Data in Brief 54 (2024) 110370

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Data on molecular docking and molecular dynamics targeting *Mycobacterium tuberculosis* shikimic acid kinase

Hinata Taniguchi, Shuhei Kawamoto, Kohei Monobe, Shunsuke Aoki*

Department of Bioscience and Bioinformatics, School of Computer Science and Systems Engineering, Kyushu Institute of Technology, 680-4 Kawazu, Iizuka-shi, Fukuoka 820-8502, Japan

ARTICLE INFO

Article history:
Received 28 September 2023
Revised 28 December 2023
Accepted 21 March 2024
Available online 26 March 2024

Dataset link: Are you linking to original data/code reported in your research, or are you linking to reference data/code which is supporting your article? Any links to data or code will be published alongside your (Original data)

Keywords:
Docking simulation
In silico screening
Molecular dynamics
Tuberculosis

ABSTRACT

We have previously performed a hierarchical *in silico* screening of a *Mycobacterium tuberculosis* shikimic acid kinase [1]. Specifically, 11 compounds were screened from a library of 154,118 compounds provided by ChemBridge [2] using UCSF DOCK [3] and the GOLD [4] program in the first and second steps, respectively. Molecular dynamic simulations were further performed on compound 2 (2-[(5Z)-5-(1-benzyl-5bromo-2-oxoindol-3-(5Z)-5-(1-benzyl-5-bromo-2-oxoindol-3-(5Z)-4-oxo-2 ylidene)-4oxo-2-sulfanylidene-1,3-thiazolidin-3-yl] acetic acid), which showed antimicrobial efficacy. These processes yielded ligand docking scores and trajectories. In this data article, we have added solvent-accessible surface area and PCA analyses, which were calculated from the raw docking scores and trajectories. Data obtained from molecular docking and molecular dynamic simulations are useful in two ways:

(1) Further support for previous work
(2) Provides a stepping stone for experimental scientists to conduct *in silico* studies and research ideas for other drug discovery researchers and computational biologists.

DOI of original article: 10.1016/j.tube.2023.102362
* Corresponding author.
E-mail address: aokis@bio.kyutech.ac.jp (S. Aoki).

https://doi.org/10.1016/j.dib.2024.110370

We believe that this article will provide an opportunity to develop new *Mycobacterium tuberculosis* therapeutics through searching for analogs and inhibitors against new targets.

Specifications Table

Subject	Computational biology
Specific subject area	Docking simulation and molecular dynamics simulation
Data format	Raw, Analyzed
Type of data	Table, Figure
Data collection	We screened 154,118 compounds downloaded from Ressource Parisienne en BioinformatiqueStructurale, ChemBridge's web-based chemical structure database by filtering for absorption, distribution, metabolism, excretion, and toxicity. We obtained the X-ray crystal structure of *Mycobacterium tuberculosis* shikimic acid kinase (MtSK) for molecular docking from the Protein Data Bank database [5] (PDB ID: 2IYQ). From the obtained structure, shikimic acid and water molecules were removed and hydrogen atoms and charges were added. Furthermore, a stable protein structure was obtained using the Energy Minimize module of MOE [6]. We used DOCK as the first step in a hierarchical screening to screen compounds from a compound library. First, we generated the protein structure surface using the DMS module of the UCSF chimera [7] to perform DOCK. The corresponding spheres were then placed in protein pockets using the SPHGEN package. A 5 Å cubic box was created around the spheres using the SHOWBOX package to limit the extent to which grids were generated, which were spaced 0.3 Å apart. After the DOCK calculations, further simulations were performed using GOLD for the 2000 compounds with DOCK scores below −45.34 kcal/mol. The center coordinates used were x: 33.808, y: −35.244, z: 29.997, using default parameters. Ten conformations were generated for each compound using MOE for ligands with ≥75 GOLD score. We reran the GOLD program on the compounds that generated the conformations, and the average score for the conformations was used as the final GOLD score. Finally, the antimicrobial efficacy of compounds with ≥75 GOLD score were tested using *Mycobacterium smegmatis*. Molecular dynamic simulations were performed for the antibacterial compound 2 (2-[(5Z)-5-(1-benzyl-5-bromo-2oxoindol-3-ylidene)-4-oxo-2-sulfanylidene-1,3-thiazolidin-3yl]acetic acid) using GROMACS 2021.4 [8]. The simulation system was created using Solution Builder on the CHARMM-GUI server [9,10] and the compound conformation output by GOLD. The system was solvated by the TIP3P water model, and sodium and chloride ions were added to balance the charges. The resulting system is a 66 Å side cube containing 24,051 water molecules, 22 sodium ions, and 30 chloride ions. Energies were minimized in up to 5,000 steps by the steepest descent method. Equilibration was performed for 100 ps under standard NVT and NPT conditions. The temperature and pressure were adjusted to 310 K and 1 bar, respectively. The final MD simulation was performed for 100 ns with a time step of 2 fs under NPT conditions.
Data source location	• Institution: Kyushu Institute of Technology • City/Town/Region: Kawazu/Iizuka/Fukuoka • Country: Japan
Data accessibility	Repository name: Molecular dynamics data for compound and *Mycobacterium tuberculosis* shikimic acid kinase targets Data identification number: 10.17632/725kx83gts.2 Direct URL to data: https://data.mendeley.com/datasets/725kx83gts/2
Related research article	This data article supports the following paper we have published: S. Kawamoto, C. Hori, H. Taniguchi, S. Okubo, S. Aoki, Identification of novel antimicrobial compounds targeting Mycobacterium tuberculosis shikimate kinase using in silico hierarchical structure-based drug screening, Tuberculosis. 141 (2023) 102362. https://doi.org/10.1016/j.tube.2023.102362.

1. Value of the Data

- This dataset gives ideas on processing and analysis for molecular docking and molecular dynamics simulations
- This dataset will not only benefit drug discovery researchers and computational biologists, but also provide a stepping stone for experimental scientists to perform *in silico* analyses.
- This data will help researchers find inhibitors of other target proteins
- Parameter files on molecular dynamics can contribute to efficient therapeutic drug design
- This dataset provides the basic data necessary for *in silico* drug design infrastructure to advance analog pairing searches and organic synthesis development and effectively use the data.

2. Data Description

This data article presents binding data for compound 2 (2-[(5Z)-5-(1-benzyl-5-bromo2-oxoindol-3-ylidene)-4-oxo-2-sulfanylidene-1,3-thiazolidin-3-yl]acetic acid) to *Mycobacterium tuberculosis* shikimic acid kinase (MtSK) not published in our previous paper. Fig. 1 shows the sphere and box used to calculate UCSF Dock. Table 1 shows the Grid_vdw, Grid_es, and GOLD Score (P450) for compounds 1–11, which were not shown in the original paper. Compounds are oriented within this sphere. However, since the sphere must be researcher-selected, this information may be useful for other studies. Fig. 2 shows the density, pressure, and temperature of the system after equilibration of the compound 2–MtSK complex before MD simulation. Fig. 3 shows the solvent-accessible surface area (SASA) [11] mutations in compound 2 during MD. Fig. 4 shows the structural mutations identified using PCA. The input and output files for the MD calculations were stored in a data repository (https://data.mendeley.com/datasets/725kx83gts/2). We divided the folders by minimization, equilibration, and production and stored the shell scripts and output files for each process. The video visualizing the MD trajectory using VMD [12] was stored as 100nsMD.mp4 in the MD_Video file (https://data.mendeley.com/datasets/725kx83gts/2).

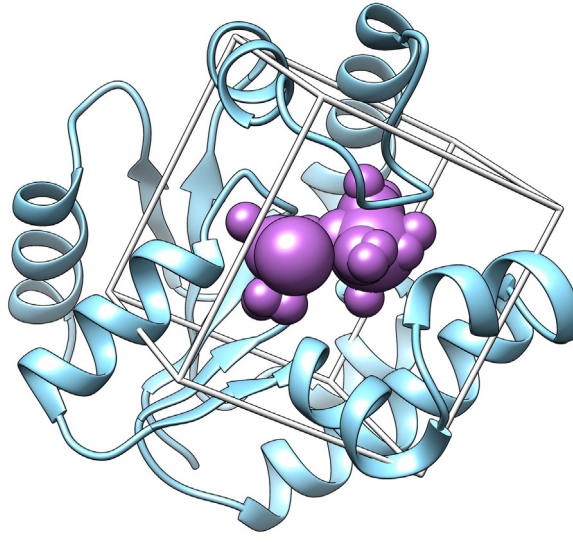

Fig. 1. The position of the probe sphere indicating the orientation of the compound and box representing the region where the grid is calculated.

Table 1
Grid_vdw Score and Grid_es Score for compounds and MtSK calculated by DOCK and GOLD Score (P450) calculated by GOLD.

Name	Chembridge ID	Grid_vdw	Grid_es	GOLD Score (P450)
compound 1	7951857	-34.760082	-10.883127	87.757
compound 2	6680698	-32.720776	-15.487597	85.469
compound 3	6345099	-38.167774	-8.109470	81.849
compound 4	7606743	-29.117647	-17.205578	82.318
compound 5	7983474	-32.058479	-14.042460	78.559
compound 6	5728840	-35.901741	-10.531392	79.375
compound 7	9022617	-35.976479	-13.731997	78.85
compound 8	7442550	-37.342285	-10.506849	77.707
compound 9	7366078	-38.750999	-7.274676	77.48
compound 10	6566230	-36.423145	-14.831291	77.325
compound 11	7763637	-36.390656	-17.316345	78.177

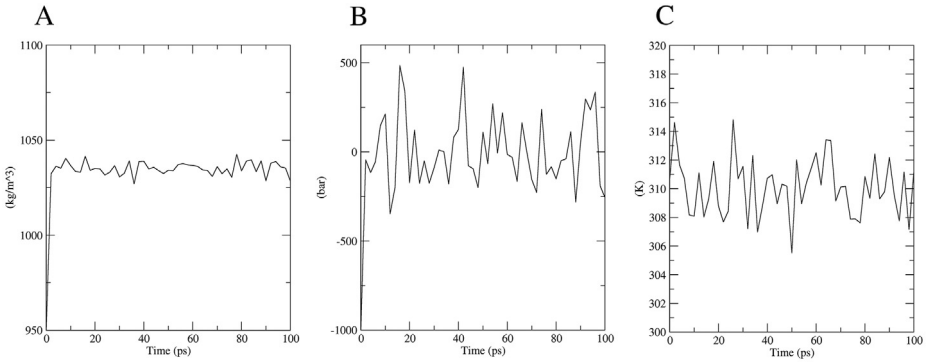

Fig. 2. Parameters after protein-ligand complex equilibration before MD run showing density gradient, pressure and temperature levels. (A) density, (B) pressure, (C) temperature.

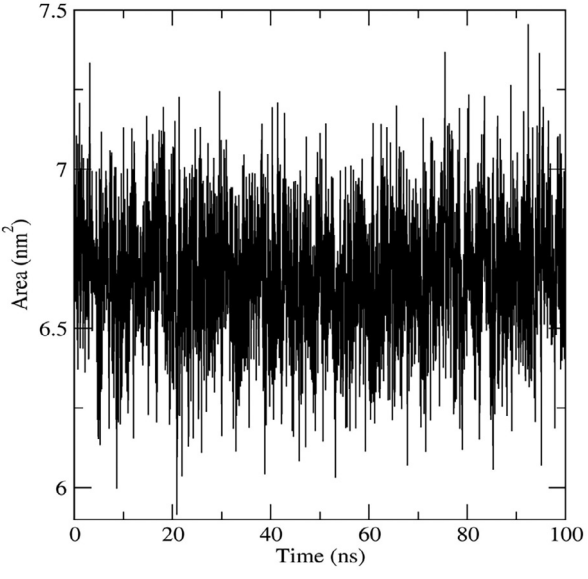

Fig. 3. Results of SASA analysis. The black line represents the area where the compound can contact the solvent during the MD simulation.

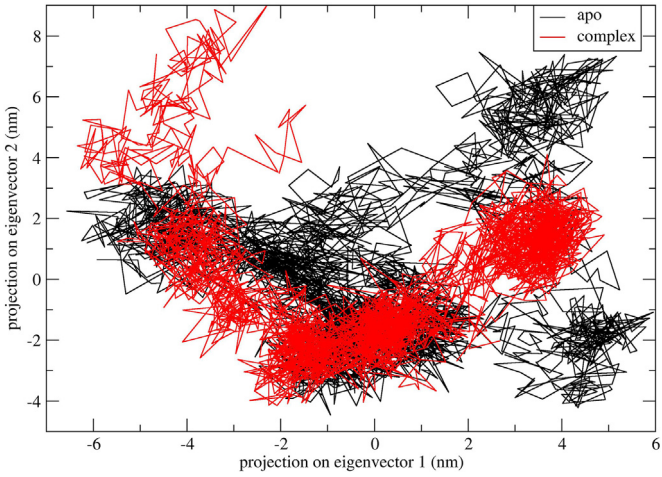

Fig. 4. Projections of trajectories of MtSK on the first two PCA eigenvectors of the trajectory. The black line shows free MtSK and the red line shows the results of PCA analysis of MtSK complexed with compound 2 (2-[(5Z)-5-(1-benzyl-5-bromo-2-oxoindol-3-ylidene)-4-oxo-2-sulfanilidene-1,3-thiazolidin-3-yl]acetic acid).

3. Experimental Design, Materials and Methods

MtSK is essential for the survival of *Mycobacterium tuberculosis* and was investigated in the paper [13] . The three-dimensional structure of MtSK, which is essential for the survival of *Mycobacterium tuberculosis*, was obtained from the Protein Data Bank database (PDB ID: 2IYQ). The MtSK–ADP complex was prepared by removing shikimic acid and water molecules from the obtained structure. The stable protein structure was obtained by adding hydrogen atoms and charges using the UCSF chimera software and minimizing the hydrogen atom energy with the MOE's energy minimization module. The first screening with 154,118 compounds was performed using the UCSF DOCK program from our laboratory library. The protein structure surface was generated using the DMS module of UCSF chimera in DOCK. The SPHGEN package in DOCK was used to place the corresponding spheres in the protein pocket. To limit the extent to which the grids were generated, a 5 Å cube box was created around the spheres using the SHOWBOX package. The grid spacing was 0.3 Å. The GOLD program was used to simulate 2000 compounds with DOCK scores of −45.34 kcal/mol or less, considering ligand flexibility. The center coordinates used were x: 33.808, y: −35.244, and z: 29.997 using default parameters. Ten conformations were generated for each compound using MOE for ligands with ≥75 GOLD score. The GOLD program was run again for the compounds that generated conformations, and the average score for the conformations was used as the final GOLD score. Compounds with ≥75 final GOLD score were validated for antimicrobial efficacy in an assay with M. smegmatis. Molecular dynamic (MD) simulations were performed for the *in silico* analysis of compound 2 (2-[(5Z)-5-(1benzyl-5-bromo-2-oxoindol-3-ylidene)-4-oxo-2-sulfanylidene-1,3-thiazolidin-3yl]acetic acid), which showed antibacterial activity in this assay. MD simulations were performed for further *in silico* analysis using GROMACS 2021.4. The simulation system was created using Solution Builder on a CHARMM-GUI server with the compound conformation outputs from GOLD. The system was solvated using the TIP3P water model, and the necessary ions were added to balance the charges. The resulting system is a 66 Å cube box with one protein, one ligand, 24051 water molecules, 22 sodium ions, and 30 chloride ions. The energies were minimized in up to 5,000 steps using the steepest descent method. Equilibration was performed for 100 ps under the standard NVT and NPT conditions. The temperature and pressure were adjusted to 310 K and 1 bar, respectively. The final MD simulation was performed for 100 ns with a time step of 2 fs under

NPT conditions using the GROMACS module for trajectory analysis of the MD simulation and VMD for visualization.

Limitations

Not applicable.

Ethics Statement

This data article does not include studies on human subjects, animal studies, or studies using data collected from social media platforms.

Data Availability

Are you linking to original data/code reported in your research, or are you linking to reference data/code which is supporting your article? Any links to data or code will be published alongside your (Original data) (Mendeley Data).

CRediT Author Statement

Hinata Taniguchi: Methodology, Validation, Writing – original draft; **Shuhei Kawamoto:** Conceptualization; **Kohei Monobe:** Methodology; **Shunsuke Aoki:** Conceptualization, Supervision.

Acknowledgements

Funding: This work was supported by a Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan [grant number 26460145].

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] S. Kawamoto, C. Hori, H. Taniguchi, S. Okubo, S. Aoki, Identification of novel antimicrobial compounds targeting Mycobacterium tuberculosis shikimate kinase using in silico hierarchical structure-based drug screening, Tuberculosis 141 (2023) 102362, doi:10.1016/j.tube.2023.102362.

[2] N. Lagarde, E. Goldwaser, T. Pencheva, D. Jereva, I. Pajeva, J. Rey, P. Tuffery, B.O. Villoutreix, M.A. Miteva, A free web-based protocol to assist structure-based virtual screening experiments, Int. J. Mol. Sci. 20 (2019) 4648, doi:10.3390/ijms20184648.

[3] P.T. Lang, S.R. Brozell, S. Mukherjee, E.F. Pettersen, E.C. Meng, V. Thomas, R.C. Rizzo, D.A. Case, T.L. James, I.D. Kuntz, DOCK 6: combining techniques to model RNA-small molecule complexes, RNA 15 (2009) 1219–1230, doi:10.1261/rna.1563609.

[4] G. Jones, P. Willett, R.C. Glen, A.R. Leach, R. Taylor, Development and validation of a genetic algorithm for flexible docking, J. Mol. Biol. 267 (1997) 727–748, doi:10.1006/jmbi.1996.0897.

[5] H.M. Berman, T. Battistuz, T.N. Bhat, W.F. Bluhm, P.E. Bourne, K. Burkhardt, Z. Feng, G.L. Gilliland, L. Iype, S. Jain, P. Fagan, J. Marvin, D. Padilla, V. Ravichandran, B. Schneider, N. Thanki, H. Weissig, J.D. Westbrook, C. Zardecki, The protein data bank, Acta Crystallogr. D. Biol. Crystallogr. 58 (2002) 899–907, doi:10.1107/S0907444902003451.

[6] S. Vilar, G. Cozza, S. Moro, Medicinal chemistry and the Molecular Operating Environment (MOE): application of QSAR and molecular docking to drug discovery, Curr. Top. Med. Chem. 8 (2008) 1555–1572, doi:10.2174/156802608786786624.
[7] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera–a visualization system for exploratory research and analysis, J. Comput. Chem. 25 (2004) 1605–1612, doi:10.1002/JCC.20084.
[8] M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, E. Lindah, GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers, SoftwareX 1–2 (2015) 19–25, doi:10.1016/J.SOFTX.2015.06.001.
[9] J. Lee, X. Cheng, J.M. Swails, M.S. Yeom, P.K. Eastman, J.A. Lemkul, S. Wei, J. Buckner, J.C. Jeong, Y. Qi, S. Jo, V.S. Pande, D.A. Case, C.L. Brooks, A.D. MacKerell, J.B. Klauda, W. Im, CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field, J. Chem. Theory. Comput. 12 (2016) 405–413, doi:10.1021/acs.jctc.5b00935.
[10] S. Jo, X. Cheng, J. Lee, S. Kim, S.J. Park, D.S. Patel, A.H. Beaven, K. Il Lee, H. Rui, S. Park, H.S. Lee, B. Roux, A.D. MacKerell, J.B. Klauda, Y. Qi, W. Im, CHARMM-GUI 10 years for biomolecular modeling and simulation, J. Comput. Chem. 38 (2017) 1114–1124, doi:10.1002/jcc.24660.
[11] F. Eisenhaber, P. Lijnzaad, P. Argos, C. Sander, M. Scharf, The double cubic lattice method: efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies, J. Comput. Chem. 16 (1995) 273–284, doi:10.1002/JCC.540160303.
[12] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, J. Mol. Graph. 14 (1996) 33–38, doi:10.1016/0263-7855(96)00018-5.
[13] J. Pereira, I. Vasconcelos, J. Oliveira, R. Caceres, W. de Azevedo, L. Basso, D. Santos, Shikimate kinase: a potential target for development of novel antitubercular agents, Curr. Drug Targets 8 (2007) 459–468, doi:10.2174/138945007780059013.