



gmXtal: Cooking Crystals with GROMACS

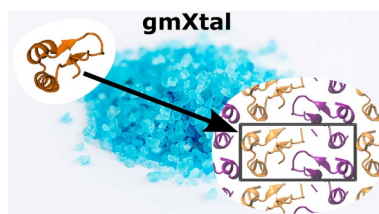
Pavel Buslaev¹ · Gerrit Groenhof¹

Accepted: 23 July 2023
© The Author(s) 2023

Abstract

Molecular dynamics (MD) simulations are routinely performed of biomolecules in solution, because this is their native environment. However, the structures used in such simulations are often obtained with X-ray crystallography, which provides the atomic coordinates of the biomolecule in a crystal environment. With the advent of free electron lasers and time-resolved techniques, X-ray crystallography can now also access metastable states that are intermediates in a biochemical process. Such experiments provide additional data, which can be used, for example, to optimize MD force fields. Doing so requires that the simulation of the biomolecule is also performed in the crystal environment. However, in contrast to simulations of biomolecules in solution, setting up a crystal is challenging. In particular, because not all solvent molecules are resolved in X-ray crystallography, adding a suitable number of solvent molecules, such that the properties of the crystallographic unit cell are preserved in the simulation, can be difficult and typically is a trial-and-error based procedure requiring manual interventions. Such interventions preclude high throughput applications. To overcome this bottleneck, we introduce **gmXtal**, a tool for setting up crystal simulations for MD simulations with GROMACS. With the information from the protein data bank (rcsb.org) **gmXtal** automatically (i) builds the crystallographic unit cell; (ii) sets the protonation of titratable residues; (iii) builds missing residues that were not resolved experimentally; and (iv) adds an appropriate number of solvent molecules to the system. **gmXtal** is available as a standalone tool <https://gitlab.com/pbuslaev/gmxtal>.

Graphical Abstract



Keywords GROMACS · Molecular dynamics · Crystal · Protein

1 Introduction

Multi-dimensional Nuclear Magnetic Resonance (NMR) spectroscopy can provide structural and dynamical information on biomolecules in solution, but this technique is

limited to macromolecules of up to 35 kDa [1]. In contrast, in X-ray crystallography there is no such limit [2, 3], but this technique can only provide static information about protein structure in a crystal environment, where there are strong inter-molecular interactions between the proteins. With the development of free electron lasers and time-resolved techniques, also protein dynamics can now be probed by X-ray crystallography, but the limitation of the non-native crystal environment remains [4].

Protein dynamics in solution have therefore been predominantly investigated by means of molecular dynamics (MD) computer simulations [5]. The initial structure for

✉ Pavel Buslaev
pavel.i.buslaev@jyu.fi

✉ Gerrit Groenhof
gerrit.x.groenhof@jyu.fi

¹ Department of Chemistry and Nanoscience Center, University of Jyväskylä, 40014 Jyväskylä, Finland

such simulations is normally prepared by placing a single asymmetric unit from an X-ray crystal structure at the center of a periodic simulation box that is subsequently filled with water molecules, ions and other constituents if needed. However, because the protein environment in the simulations differs significantly from that in the crystallography experiment, comparing observables obtained from MD simulations with observables from crystallography remains challenging. Already in the early days of MD, Herman Berendsen et al. had addressed this issue and compared trajectories of proteins in solution and crystal environments [6, 7]. Later, it was demonstrated that proteins simulated in solution can relax into a conformation that is different from the X-ray structure [8–10]. Thus, to validate MD simulations in general and force fields in particular, with either static or dynamic X-ray data, simulations would need to be performed of protein in crystals, rather than solution [11–20].

To routinely perform MD simulations of protein crystals, efficient standardized approaches are needed for building crystal models. Whereas for setting up simulations of proteins in solution, a wide range of user-friendly methods has been introduced [21–23], the number of tools available to create a crystal model is much more limited. Although good guidelines exist, for example the tutorial of Cerutti and Case [24], and also CHARMM-GUI can generate protein crystals [21], these approaches are difficult to scale up for high-throughput applications, which would be needed for a systematic validation and improvement of force fields based on X-ray data. Furthermore, because not all solvent molecules are typically resolved in X-ray structures, determining how many solvent molecules are needed to preserve the volume of the unit cell while sampling the isothermal-isobaric ensemble in MD simulations under periodic boundary conditions, remains a major challenge that is currently not resolved. Instead, the addition of solvent molecules, which is essential to preserve the volume of the unit-cell, requires manual intervention, often by trial and error.

To overcome these challenges and provide the community with an efficient tool for setting up protein crystals for MD simulations in an automated and hence reproducible manner, we have implemented **gmXtal**, a combination of python scripts that automatically constructs the unit-cell, estimates the correct number of solvent molecules to be added and prepares the workflow for running MD simulations of the hydrated crystal with the GROMACS MD program [25, 26]. This paper is organized as follows: we first introduce the steps of the **gmXtal** workflow, we then provide the technical details of the simulations we performed, and finally demonstrate how **gmXtal** works by building crystals of three representative biomolecules.

2 Methods

2.1 gmXtal

gmXtal is a python-based toolbox, which automatically builds structures, generates topologies, and prepares scripts for simulating biomolecular crystals with GROMACS [25, 26]. **gmXtal** consists of three modules: (i) a preparation module, (ii) a GROMACS workflow, and (iii) a check module (Fig. 1), which are explained in detail below. **gmXtal** can also be used for setting up simulations of biomolecules in solution, if requested. Building crystals of membrane proteins is currently not supported.

2.1.1 Preparation Module

The preparation module cleans the structure, builds the crystal and prepares a bash script for running a GROMACS workflow. The structure can either be downloaded from the protein data bank [27], or provided by the user. **gmXtal** works with PDBx/mmCIF format [28], which is currently the standard file format of the protein data bank. Structure cleaning includes (i) selecting alternative conformations and preferred conformers of residues, which are decided by the user, (ii) fixing missing residues based on the sequence information, and (iii) determining the optimal protonation state of titratable amino acids at the pH value selected by the user. Addition of missing residues is required to complete protein chains. Residues that are missing in the experimental structures, are from parts of the biomolecule that are less structured and hence do not give rise to Bragg peaks. Therefore, the resulting model structure might deviate from the real conformation, which can lead to bias in the simulation. Because adding missing residues can furthermore introduce steric clashes, **gmXtal** checks for such clashes and notifies the user. Structure cleaning also takes care of crystal waters, ions, and ligands, if a user wants these to remain included in the simulation box. Note, however, that **gmXtal** expects the ions to be part of the force field used. Also, if ligands are selected for processing, the user is expected to provide the structure and force field parameters of the ligand, including hydrogens, because at this stage, **gmXtal** cannot automatically parameterise or add hydrogen atoms to ligands.

The mmCIF files are parsed with the BIO.PDB [29] and PDBeCIF [30] packages. The cleaning of the input structure is performed with standard routines from the Bio.PDB [29] and MDAnalysis [31, 32] packages. If needed, the structure is fixed either with *pdbfixer* [33], or with *Modeller* [34]. The protonation states of titratable residues can either be decided upon by the user, or automatically with *pdb2gmX*, based on pKa predictions by PROPKA [22, 35] or pKa-ANI [36].

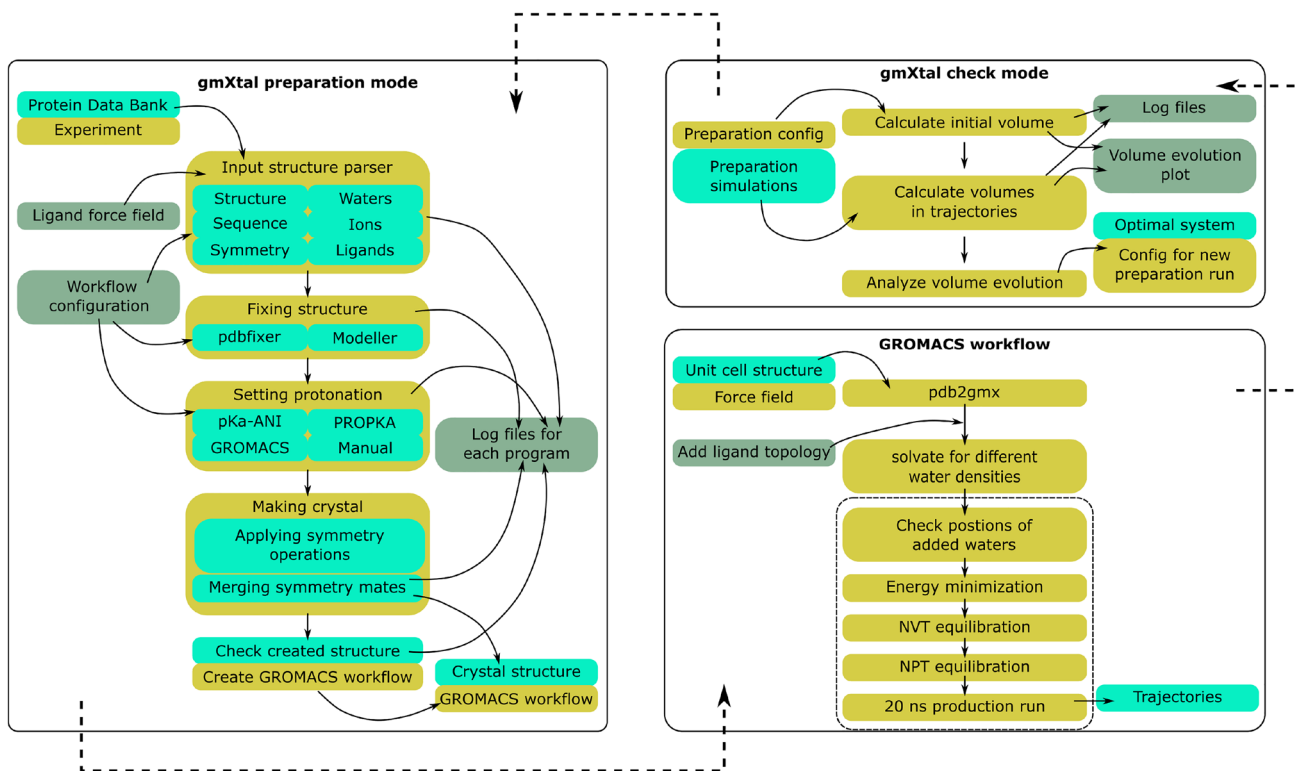


Fig. 1 gmXtal consists of three main parts: (i) a preparation module, (ii) a GROMACS [25, 26] workflow, and (iii) a check module. The preparation module retrieves the structure from the protein data bank [27] or uses a structure provided by the user, cleans the structure, builds the crystal and creates a bash script that sets up a GROMACS workflow. The GROMACS workflow solvates the crystal structure at different water concentrations, runs minimization, NVT and NPT equilibrations, and 20 nanosecond production runs. The check mod-

ule analyses the evolution of the simulation box volume for all production runs and compares the volume of systems with different numbers of added water molecules to the unit cell volume from the X-ray structure file. If none of the simulated systems preserves the experimental volume, the check module will suggest new parameters for the preparation module. Otherwise, it will recommend using the system that preserves the volume, for further production runs

Crystals are reconstructed using routines from the GEMMI library [37], which generates symmetry matrices from the symmetry information in the structural files available in PDBx/mmCIF format under *_symmetry:space_group_name_H-M* entry. GEMMI generates all symmetry matrices that correspond to the particular symmetry group. The generated matrices are so-called Seitz matrices $\{\mathbf{R}|\mathbf{t}\}$, which are transformation matrices in the cell coordinate space and thus, have to be translated into Cartesian space, \mathbf{M} . This is achieved by applying the following transformations:

$$\mathbf{M} = \mathbf{A} \cdot \mathbf{R} \cdot \mathbf{A}^{-1} \tag{1}$$

Here, \mathbf{R} is the rotational part of the Seitz matrix, and \mathbf{A} is the coordinate transformation matrix:

$$\mathbf{A} = \begin{pmatrix} a & b \cos(\gamma) & c \cos(\beta) \\ 0 & b \sin(\gamma) & -c \sin(\beta) \cos(\alpha^*) \\ 0 & 0 & c \sin(\beta) \sin(\alpha^*) \end{pmatrix} \tag{2}$$

with a, b, c the lengths of the unit cell vectors, and $\alpha, \beta,$ and γ the angles between these vectors. Together these six parameters determine the geometry of the unit cell. The parameter α^* is defined as

$$\cos(\alpha^*) = \frac{\cos(\beta) \cos(\gamma) - \cos(\alpha)}{\sin(\beta) \sin(\gamma)} \tag{3}$$

The translation vectors are computed from the translation part \mathbf{t} of the Seitz matrices as follows:

$$\mathbf{V} = \mathbf{A} \cdot \mathbf{t} \tag{4}$$

With the matrices \mathbf{M} and translation vectors \mathbf{V} , all crystallographic copies of asymmetric unit are created and then merged into a complete unit cell using routines from MDAnalysis [31, 32]. After the crystal structure is prepared, gmXtal writes a bash script for the GROMACS workflow.

2.1.2 GROMACS Workflow

Simulations of crystals are simulations of crystallographic unit cell of proteins. **gmXtal** prepares files for simulations of such system with GROMACS. By default, all simulations are performed with periodic boundary conditions, with the initial box dimensions set to experimental crystallographic unit cell parameters. The preparation module generates the symmetry mates of proteins, DNA chains and cofactors. Water molecules and ions, which are not visible in the experimental structure, still have to be added. After the addition of water and ions, the system has to relax and equilibrate to the desired temperature and pressure. All these steps are performed in GROMACS workflow. The GROMACS workflow includes the following steps: (i) adding hydrogens to the biomolecule with *gmx pdb2gmx* tool; (ii) adding water (solvent) molecules to the system; (iii) adding ions; (iv) running energy minimization; (v) running NVT and NPT equilibration; (vi) and performing a production run. The parameters for the energy minimization, equilibration, and production runs are provided as a part of **gmXtal** package, but the user can also provide own GROMACS parameter files (i.e., **.mdp**) for each of these steps.

Although not all solvent molecules are resolved in a crystal structure, these molecules determine density of the biomolecules in the crystal. Therefore, it is imperative that these missing solvent molecules are added to the unitcell. To determine the number of solvent molecules required to maintain the correct density, and hence unit cell volume, **gmXtal** creates multiple simulation boxes with different numbers of solvent molecules and checks which number preserves the unit cell volume during MD simulations. The number of solvent molecules is controlled with the **-scale** option of the *gmx solvate* routine. While the default value of 0.57 is suitable for creating simulation boxes of solvated proteins with the correct water density, this value may not be suitable for proteins in crystals (see results). Therefore, **gmXtal** probes several values of **-scale** parameter around 0.57 and recommends the optimal value based on an analysis of the MD simulations performed by the check module.

2.1.3 Check Module

The check module of **gmXtal** analyses the time-evolution of the unit cell volume in the production runs of systems with different numbers of added solvent molecules. For each frame the volume is computed as

$$V = a \times b \times c \times \sqrt{1 - \cos(\alpha)^2 - \cos(\beta)^2 - \cos(\gamma)^2 + 2 \cos(\alpha) \cos(\beta) \cos(\gamma)} \quad (5)$$

If the average unit cell volume in a production run is within 0.5% of the experimental value, the system is considered optimal and **gmXtal** will recommend the user to proceed with the corresponding parameters. If the average volume is too far from the experimental volume in all trajectories, **gmXtal** will propose to rerun the whole workflow with a different set of **-scale** parameters for the *gmx solvate* routine. In this case **gmXtal** will output a file with the suggested new input parameters, for the preparation module.

2.1.4 gmXtal Input

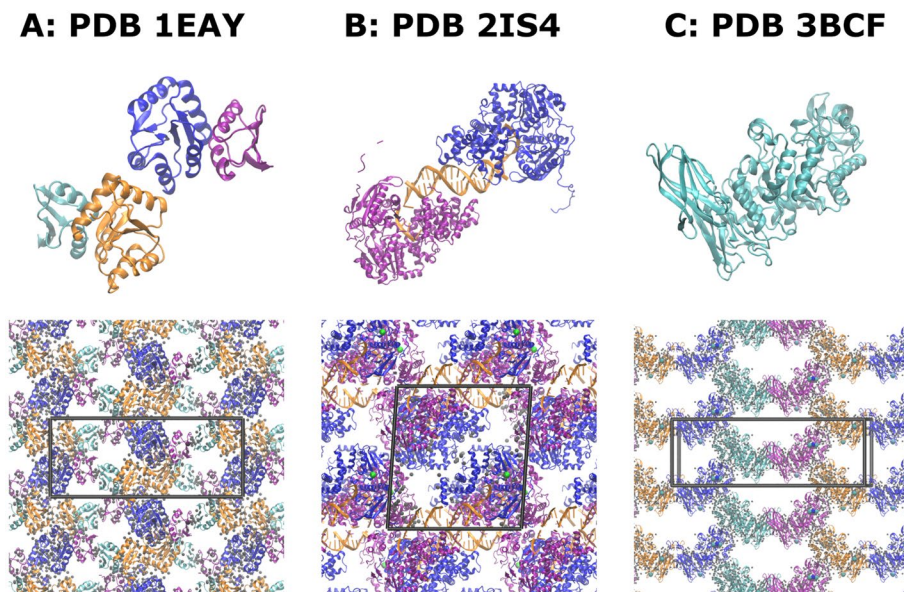
The input required to run **gmXtal** in both preparation and check modes can either be provided interactively via the user interface, or read in from a file in **.yaml** format. If the input is provided interactively via the user-interface, **gmXtal** will save the input parameters locally in **.yaml** format, which can be reused or modified if needed. Examples of **.yaml** files are provided alongside with the script at <https://gitlab.com/pbuslaev/gmxtal>.

2.2 Simulation Setup

To validate the workflow implemented in **gmXtal** we set up crystals for MD simulations of (i) chey-binding (P2) domain of chea in complex with chey from *Escherichia coli* (PDB ID: 1EAY) [38], of (ii) UvrD-DNA-ADPNP ternary complex (PDB ID: 2IS4) [39], and of (iii) alpha-amylase B from *Halothermothrix orenii* (PDB ID: 3BCF) [40]. The 1EAY structure is a hetero-dimer, has a symmetry $P_{2_12_12_1}$, and has two copies of the protein dimer in the asymmetric unit. It also includes 124 structural water molecules. The 2IS4 structure is a homo-dimer that includes protein and DNA, has a symmetry P_{12_11} , and also includes two magnesium structural ions and 118 structural water molecules. The 3BCF structure is a monomer with C_{121} symmetry. The structure contains also four calcium and one sodium structural ions, as well as 256 crystallographic water molecules. The structures of the initial complexes as well as the fully solvated crystals generated by **gmXtal** are shown in Fig. 2.

After the construction of the crystals, these systems were simulated with GROMACS2022 [25, 26] using periodic boundary conditions. The Amber ff99SB-ildn force field was used to model the interactions between the atoms [41]. The parameters used in our simulations are standard parameters recommended for using the AMBER family force fields with the GROMACS software [42]. Water was modeled with the TIP3P water model [43]. To all systems Na^+ and Cl^- ions were added at 0.15 M concentration to neutralize the simulation box. Coulomb interactions were treated with the smooth PME method using a real-space cut-off of 1.0 nm and a grid spacing of 0.16 nm [44, 45]. Van der Waals interactions were modeled with a Lennard–Jones potential that was truncated

Fig. 2 Structures of simulated systems (top row) and crystals build by gmxtal (bottom row). Different chains are shown in different colors. The unit cell is indicated by the black box. Structural waters are shown as gray spheres, ions as colored spheres



at 1 nm and shifted using the Verlet modifier. The temperature was kept constant at 300 K with the v-rescale thermostat [46] using a time constant 0.5 ps^{-1} . The pressure was kept constant at 1 bar with c-rescale barostat [47] using relaxation time of 5.0 ps. Trajectories were calculated using a leap-frog algorithm with a time step of 2 fs. The LINCS algorithm was used to constrain the lengths of bonds involving hydrogens in the protein and DNA [48, 49], while SETTLE was used to constrain the internal degrees of freedom of the water molecules [50]. All production runs were simulated for 20 ns. Prior to the production runs, the energy of all systems was minimized with the steepest descent method, first followed by a 100 ps simulation at constant temperature and volume, and then by a 100 ps simulation at constant temperature and pressure. All parameter files used for these simulations are provided as a part of **gmXtal** at <https://gitlab.com/pbuslaev/gmxtal>.

3 Results

We performed simulations of three systems that have different symmetries and content of the asymmetric unit (Fig. 2). These systems were selected to demonstrate that **gmXtal** is applicable to a wide range of biomolecular complexes and crystallographic symmetries. Thus, we selected (i) a heterodimer and (ii) a protein-DNA complex with two copies of the asymmetric unit in the unit cell and (iii) a large enzyme with four copies of the asymmetric unit in the unit cell. Figure 2 shows the structural organization of the crystals that were build by **gmXtal**. The differences in crystallographic symmetries lead to differences in the empty spaces between

the proteins, which **gmXtal** fills up with water during the preparation phase.

To determine the correct number of solvent molecules, we ran **gmXtal** with five different values of the **-scale** option of the *gmx solvate* program: 0.53, 0.55, 0.57, 0.59, and 0.61. These values determine the initial water density in the crystals. For each system a 20 ns MD trajectory was computed and the time-evolution of the volume was analyzed (Fig. 3).

For systems 2IS4 and 3BCF the experimental volume was reproduced with the **-scale** option set to 0.53. For 1EAY system, the experimental volume was not reproduced fairly well for any of the selected densities of added solvent. Instead, **gmXtal** recommended to prepare new systems with the **-scale** option in the range from 0.53 to 0.55, and run a new set of simulations. After this second iteration, we found that the optimal water density was achieved with the **-scale** option set to 0.54. The three examples suggest that, while the number of added solvent molecules cannot be predicted in advance, **gmXtal** provides a means to estimate that number. For crystals the optimal value for the **-scale** option of *gmx solvate* tool can be different from the recommended value for solvating proteins in water. The crystals created by **gmXtal** provide the starting point for investigating the properties of biomolecular crystals by means of MD simulations, just as Herman Berendsen et al. did, almost four decades ago [6].

4 Conclusion

To summarize, we have presented **gmXtal**, a python-based toolbox for automatically generating hydrated crystals from protein data bank files. We believe that this tool can be useful for performing simulations of crystals as it avoids the

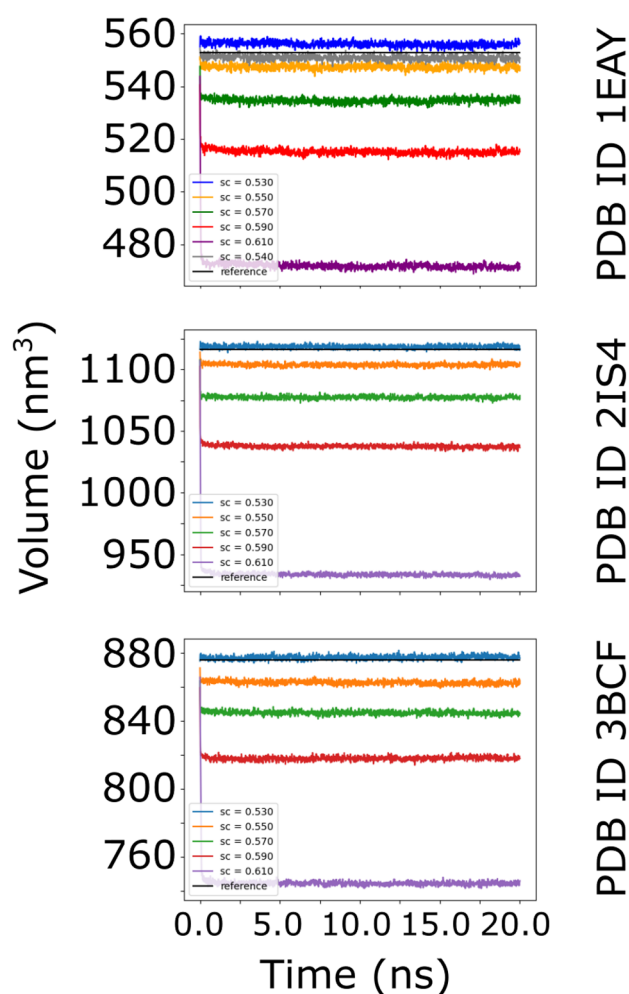


Fig. 3 Time-evolution of the unit cell volume in simulations with varying numbers of water molecules added to the system. The water molecules were added with *gmX solvate* using different values for the *-scale* option: 0.53 — blue curve, 0.55 — orange curve, 0.57 — green curve, 0.59 — red curve, and 0.61 — purple curve. For 2IS4 and 3BCF the optimal *-scale* parameter is 0.53. For 1EAY system 0.53 is quite close to preserving the volume, but *gmXtal* nevertheless suggested running a second iteration with *-scale* values in-between 0.53 and 0.55. After the second iteration, *gmXtal* recommends *-scale* parameter to be set to 0.54 for 1EAY system. The volume evolution of 1EAY at 0.54 value of *-scale* parameter is shown in the top panel with gray color (Color figure online)

tedious procedure of setting up crystal models manually and in particular determining the number of solvent molecules required to maintain the experimental density and unit cell volume. We expect that our tool will pave the way to systematically explore the effects of the crystal environment on the properties of biomolecular systems, including catalytic activity and pK_a s of titratable amino acids, which can be dramatically different between solution and crystal, as for instance in PYP, for which it was suggested that an Arginine is deprotonated in the crystal [51].

Acknowledgements The authors thank Noora Aho, Dmitry Morozov, Vaibhav Modi, and Ilya Sokolovskii for feedback on discussing and using *gmXtal*. This work was supported by the Academy of Finland (Grant 342908, 332743). The simulations were performed on resources provided by the CSC-IT Center for Science, Finland.

Author Contributions PB implemented *gmXtal*. PB and GG designed the research and wrote the manuscript.

Funding Open Access funding provided by University of Jyväskylä (JYU).

Declarations

Competing interests The authors have no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Yu H (1999) Extending the size limit of protein nuclear magnetic resonance. *Proc Natl Acad Sci USA* 96(2):332–334
2. Drenth J (2007) Principles of protein X-ray crystallography. Springer, New York
3. Fischer M (2021) Macromolecular room temperature crystallography. *Q Rev Biophys* 54:1
4. Brändén G, Neutze R (2021) Advances and challenges in time-resolved macromolecular crystallography. *Science* 373(6558):0954
5. Hollingsworth SA, Dror RO (2018) Molecular dynamics simulation for all. *Neuron* 99(6):1129–1143
6. Van Gunsteren W, Berendsen H, Hermans J, Hol W, Postma J (1983) Computer simulation of the dynamics of hydrated protein crystals and its comparison with x-ray data. *Proc Natl Acad Sci USA* 80(14):4315–4319
7. Van Gunsteren W, Berendsen H (1984) Computer simulation as a tool for tracing the conformational differences between proteins in solution and in the crystalline state. *J Mol Biol* 176(4):559–564
8. Anselmi M, Brunori M, Vallone B, Di Nola A (2008) Molecular dynamics simulation of the neuroglobin crystal: comparison with the simulation in solution. *Biophys J* 95(9):4157–4162
9. Amemiya T, Koike R, Fuchigami S, Ikeguchi M, Kidera A (2011) Classification and annotation of the relationship between protein structural change and ligand binding. *J Mol Biol* 408(3):568–584
10. Terada T, Kidera A (2012) Comparative molecular dynamics simulation study of crystal environment effect on protein structure. *J Phys Chem B* 116(23):6810–6818
11. Pande K, Hutchison CD, Groenhof G, Aquila A, Robinson JS, Tenboer J, Basu S, Boutet S, DePonte DP, Liang M (2016) Femtosecond structural dynamics drives the trans/cis isomerization in photoactive yellow protein. *Science* 352(6286):725–729

12. Walser R, Hünenberger PH, Gunsteren WF (2001) Comparison of different schemes to treat long-range electrostatic interactions in molecular dynamics simulations of a protein crystal. *Proteins* 43(4):509–519
13. Kriz K, Schmidt L, Andersson AT, Walz M-M, Spoel D (2023) An imbalance in the force: the need for standardized benchmarks for molecular simulation. *J Chem Inf Model* 63(2):412–431
14. Cerutti DS, Freddolino PL, Duke RE Jr, Case DA (2010) Simulations of a protein crystal with a high resolution x-ray structure: evaluation of force fields and water models. *J Phys Chem B* 114(40):12811–12824
15. Schmidt L, Spoel D, Walz M-M (2022) Probing phase transitions in organic crystals using atomistic md simulations. *ACS Phys Chem Au* 3(1):84–93
16. Janowski PA, Liu C, Deckman J, Case DA (2016) Molecular dynamics simulation of triclinic lysozyme in a crystal lattice. *Protein Sci* 25(1):87–102
17. Janowski PA, Cerutti DS, Holton J, Case DA (2013) Peptide crystal simulations reveal hidden dynamics. *J Am Chem Soc* 135(21):7938–7948
18. Walser R, Hünenberger PH, Gunsteren WF (2002) Molecular dynamics simulations of a double unit cell in a protein crystal: volume relaxation at constant pressure and correlation of motions between the two unit cells. *Proteins* 48(2):327–340
19. Ma P, Xue Y, Coquelle N, Haller JD, Yuwen T, Ayala I, Mikhailovskii O, Willbold D, Colletier J-P, Skrynnikov NR (2015) Observing the overall rocking motion of a protein in a crystal. *Nat Commun* 6(1):8361
20. Kurauskas V, Izmailov SA, Rogacheva ON, Hessel A, Ayala I, Woodhouse J, Shilova A, Xue Y, Yuwen T, Coquelle N (2017) Slow conformational exchange and overall rocking motion in ubiquitin protein crystals. *Nat Commun* 8(1):145
21. Jo S, Kim T, Iyer VG, Im W (2008) Charmm-gui: a web-based graphical user interface for charmm. *J Comput Chem* 29(11):1859–1865
22. Sommer B, Dingersen T, Gamroth C, Schneider SE, Rubert S, Krüger J, Dietz K-J (2011) Cellmicrocosmos 2.2 membraneeditor: a modular interactive shape-based software approach to solve heterogeneous membrane packing problems. *J Chem Inf Model* 51(5):1165–1182
23. Martínez L, Andrade R, Birgin EG, Martínez JM (2009) Packmol: a package for building initial configurations for molecular dynamics simulations. *J Comput Chem* 30(13):2157–2164
24. Cerutti D, Case D (2023) Simulating crystals with the amber molecular dynamics package. <https://ambermd.org/tutorials/advanced/tutorial13/XtalTutorial1.html>. Accessed 28 April 2023
25. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E (2015) Gromacs: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1:19–25
26. Páll S, Zhmurov A, Bauer P, Abraham M, Lundborg M, Gray A, Hess B, Lindahl E (2020) Heterogeneous parallelization and acceleration of molecular dynamics simulations in Gromacs. *J Chem Phys* 153(13):134110
27. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. *Nucleic Acids Res* 28(1):235–242
28. Westbrook JD, Young JY, Shao C, Feng Z, Guranovic V, Lawson CL, Vallat B, Adams PD, Berrisford JM, Bricogne G (2022) Pdbx/mmCIF ecosystem: foundational semantic tools for structural biology. *J Mol Biol* 434(11):167599
29. Hamelryck T, Manderick B (2003) Pdb file parser and structure class implemented in python. *Bioinformatics* 19(17):2308–2310
30. Ginkel G, Pravda L, Dana JM, Varadi M, Keller P, Anyango S, Velankar S (2021) Pdbecif: an open-source MMCIF/CIF parsing and processing package. *BMC Bioinform* 22(1):1–7
31. Michaud-Agrawal N, Denning EJ, Woolf TB, Beckstein O (2011) Mdanalysis: a toolkit for the analysis of molecular dynamics simulations. *J Comput Chem* 32(10):2319–2327
32. Gowers RJ, Linke M, Barnoud J, Reddy TJ, Melo MN, Seyler SL, Domanski J, Dotson DL, Buchoux S, Kenney IM (2016) Mdanalysis: a python package for the rapid analysis of molecular dynamics simulations. In: *Proceedings of the 15th Python in science conference*. SciPy Austin, TX, vol 98, p 105
33. Eastman P, Swails J, Chodera JD, McGibbon RT, Zhao Y, Beauchamp KA, Wang L-P, Simmonett AC, Harrigan MP, Stern CD (2017) Openmm 7: rapid development of high performance algorithms for molecular dynamics. *PLoS Comput Biol* 13(7):1005659
34. Šali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234(3):779–815
35. Olsson MH, Sørensen CR, Rostkowski M, Jensen JH (2011) Propka3: consistent treatment of internal and surface residues in empirical pKa predictions. *J Chem Theory Comput* 7(2):525–537
36. Gokcan H, Isayev O (2022) Prediction of protein pKa with representation learning. *Chem Sci* 13(8):2462–2474
37. Wojdyr M (2022) GEMMI: a library for structural biology. *J Open Source Softw* 7(73):4200
38. McEvoy MM, Hausrath AC, Randolph GB, Remington SJ, Dahlquist FW (1998) Two binding modes reveal flexibility in kinase/response regulator interactions in the bacterial chemotaxis pathway. *Proc Natl Acad Sci* 95(13):7333–7338
39. Lee JY, Yang W (2006) UvrD helicase unwinds DNA one base pair at a time by a two-part power stroke. *Cell* 127(7):1349–1360
40. Tan T-C, Mijts BN, Swaminathan K, Patel BK, Divne C (2008) Crystal structure of the polyextremophilic α -amylase AmyB from halothermothrix orenii: details of a productive enzyme-substrate complex and an n domain with a role in binding raw starch. *J Mol Biol* 378(4):852–870
41. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE (2010) Improved side-chain torsion potentials for the amber ff99sb protein force field. *Proteins* 78(8):1950–1958
42. Sorin EJ, Pande VS (2005) Exploring the helix-coil transition via all-atom equilibrium ensemble simulations. *Biophys J* 88(4):2472–2493
43. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79(2):926–935
44. Darden T, York D, Pedersen L (1993) Particle mesh ewald: an nlog(n) method for Ewald sums in large systems. *J Chem Phys* 98(12):10089–10092
45. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) A smooth particle mesh Ewald method. *J Chem Phys* 103(19):8577–8593
46. Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. *J Chem Phys* 126(1):014101
47. Bernetti M, Bussi G (2020) Pressure control using stochastic cell rescaling. *J Chem Phys* 153(11):114107
48. Hess B, Bekker H, Berendsen HJ, Fraaije JG (1997) Lincs: a linear constraint solver for molecular simulations. *J Comput Chem* 18(12):1463–1472
49. Hess B (2008) P-lincs: a parallel linear constraint solver for molecular simulation. *J Chem Theory Comput* 4(1):116–122
50. Miyamoto S, Kollman PA (1992) Settle: an analytical version of the shake and rattle algorithm for rigid water models. *J Comput Chem* 13(8):952–962
51. Yamaguchi S, Kamikubo H, Kurihara K, Kuroki R, Niimura N, Shimizu N, Yamazaki Y, Kataoka M (2009) Low-barrier hydrogen bond in photoactive yellow protein. *Proc Natl Acad Sci USA* 106(2):440–444