

Synthesis and Computational Analysis of Novel IspF Inhibitors

Daniel Harper

Department of Chemistry, Point Loma Nazarene University, San Diego, California 92106

ABSTRACT: A potent inhibitor of 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (IspF) has the potential to serve as a new class of antibiotic. In this project, we explored the development of such an inhibitor using a mixed experimental/computational approach. A promising lead was identified using chemical intuition and previously published research, but it failed to show significant activity toward IspF. The computational results explained this unexpected lack of inhibition. The lead compound cannot adopt the anticipated bidentate binding motif without substantial steric and angle strain. Modifications to reduce these sources of strain are necessary if a potential drug candidate is to be identified. The computational results also provided parameters for the zinc atom in IspF's active site in the trigonal bipyramidal geometry. These parameters will allow for the active site to be probed using molecular dynamics calculations.

Traditional medicinal chemistry first selects a biological target and then uses a combination of high-throughput screening (HTS) and structure activity relationship (SAR) to identify chemical agents which are active against it. HTS is an approach where "leads" are identified by testing many candidate molecules with the expectation that only a small number will yield significant results. Its practicality is often aided by automation. SAR takes place after an initial molecule, or "lead," is identified. In this process, the lead molecule undergoes a series of slight modifications to produce a library of derivative compounds. If the modifications improve the lead compound's activity, they are retained through further iterations. The SAR process can be repeated as many times as necessary to produce a compound with sufficient activity to serve as a drug candidate.

In this study, 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (IspF) was selected as the target. This enzyme is used by pathogens in the synthesis of isoprenoids, an essential class of biomolecule. Meanwhile, mammalian cells also produce isoprenoids, but using a distinctly different pathway. Therefore, inhibiting IspF would selectively deprive these pathogens of isoprenoids, leading to antibiotic properties.

IspF contains a zinc atom in its active site. Therefore, a compound which strongly chelates zinc has the potential to serve as a potent inhibitor. Using this chemical intuition, the molecule shown in Figure 1 was selected as a lead, eliminating the need for HTS. Both nitrogen atoms are good zinc binding groups, allowing it to act as a bidentate ligand. Additionally, previous research has established that this compound will adopt a bidentate binding motif in solution with free zinc ions.¹

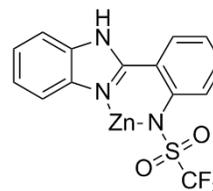


Figure 1. The selected lead compound binding zinc through both nitrogen atoms.

This study also benefitted from the application of computational methodologies. These are particularly valuable in situations where experimental methods are excessively time-intensive, cost prohibitive, or physically impractical.² They can also provide insight regarding what occurs at the atomic scale and inform more efficient drug development. One common application of computational methods is molecular dynamics (MD). In MD, modelling of the atomic realm is done classically. Covalent bonds, electrostatic interactions, Van der Waals forces, etc. are treated with a classical equation.³ Practically, this is often done using force fields. The AMBER force field is particularly common for systems involving proteins. It predicts the energy potential of a conformation according to the following equation:³

$$V_{amber} = \sum_i^{n_{bonds}} b_i (r_i - r_{i,eq})^2 + \sum_i^{n_{angles}} a_i (\theta_i - \theta_{i,eq})^2 + \sum_i^{n_{dihedrals}} \sum_n^{n_{i,max}} \left(\frac{V_{i,n}}{2} \right) [1 + \cos(n\phi_i - \gamma_{i,n})] + \sum_{i < j}^{n_{atoms}} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + \sum_{i < j}^{n_{atoms}} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad (1)$$

The terms in the equation correspond to bond, angles, dihedral angles, the Pauli exclusion force, and the Coulomb interaction, respectively. The constants in this equation, called parameters, frequently benefit from transferability between similar systems.³ For example, the value of b_i is known for the carbon-oxygen double bond in alanine and

it can be reasonably applied to most proteins without further modification. In many applications, utilizing the transferability of these parameters has been found to be sufficiently accurate.⁴

However, transferability does not hold for all cases, particularly metal ions. The parameters for metal containing systems must be resolved for each system individually.⁵ This calls for a deeper level of theory.

It has been found that quantum mechanics can accurately predict suitable parameters from metal ions in a wide variety of situations. More specifically, the MCPB.py workflow has been developed specifically for this application.⁶ Once these parameters are found, they can be applied in classical molecular dynamics simulations.

In this experiment, 4 similar molecules were synthesized and tested for activity against IspF. This follows the approach used in traditional medicinal chemistry. Then, the zinc-containing active site was analyzed using the MCPB.py workflow. This provided accurate predictions of the active site's geometry and provided the parameters for future MD simulations.

Experimental Materials and Methods

Chemical synthesis was carried out according to the scheme depicted in Figure 2.

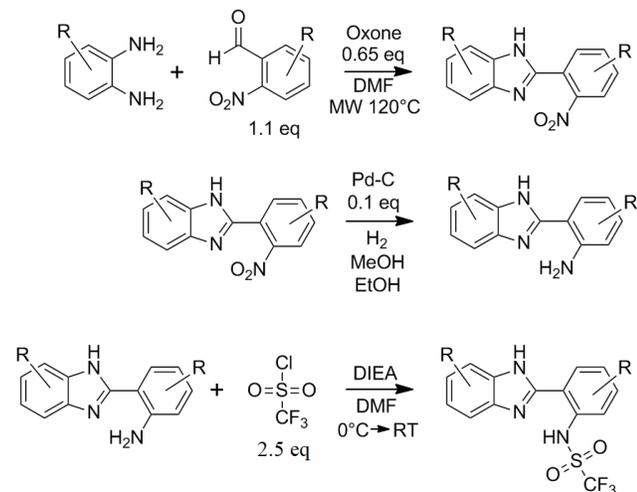


Figure 2. General scheme for chemical synthesis of derivatives. The “R” represents a variable group which may be attached at any point of the ring.

In the first step, diamine and aldehyde compounds were combined via reductive amination using oxone as an oxidant. Both starting materials are widely commercially available with various substituents in the “R” positions. This allows for a large number of derivative molecules to be produced with little or no adjustments to the synthetic pathway.

In the second step, the nitro group was reduced using palladium-on-carbon and hydrogen gas. The step proceeds with high yields (>90%) and very few side products. In many cases, purification was not necessary.

In the final step, a sulfonyl-chloride coupling was used to add the SO_2CF_3 group. This step is complicated by the highly volatile and reactive nature of the sulfonyl chloride reagent, which degrades in air. The reaction must be carried out under argon, but the sulfonyl chloride will evaporate if continuously flowing argon is used. Instead, the core molecule was dissolved in DMF and placed under argon in a round bottom flask with a septa. The sulfonyl chloride was then added dropwise through the septa using a syringe over the course of approximately five minutes. The reaction proceeds to completion with a large number of side products and low yields (<30%).

All solvents and reagents were sourced from Sigma-Aldrich. After each reaction step the product was characterized by H-NMR and purified if necessary. Purification was accomplished by recrystallization from ethyl acetate or column chromatography.

Four final compounds were produced using this methodology, as shown in Figure 3.

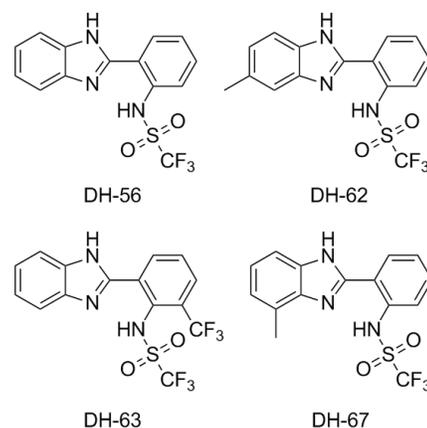


Figure 3. The 4 final compounds produced in this experiment. DH-56 is the “lead” compound, while the remaining three are derivatives.

Final compounds were tested for activity against the IspF enzyme in a surface plasmon resonance (SPR) assay.

Computational Methods

The crystal structure of IspF was taken from the Protein Data Bank, code 3iew.⁷ The first monomer of this crystal structure features the zinc in a trigonal-bipyramidal geometry, while zinc adopts the tetrahedral geometry in the remaining two monomers. Considering the anticipated binding motif of the lead molecule, the first monomer was used for future steps. Further, all analysis was carried out on the apo and holo system of the protein in parallel. The apo system was generated by removing the existing ligand and placing 2 hydroxide anions in appropriate proximity to the zinc. To generate the holo system, the lead molecule was first drawn using the Hyperchem software. Then, it was aligned to the existing ligand using Visual Molecular Dynamics (VMD) with the RMSD calculator extension.⁸ Finally, the .pdb file was edited to replace the existing ligand with the lead molecule, generating the holo system.

Next, .mol2 files were generated for the zinc, the ligating hydroxide anions, and the lead compound using the antechamber software.⁹ In the case of the zinc, the charge in the .mol2 file was manually modified from “0” to “2.” Additionally, parmchk2¹⁰ was used to generate .frcmod files for all ligating molecules. As for the protein, the monomer’s crystal structure was cleaned and protonated with the H++ webserver (version 3.2).¹¹⁻¹³ The default settings of the H++ webserver were used. This process added hydrogens to the protein, but removed the zinc and ligands from the active site, leading to incorrect protonation states for the active site’s 2 histidine residues. The .pdb file was manually edited to correct these protonation states. Then, the zinc and ligating residues were added back in. The resulting file was renumbered using pdb4amber¹⁰ for both the apo and holo systems, yielding the necessary files for the MCPB.py workflow.⁶

The first step of MCPB.py automatically generated the input files to run the active sites of the holo and apo structures through Gaussian 09.¹⁴ This program used a Hartree-Fock calculation and the 631G* basis set to calculate the geometry, bond force constants, and charge distribution within the active site. The formchk command generated an additional .fchk file for use in future steps of MCPB.py. At this point, MCPB.py’s associated literature suggests applying a bug fix. This step is out of date and no longer required. In fact, applying the bug fix leads to additional errors. For this project, the bug fix was intentionally not taken.

Finally, the resulting output files were run through the remaining 3 steps in the MCPB.py workflow. This produced a .prmtop and .incrdp file, the two inputs for a molecular dynamics simulation. Parmed¹⁰ was used to extract the zinc parameters from the .prmtop file.

Results and Discussion

Experimental results found that none of the 4 compounds displayed significant inhibition toward the IspF enzyme. The measurements were recorded in terms of IC₅₀ values. These are the concentration of compound required to reduce the enzyme’s activity by 50%. Less than 10 μm would have been considered a “hit,” and less than 100 μm would have registered in the assay. But, it was found that the IC₅₀ value of each compound must be greater than 100 μm.

Table 1. Experimentally Measured IC₅₀ Values

Compound	IC ₅₀ (μm)
DH-56	> 100
DH-62	> 100
DH-63	> 100
DH-67	> 100

The results of Gaussian 09’s geometry optimization reveal a possible explanation for the compounds’ unexpectedly poor inhibition of the protein. As shown in Figure 4, calculations using quantum mechanics predict that the compounds will not bond the zinc through both nitrogen atoms as previously expected. Instead, the compounds bond only through the aromatic nitrogen.

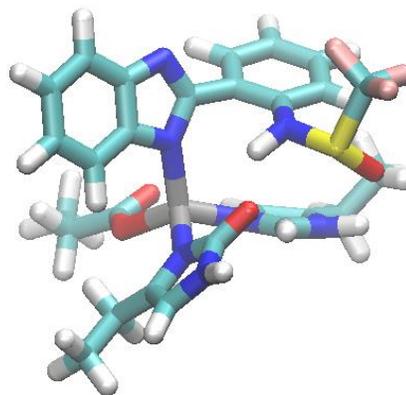


Figure 4. Geometry predicted by quantum mechanical calculations with DH-56 in the active site. The zinc adopts the tetrahedral geometry, only binding the ligand a single time.

However, calculations involving the apo system (wherein the zinc is ligated by hydroxides) found that the zinc maintained its trigonal bipyramidal structure. This is shown in Figure 5.

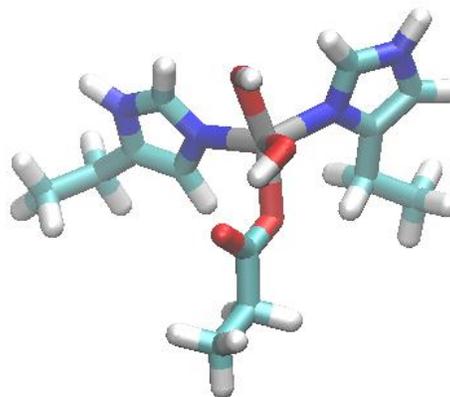


Figure 5. Predicted geometry of the apo active site. The zinc maintains its trigonal bipyramidal geometry.

Apparently, the active site can adopt a trigonal bipyramidal geometry, but will not do so in the case of DH-56. There are two likely explanations for this. The first is steric hinderance. As shown in Figure 6, substantial steric hinderance is required for both nitrogen atoms to bind the zinc.

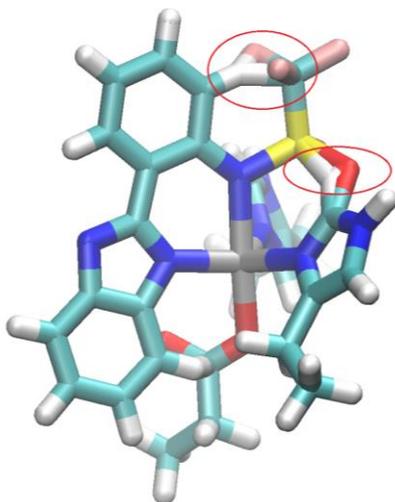


Figure 6. DH-56 binding the zinc through both nitrogen atoms. This is the initial geometry provided for the Gaussian 09 calculations. Note the steric interactions in the upper right corner of the image. The quantum calculations found this geometry to be unstable, preferring that of Figure 3.

The second possible explanation is related to angle strain. As seen in Figure 5, when the zinc adopts a trigonal bipyramidal geometry it leaves an axial and an equatorial position available. The expected angle between these two positions is 90° (see Figure 7).

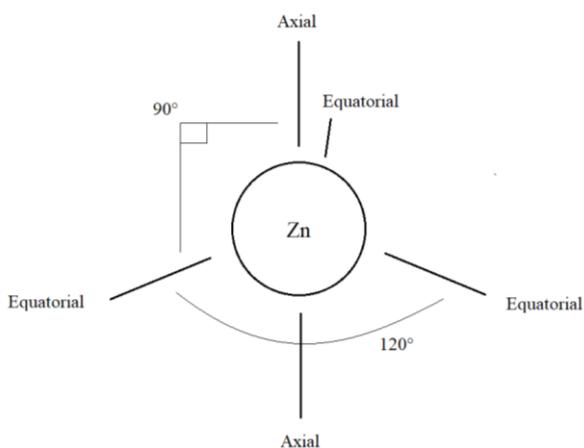


Figure 7. Trigonal bipyramidal geometry of zinc. The expected angle between 2 equatorial groups is 120° while the expected angle between an equatorial and axial is 90° .

However, if both nitrogen atoms bind the zinc, a six-membered ring is formed. The ideal internal angle of a six-membered ring is 120° . In order for both nitrogen atoms to bind, they must overcome the strain of altering the ring's internal angle.

The steric and angle strain factors almost certainly both contribute to the actual cause of the poor inhibition of DH-56 and its derivatives. Future compounds may exhibit improved inhi-

bition if smaller rings (i.e. 5 or 4-member rings) and compounds with reduced bulk are used.

In consideration both issues, several new potential lead molecules are proposed. These are outlined in Figure 8.

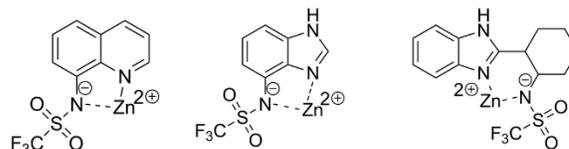


Figure 8. Three new proposed lead candidates in consideration of steric strain and angle strain.

All three of the compounds are reasonable synthetic targets and could offer improved activity against IspF. Future studies in this area should focus on these new lead molecules.

In addition to providing an explanation for the DH-56's poor activity, the computational results also provided the parameters for modeling the apo system's zinc. These parameters allow the active site to be probed using MD, speeding the development of future IspF inhibitors. The parameters are particularly notable because they represent the trigonal bipyramidal geometry, rather than the more typical tetrahedral geometry.

ASSOCIATED CONTENT

Associated content is available online at:

<https://drive.google.com/drive/folders/1YOXgLt7Vj-yv9TadGYno8rCCK7LJ2Kr?usp=sharing>

Parameters for the zinc atom in the holo system in Figure 4.

Parameters for the zinc atom in the apo system in Figure 5.

Optimized geometry for the holo system if the DH-56 ligand is rotated 180° . This geometry is thought to be unrealistic.

AUTHOR INFORMATION

Corresponding Author

danielharper206@pointloma.edu

ACKNOWLEDGMENT

The experimental portion of this research was conducted with the funding and resources of Point Loma Nazarene University. I am also grateful to Dr. Lane Votapka and Dr. Matthieu Rouffet for their advice and mentorship. Finally, I would like to thank the Amaro lab at UCSD for use of their computer systems and Gaussian 09 license.

ABBREVIATIONS

IspF, 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; HTS, High Throughput Screening; SAR, Structure Activity Relationship; MD, Molecular Dynamics; SPR, Surface Plasmon Resonance; VMD, Visual Molecular Dynamics

REFERENCES

1. Martin, D.; Rouffet, M.; Cohen, S. Illuminating Metal-Ion Sensors: Benzimidazolesulfonamide Metal Complexes. *Inorg. Chem.* 2010, 49, 10226-10228.
2. Tantillo, D. Faster, Catalyst! React! React! Exploiting Computational Chemistry for Catalyst Development and Design. *Acc. Chem. Res.* 2016, 49 (6), 1079
3. Cramer, C. *Essentials of Computational Chemistry*; John Wiley & Sons Ltd.: Chichester, 2004.
4. Branco, R.; Fernandes, P.; Ramos, M. Molecular Dynamics Simulations of the Enzyme Cu, Zn Superoxide Dismutase. *J. Phys. Chem.* 2006, 110 (3), 16754-16762
5. Lin, F. Wang, R. Systematic Derivation of AMBER Force Field Parameters Applicable to Zinc-Containing Systems. *J. Chem. Theory Comput.* 2009, 14(1), 1852-1870
6. Li, P.; Merz, K.; MCPB.py: A Python Based Metal Center Parameter Builder. *J. Chem. Inf. Model.* 2016, 56(4), 599-604.
7. Begley, D.; Hartley, R.; Davies, D. Leveraging structure determination with fragment screening for infectious disease drug targets: MECP synthase from *Burkholderia pseudomallei*. *J Struct Funct Genomics.* 2011, 12, 63-76.
8. Hmuphrey, W.; Dalke, A; Schulten, K. VMD-Visual Molecular Dynamics. *J Molec Graphics.* 1996, 14, 33-38.
9. Wang, J.; Wang, W.; Kollman, P.; Case, D. Automatic atom type and bond type perception in molecular mechanical calculations. *J Mol Graph Model.* 2006, 25, 247260.
10. D.A. Case, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, D. Greene, N. Homeyer, S. Izadi, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, L. Xiao, D.M. York and P.A. Kollman (2017), AMBER 2017, University of California, San Francisco.
11. Anandakrishnan, R.; Aguilar, B.; Onufriev, A. H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulation. *Nucleic Acids Res.* 2012, 40(W1), W537-541.
12. Meyers, J.; Grothaus, G.; Narayanan, S.; Onufriev, A. A simple clustering algorithm can be accurate enough for use in calculations of pKs in macromolecules. *Proteins.* 2006, 63, 928-938.
13. Gordon, J.; Myers, J.; Folta, T.; Shoja, V.; Heath, L.; Onufriev, A. H++: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res.* 2005, 33, W368-371.
14. Gaussian 09, Revision A.02, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2016.