) that enforces loop closure. Optionally, the loop side chains are placed in a subsequent stage (following an incremental

procedure). The reader is referred to our previous publication (Barozet *et al.*, 2020) for additional detail about the method.

3 A typical use-case

Thiamin phosphate synthase is one among many examples of enzymes involving flexible loops in the active site (Beglov *et al.*, 2018). Its structure was determined by X-ray crystallography in different conditions (Peapus *et al.*, 2001): structures were determined for the unliganded protein (PDB ID: 1G4E), and for complexes with substrates, intermediates and products (PDB IDs: 1G4P, 1G4S, 1G4T, 1G67, 1G69, 1G6C). Overall, the TIM-barrel-like fold remains unchanged in all the cases, but some loops show conformational changes, in particular loop 6 (residues 150-169). Due to its intrinsic flexibility, the full structure of the loop cannot be determined when its conformation is not stabilized by interactions with the thiazole group of the ligand. A large fragment of the loop is missing in the unliganded structure 1G4E. A comparison with the available structures for the liganded protein shows that the conformations of a few residues at the beginning and at the end of the loop remain relatively unchanged, whereas the middle fragment, bounded by residues Tyr154 and Ala164, is flexible.

Figure 1 illustrates the application of MoMA-LoopSampler to generate a conformational ensemble model of loop 6. First, the name `TPsynthaseLoop6` is assigned to the job in order to identify it. Next, the user must provide the PDB file of the protein scaffold and the chain identifier (if the file contains several chains)¹, as well as the numbers of the anchor residues: 154 and 164. In the case of a missing loop, the amino acid sequence of the protein fragment between the anchor residues must be provided: `PTETKQDTR` in this case. This is not mandatory if the coordinates of a loop conformation are present in the input PDB file. Finally, the user must specify the desired number of samples (up to 5000 in the current implementation of the server). Other parameters/options (not represented in the figure) are available via the web interface: (1) The sampled loop states are provided in a single PDB file, using the MODEL record type as separator. The generation of a single PDB per sampled state is an option. (2) By default, sampled loop states returned as output do not contain atomic models of the side chains (this is an option). Nevertheless, a coarse-grained model of the loop side chains (a residue-dependent pseudo atom at the $C\beta$ position) is considered in the sampling process. (3) The method rejects configurations involving collisions, which is determined using 70% of the atoms van der Waals radii. This is a tunable parameter. (4) At each iteration of the overall sampling algorithm, the IK solver may provide several solutions for the angles of the loop-closing tripeptides. By default, only one collision-free solution is saved, but the user can choose to keep all valid solutions. Finally note that MoMA-LoopSampler may slightly perturb the conformation of one or two residues at both ends of the loop (involving the anchor residues) due to the sampling procedure that considers three-residue fragments.

The results for this experiment are detailed in the Supplementary Data. An additional use-case of MoMA-LoopSampler is also included.

¹ We recommend the user to clean the input PDB file and to keep only the chain of interest. Ligands or other small molecules, water molecules and ions are automatically removed.

MoMA-LoopSampler

Name of the job

PDB file
 1g4eA.pdb

Chain

N-ter anchor

C-ter anchor

Loop sequence (for missing loops)

Number of sampled states

... other parameters/options ...

Fig. 1. Graphical explanation of the application of MoMA-LoopSampler to generate a conformational ensemble model of a missing loop (loop 6) from the unliganded structure of thiamin phosphate synthase.

Funding

This work has been supported by the French ANR through grant ANR-19-PI3A-0004.

Conflict of interest: None declared.

References

- Barozet, A., Molloy, K., Vaisset, M., Siméon, T., and Cortés, J. (2020). A reinforcement-learning-based approach to enhance exhaustive protein loop sampling. *Bioinformatics*, **36**(4), 1099–1106.
- Barozet, A., Bianciotto, M., Vaisset, M., Siméon, T., Minoux, H., and Cortés, J. (2021). Protein loops with multiple metastable conformations: a challenge for sampling and scoring methods. *Proteins*, **89**(2), 218–231.
- Beglov, D., Hall, D. R., Wakefield, A. E., Luo, L., Allen, K. N., Kozakov, D., Whitty, A., and Vajda, S. (2018). Exploring the structural origins of cryptic sites on proteins. *Proc. Natl. Acad. Sci. U.S.A.*, **115**(15), E3416–E3425.
- Cortés, J., Siméon, T., Remaud-Siméon, M., and Tran, V. (2004). Geometric algorithms for the conformational analysis of long protein loops. *J. Comput. Chem.*, **25**(7), 956–967.
- Kundert, K. and Kortemme, T. (2019). Computational design of structured loops for new protein functions. *Biol. Chem.*, **400**(3), 275–288.
- Marks, C., Shi, J., and Deane, C. M. (2018). Predicting loop conformational ensembles. *Bioinformatics*, **34**(6), 949–956.
- Papaleo, E., Saladino, G., Lambrugh, M., Lindorff-Larsen, K., Gervasio, F. L., and Nussinov, R. (2016). The role of protein loops and linkers in conformational dynamics and allostery. *Chem. Rev.*, **116**(11), 6391–6423.
- Peapus, D. H., Chiu, H.-J., Campobasso, N., Reddick, J. J., Begley, T. P., and Ealick, S. E. (2001). Structural characterization of the enzyme-substrate, enzyme-intermediate, and enzyme-product complexes of thiamin phosphate synthase. *Biochemistry*, **40**(34), 10103–10114.
- Shehu, A. and Kavraki, L. E. (2012). Modeling structures and motions of loops in protein molecules. *Entropy*, **14**(12), 252–290.