SUPPLEMENTARY MATERIAL

Structure modelling of the novel AcDODA enzyme

Structure homology of the DODA enzyme from *A. cylindrica* was computed by the SWISS-MODEL (Swiss Institute of Bioinformatics, Biozentrum, University of Basel, Switzerland) homology server (Bienert et al., 2017; Waterhouse et al., 2018) which relies on ProMod3, an inhouse comparative modelling engine based on OpenStructure (Biasini et al., 2013). The SWISS-MODEL template library (SMTL version 2020-03-04, PDB release 2020-02-28) was searched with BLAST (Camacho et al., 2009) and HHBlits (Remmert, Biegert, Hauser, & Söding, 2012) for evolutionary related structures matching the target sequence.

ProMod3 extracts initial structural information from the selected template structure. Insertions and deletions, as defined by the sequence alignment, are resolved by first searching for viable candidates in a structural database. Final candidates are then selected using statistical potentials of mean force scoring methods. If no candidates can be found, a conformational space search is performed using Monte Carlo techniques. Non-conserved side chains are modelled using the 2010 backbone-dependent rotamer library from the Dunbrack group (Shapovalov & Dunbrack, 2011). The optimal configuration of rotamers is estimated using the graph-based TreePack algorithm (McPake, Murray, & Sandford, 2009) by minimising the SCWRL4 energy function (Krivov, Shapovalov, & Dunbrack, 2009). As a final step, small structural distortions, unfavourable interactions or clashes introduced during the modelling process are resolved by energy minimisation. ProMod3 uses the OpenMM library (Eastman et al., 2017) to perform the computations and the CHARMM27 force field (Mackerell, Feig, & Brooks, 2004) for parameterisation.

In SWISS-MODEL, the quaternary structure annotation of the template is used to model the target sequence in its oligomeric form. The method used is based on a supervised machine learning algorithm, Support Vector Machines (SVM), which combines interface conservation, structural clustering, and other template features to provide a quaternary structure quality estimate (QSQE) (Bertoni, Kiefer, Biasini, Bordoli, & Schwede, 2017). The QSQE score is only computed if it is possible to build an oligomer and only for the top ranked templates. In this model, the in silico results were corroborated by molecular characterization of the expressed and purified proteins.

The GMQE and QMEAN parameters were determined to ascertain enough model quality. GMQE provides a quality estimation which combines properties from the target-template alignment and the template search method. The resulting GMQE score is expressed as a number between 0 and 1, reflecting the expected accuracy of a model built with that alignment and template and the coverage of the target. Higher numbers indicate higher reliability. The QMEAN parameter is a composite estimator based on different geometrical properties and provides both global (entire structure) and local (per residue) absolute quality estimates on the basis of the single model used (Benkert, Biasini, & Schwede, 2011). The QMEAN score provides an estimate of the "degree of nativeness" of the structural features observed in the model on a global scale. It indicates whether the model is comparable to what one would expect from experimental structures of similar size. If QMEAN scores between zero and -4.0 indicate good agreement between the model structure and experimental structures of similar size (higher quality around 0.0). Scores of -4.0 or below are indicative of models of low quality.

QMQE and QMEAN values obtained to determine AcDODA as a homodimer related to a dioxygenase enzyme characterized from *Nostoc punctiforme* PCC 73102 were 0.85 and -0.28, respectively.

REFERENCES

Benkert, P., Biasini, M., & Schwede, T. (2011). Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*, *27*(3), 343–350. https://doi.org/10.1093/bioinformatics/btq662

Bertoni, M., Kiefer, F., Biasini, M., Bordoli, L., & Schwede, T. (2017). Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-09654-8

Biasini, M., Schmidt, T., Bienert, S., Mariani, V., Studer, G., Haas, J., … Schwede, T. (2013). OpenStructure: An integrated software framework for computational structural biology. *Acta Crystallographica Section D: Biological Crystallography*, *69*(5), 701–709. https://doi.org/10.1107/S0907444913007051

Bienert, S., Waterhouse, A., De Beer, T. A. P., Tauriello, G., Studer, G., Bordoli, L., & Schwede, T. (2017). The SWISS-MODEL Repository-new features and functionality. *Nucleic Acids Research*, *45*(D1), D313–D319. https://doi.org/10.1093/nar/gkw1132

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*(1), 421. https://doi.org/10.1186/1471-2105-10-421

Eastman, P., Swails, J., Chodera, J. D., McGibbon, R. T., Zhao, Y., Beauchamp, K. A., … Pande, V. S. (2017). OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Computational Biology*, *13*(7). https://doi.org/10.1371/journal.pcbi.1005659

Krivov, G. G., Shapovalov, M. V., & Dunbrack, R. L. (2009). Improved prediction of protein side-chain conformations with SCWRL4. *Proteins: Structure, Function and Bioinformatics*, *77*(4), 778–795. https://doi.org/10.1002/prot.22488

Mackerell, A. D., Feig, M., & Brooks, C. L. (2004). Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulation. *Journal of Computational Chemistry*, *25*(11), 1400–1415. https://doi.org/10.1002/jcc.20065

McPake, C. B., Murray, C. B., & Sandford, G. (2009). Epoxidation of alkenes using HOF·MeCN by a continuous flow process. *Tetrahedron Letters*, *50*(15), 1674–1676. https://doi.org/10.1016/j.tetlet.2008.12.073

Remmert, M., Biegert, A., Hauser, A., & Söding, J. (2012). HHblits: Lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nature Methods*, *9*(2), 173–175. https://doi.org/10.1038/nmeth.1818

Shapovalov, M. V, & Dunbrack, R. L. (2011). A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions. *Structure*, *19*(6), 844–858. https://doi.org/10.1016/j.str.2011.03.019

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., … Schwede, T. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research*, *46*(W1), W296–W303. https://doi.org/10.1093/nar/gky427