BIOCHEMISTRY
including biophysical chemistry & molecular biology

Rapid Report

pubs.acs.org/biochemistry

Predicting Enzyme–Substrate Specificity with QM/MM Methods: A Case Study of the Stereospecificity of D-Glucarate Dehydratase

Boxue Tian,[†,‡] Frank Wallrapp,[‡] Chakrapani Kalyanaraman,[‡] Suwen Zhao,[‡] Leif A. Eriksson,*[,§] and Matthew P. Jacobson*[,‡]

[†]School of Chemistry, National University of Ireland, Galway, Galway, Ireland

[‡]Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, California 94158, United States

[§]Department of Chemistry and Molecular Biology, University of Gothenburg, 412 96 Göteborg, Sweden

Supporting Information

ABSTRACT: The stereospecificity of D-glucarate dehydratase (GlucD) is explored by QM/MM calculations. Both the substrate binding and the chemical steps of GlucD contribute to substrate specificity. Although the identification of transition states remains computationally intensive, we suggest that QM/MM computations on ground states or intermediates can capture aspects of specificity that cannot be obtained using docking or molecular mechanics methods.

Understanding enzyme specificity at the atomic level is important for enzyme function prediction and redesign of enzymes. Substrate binding and the chemical steps are critical for specificity if product release is not rate-limiting, and k_{cat}/K_M is often used to quantify the enzymatic proficiency for a given substrate. Nowadays, computer simulations of enzyme-ligand systems have become routine.[1−8] To evaluate k_{cat}/K_M with computational approaches, one needs to model both substrate binding and chemical steps. However, most computational studies have only focused on one of these aspects. In most cases, the enzyme (in vitro) function prediction problem is equivalent to studying enzyme substrate specificity, where the key is to find a ligand with high k_{cat}/K_M values. Therefore, combining properties for the substrate-binding step (e.g., binding affinity) and for the chemical steps (e.g., activation energy) should give more reasonable results in determining whether or not a ligand is a substrate.

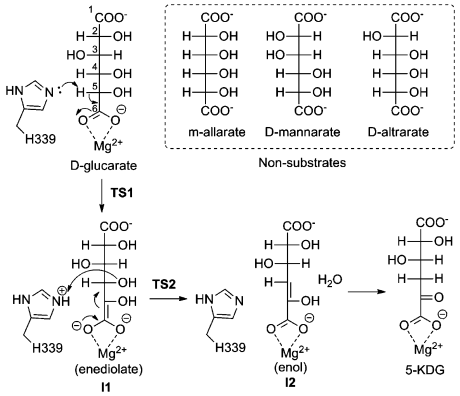
Quantum mechanical/molecular mechanical (QM/MM) methods have been widely used to investigate enzyme mechanisms, and computationally proposed mechanisms can be compared with experimental data such as kinetic constants and kinetic isotope effects.[4] Here we make use of QM/MM methods to study instead the specificity of an enzyme for closely related substrates. The enzyme in question is D-glucarate dehydratase (GlucD), a well-studied member of the mechanistically diverse enolase superfamily.[9−11] We have previously studied the substrate specificity of GlucD and other acid sugar dehydratases of the enolase superfamily (among many other enzymes) using molecular docking methods.[12,13] The concept is that docking methods more commonly used in computer-aided drug design can be productively used to identify potential substrates (and eliminate implausible ones) by crudely estimating relative binding affinities. Many retrospective[12−16] and prospective studies[17] have demonstrated the utility of this approach, primarily to reduce the number of substrates for experimental testing and to identify enzymes likely to have novel substrates. Nonetheless, it is clear that these methods have many limitations, including well-documented challenges in estimating relative binding affinities. Arguably the most fundamental limitation is that molecular mechanics methods, no matter how sophisticated the energy models, cannot model the reactive step, or more precisely the transition states that may limit the overall reaction rate, although a useful heuristic has been successfully employed in which "high-energy intermediates" are docked rather than substrates or products.[18]

Empirically, we have found that metabolite docking against GlucD and other acid sugar dehydratases of the enolase superfamily routinely ranks the "correct" substrate (as judged by in vitro k_{cat}/K_M) among the top metabolites from large libraries.[12−17] However, as in computer-aided drug design, there are many "false positives". Some of these can be eliminated as implausible based on other criteria such as lacking a required chemical group (proton α to a carboxylate in the case of the enolase superfamily) or when the metabolite docks in a pose inconsistent with catalysis. However, some false positives are plausible and closely related to the true substrate,; for example, we invariably find many acid sugars among the top-ranked metabolites for acid sugar dehydratases. While predicting even this relatively crude level of selectivity is clearly useful (e.g., likely 6 carbon diacid sugar dehydratase), more precise predictions of details such as stereochemistry would be of course preferable.

GlucD catalyzes the dehydration of D-glucarate or L-idarate to give 5-keto-4-deoxy-D-glucarate (5-KDG), as well as the interconversion between D-glucarate and L-idarate.[9−11,19−21] The overall reaction mechanism is unambiguous,[10,11] proceeding through an enolate intermediate, as is believed to be the case for all known members of the superfamily. In the dehydration of D-glucarate, an active site histidine (H339) abstracts a proton from C5 to give the enediolate intermediate **I1** (Scheme 1). The protonated H339 then transfers the proton to the hydroxyl on C4 to give the enol intermediate **I2**, which converts to the final product 5-KDG (Scheme 1).

Received: April 30, 2013
Revised: July 16, 2013
Published: July 31, 2013

Scheme 1. Catalytic Mechanism of GlucD and the Structures of the Three Nonsubstrates



GlucD, like other acid sugar dehydratases in the superfamily, has high stereospecificity. Of the 12 stereoisomers of glucarate, dehydration activity has only been observed for glucarate and idarate (epimers at C5),[20] although other stereoisomers are known to be substrates for other acid sugar dehydratases in the enolase superfamily.[9] Our focus here is stereospecificity at the noncatalytic positions C2 and C3 (*m*-allarate, D-mannarate, and D-altrarate; Scheme 1).[20]

Table 1 shows that docking-based methods do not unambiguously identify glucarate as the "correct" substrate.

Table 1. Relative Binding Scores of the Putative Reactive Poses of D-Glucarate and Nonsubstrates[a]

ligands	Glide-SP	Glide-XP	MM/GBSA[b]	QM/MM[c]
D-glucarate	0.0	0.0	0.0	0.0
m-allarate	0.7	−1.9	−3.7	6.6
D-mannarate	0.1	−2.4	3.5	9.9
D-altrarate	0.2	−1.7	10.7	13.9

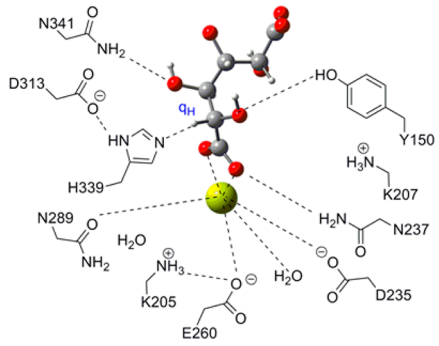
[a]Energies are in kcal/mol. [b]Binding scores of the Glide-XP poses are calculated at the MM/GBSA level using Prime, allowing relaxation of the receptor−ligand complexes (see SI for details). [c]Relative QM/MM energies of the receptor−ligand complexes.

Computational details can be found in Supporting Information (SI). Protonation states of H339 and K207 were assigned according to the Michaelis complex for D-glucarate, as the reactive regions of the four ligands in Scheme 1 are identical. The putative reactive poses of the four ligands were obtained by superimposing these ligands to the product ligand in the crystal structure (PDB ID: 1ECQ), with refinements using Glide.

Relative binding scores for the four ligands are shown in Table 1, using three different scoring functions: the SP and XP scoring functions from Glide, and the molecular mechanics/implicit solvent based MM/GBSA method, which we have used in much of our previous work on predicting enzyme substrates (see SI for details). D-Glucarate is only slightly more favored than the other three ligands in the Glide-SP calculations and is not predicted to be the best binder in the Glide-XP and MM/GBSA calculations (Table 1). Even if the different scoring functions did generate identical rankings (and even if these were correct), these results alone cannot distinguish between substrates and inhibitors.

To investigate the relative energies of the transition states for the chemical steps, the QM/MM [ONIOM (B3LYP:AMBER)] method was used, with the QM layer as described in Scheme 2. All the QM/MM energies reported here are relative energies at

Scheme 2. QM Region in the QM/MM Calculations



the [B3LYP/6-311+G(2d,p):AMBER] level. Although there is experimental evidence to support the catalytic mechanism of GlucD (as described in Scheme 1),[10,11] no computational study on this mechanism has yet been reported. QM/MM calculations have been reported for two related enzymes in the same superfamily as GlucD, mandelate racemase (MR),[22−24] and enolase (the namesake of the superfamily).[25,26] GlucD shares more similarities with enolase in terms of the substrate-binding mode, in which the carboxylate group of the substrate is a bidentate ligand to the Mg^{2+} (Scheme 2).

The QM/MM calculations on the D-glucarate substrate predict that **TS2** is the rate-determining state (Figure 1), and

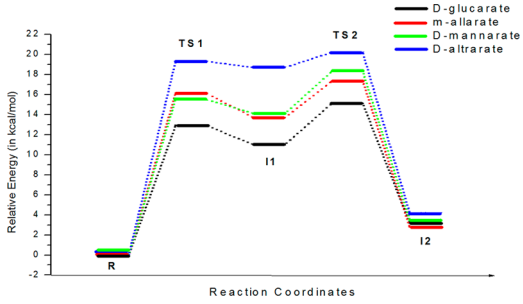


Figure 1. Energy profiles (in kcal/mol) along the reaction coordinate for D-glucarate (in black), *m*-allarate (in red), D-mannarate (in green), and D-altrarate (in blue).

the obtained energy barrier of 15.1 kcal/mol is consistent with the phenomenological activation energy ∼15 kcal/mol (at 295.15 K, k_{cat} = 35 s^{-1} for D-glucarate[10]). The barrier height for the proton abstraction step is also significant (**TS1**, 12.9 kcal/mol; Figure 1). Although no experimental results are available to verify the rate-limiting step, it has been suggested that the proton abstraction step is at least partially rate-limiting,[10,11] because all the enzymes in the EN superfamily share this common step.

We thus examine both transition states (**TS1** and **TS2**) in order to study the specificity at the chemical steps. D-Glucarate is the best substrate in terms of energies for both **TS1** and **TS2** (Figure 1 and Table 2). The energy difference of the transition states between D-glucarate and the three nonsubstrates is in the range of 2−7 kcal/mol (Table 2), which is large enough to conclude that specificity is at least partially encoded in the chemical steps. Comparison of the energy profiles and the optimized geometries of the four ligands can be found in Figures S2−S3. Because of the limitations of the QM/MM methodology, including finite QM region and basis sets, and neglect of entropic effects, we interpret the results in Table 2 qualitatively rather than quantitatively.

dx.doi.org/10.1021/bi400546j | *Biochemistry* 2013, 52, 5511−5513

Table 2. Relative Energies of TS1 and TS2 for D-Glucarate and Three Nonsubstrates

ligands	ΔE_{TS1} [a]	$\Delta\Delta E_{TS1}$ [b]	ΔE_{TS2} [a]	$\Delta\Delta E_{TS2}$ [b]	q_H [c]
D-glucarate	12.9	0.0	15.1	0.0	+0.78
m-allarate	16.1	3.2	17.3	2.2	+0.33
D-mannarate	15.5	2.6	18.3	3.2	+0.61
D-altrarate	19.3	6.4	20.2	5.1	+0.02

[a]B3LYP/6-311+G(2d,p):AMBER energies in kcal/mol. [b]$\Delta\Delta E_{TS} = \Delta E_{TS}$[ligand] − ΔE_{TS}[D-glucarate] in kcal/mol. [c]ESP charges of Michaelis complexes at B3LYP/6-311+G(2d,p): AMBER level, details in Figure S4.

Because identifying transition states using QM/MM methods remains computationally expensive, we also investigated whether the quantum results on the Michaelis complexes provided information about specificity. First, the relative energies of the substrate complexes using QM/MM did rank D-glucarate as optimal (Table 1). Unlike the other scoring functions shown in Table 1, quantum methods allow electron distributions to change in response to the atomic environment. To quantify this effect, we computed atomic charges derived from electrostatic potentials (ESP), Figure S4. Significant differences among the ligands are observed for atoms involved in the chemical steps, with the substrate GlucD showing a larger positive charge on the proton abstracted in the first chemical step (Table 2). The trends in the partial charges are in qualitatively accordance with the energies of **TS1** (Table 2). As the optimized geometries at the reactive region are very similar for the four ligands, we suggest that the energies of **TS1** are mainly affected by the variations of the overall electrostatic environment. As optimizations of the Michaelis complexes are computationally more efficient than optimizations of the transition state, the ESP atomic charges of the Michaelis complex might be useful for excluding poor substrate candidates when the computational resources are limited. However, we cannot yet assess the generality of this heuristic beyond this specific case.

In conclusion, docking and QM/MM calculations are performed to study the GlucD stereospecificity at the non-reactive regions. Our results show that GlucD specificity is likely encoded in both the substrate binding and the chemical steps. We emphasize that enzyme substrate specificity is a two parameter (k_{cat}/K_M) problem, and computational studies should investigate properties for both steps.

■ ASSOCIATED CONTENT

Supporting Information

Computational details; images for intermediates and TSs; ESP charges; docking results; Cartesian coordinates and energies. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: matt.jacobson@ucsf.edu; phone: 415-514-9811.

Funding

This work was supported by the Enzyme Function Initiative, NIH grant U54-GM093342 (to M.P.J.) and the National University of Ireland-Galway (to L.A.E. and B.T.).

Notes

The authors declare the following competing financial interest(s): M.J.P. is a consultant for Schrodinger LLC.

■ ACKNOWLEDGMENTS

The SFI/HEA Irish Centre for High-End Computing (ICHEC) is acknowledged for the provision of computational facilities and support. We thank John Gerlt (U. Illinois) for helpful discussions.

■ REFERENCES

(1) Garcia-Viloca, M., Gao, J., Karplus, M., and Truhlar, D. G. (2004) *Science 303*, 186−195.
(2) Marti, S., Roca, M., Andres, J., Moliner, V., Silla, E., Tunon, I., and Bertran (2004) *J. Chem. Soc. Rev. 33*, 98−107.
(3) Warshel, A. (2003) *Annu. Rev. Biophys. Biomol. Struct. 32*, 425−443.
(4) Gao, J., Ma, S., Major, D. T., Nam, K., Pu, J., and Truhlar, D. G. (2006) *Chem. Rev. 106*, 3188−3209.
(5) Kua, J., Zhang, Y. K., and McCammon, J. A. (2002) *J. Am. Chem. Soc. 124*, 8260−8267.
(6) Jorgensen, W. L. (2004) *Science 303*, 1813−1818.
(7) Mulholland, A. J. (2005) *Drug Discovery Today 10*, 1393−1402.
(8) Ramos, M. J., and Fernandes, P. A. (2008) *Acc. Chem. Res. 41*, 689−698.
(9) Gerlt, J. A., Babbitt, P. C., and Rayment, I. (2005) *Arch. Biochem. Biophys. 433*, 59−70.
(10) Gulick, A. M., Hubbard, B. K., Gerlt, J. A., and Rayment, I. (2000) *Biochemistry 39*, 4590−4602.
(11) Gulick, A. M., Hubbard, B. K., Gerlt, J. A., and Rayment, I. (2001) *Biochemistry 40*, 10054−10062.
(12) Rakus, J. F., Kalyanaraman, C., Federov, A. A., Federov, E. V., Mills-Groninger, F. P., Toro, R., Bonanno, J., Bain, K., Sauder, M., Burley, S. K., Almo, S. C., Jacobson, M. P., and Gerlt, J. A. (2009) *Biochemistry 48*, 11546−11558.
(13) Kalyanaraman, C., and Jacobson, M. P. (2010) *Biochemistry 49*, 4003−4005.
(14) Song, L., Kalyanaraman, C., Fedorov, A. A., Fedorov, E. V., Glasner, M. E., Brown, S., Imker, H. J., Babbitt, P. C., Almo, S. C., Jacobson, M. P., and Gerlt, J. A. (2007) *Nat. Chem. Biol. 3*, 486−491.
(15) Kalyanaraman, C., Imker, H. J., Fedorov, A. A., Fedorov, E. V., Glasner, M. E., Babbitt, P. C., Almo, S. C., Gerlt, J. A., and Jacobson, M. P. (2008) *Structure 16*, 1668−1677.
(16) Hermann, J. C., Marti-Arbona, R., Fedorov, A. A., Fedorov, E., Almo, S. C., Shoichet, B. K., and Raushel, F. M. (2007) *Nature 448*, 775−781.
(17) Lukk, T., Sakai, A., Kalyanaraman, C., Brown, S. D., Imker, H. J., Song, L., Fedorov, A. A., Fedorov, E. V., Toro, R., Hillerich, B., Seidel, R., Patskovsky, Y., Vetting, M. W., Nair, S. K., Babbitt, P. C., Almo, S. C., Gerlt, J. A., and Jacobson, M. P. (2012) *Proc. Natl. Acad. Sci. U. S. A. 109*, 4122−4127.
(18) Fan, H., Hitchcock, D. S., Seidel-II, R. D., Hillerich, B., Lin, H., Almo, S. C., Sali, A., Shoichet, B. K., and Raushel, F. M. (2013) *J. Am. Chem. Soc. 135*, 795−803.
(19) Gulick, A. M., Palmer, D. R. J., Babbitt, P. C., Gerlt, J. A., and Rayment, I. (1998) *Biochemistry 37*, 14358−14368.
(20) Palmer, D. R. J., Wieczorek, S. J., Hubbard, B. K., Mrachko, G. T., and Gerlt, J. A. (1997) *J. Am. Chem. Soc. 119*, 9580−9581.
(21) Wieczorek, S. J., Kalivoda, K. A., Clifton, J. G., Ringe, D., Petsko, G. A., and Gerlt, J. A. (1999) *J. Am. Chem. Soc. 121*, 4540−4541.
(22) Garcia-Viloca, M., Gonzalez-Lafont, A., and Lluch, J. M. (2001) *J. Am. Chem. Soc. 123*, 709−721.
(23) Prat-Resina, X., Garcia-Viloca, M., Gonzalez-Lafont, A., and Lluch, J. M. (2002) *Phys. Chem. Chem. Phys. 4*, 5365−5371.
(24) Prat-Resina, X., Gonzalez-Lafont, A., and Lluch, J. M. (2005) *J. Phys. Chem. B 109*, 21089−21101.
(25) Alhambra, C., Gao, J. L., Corchado, J. C., Villa, J., and Truhlar, D. G. (1999) *J. Am. Chem. Soc. 121*, 2253−2258.
(26) Liu, H. Y., Zhang, Y. K., and Yang, W. T. (2000) *J. Am. Chem. Soc. 122*, 6560−6570.