

## Status report on the quantum chemical cluster approach for modeling enzyme reactions

Fahmi Himo <sup>1</sup>✉ & Sam P. de Visser <sup>2</sup>✉

The cluster approach is a very valuable technique for elucidating reaction mechanisms of enzymes. Here, the authors discuss the current status of this methodology, highlighting its strengths and weaknesses, and argue that it should be the method of choice for investigating enzymatic reaction mechanisms.

### Preface

The cluster approach is a very valuable technique for elucidating the reaction mechanisms of enzymes. Here, the authors discuss the current status of this methodology, highlighting its strengths and weaknesses, and argue that it should be the method of choice for investigating enzymatic reaction mechanisms.

Enzymes are Nature's sophisticated catalytic machines with functions related to biosynthesis and biodegradation of natural products. Their size and complexity are daunting for computational chemists. Nevertheless, over the years computational chemists have developed an array of approaches that are able to model various aspects of enzymes' functions and properties, at different length and time scales. These methods are so powerful these days that the computations can be used as a predictive technique that can guide experimental work.

### The cluster approach

The quantum chemical cluster approach is a technique devised to investigate reaction mechanisms and spectroscopic properties of enzyme active sites<sup>1–3</sup>. It relies on the use of relatively accurate quantum chemical methods that can predict bond strengths, physicochemical properties, and spectroscopic features of the active site of the enzyme. Most common approaches use density functional theory (DFT) methods and share many common features with how homogeneous catalysis and biomimetic models are calculated.

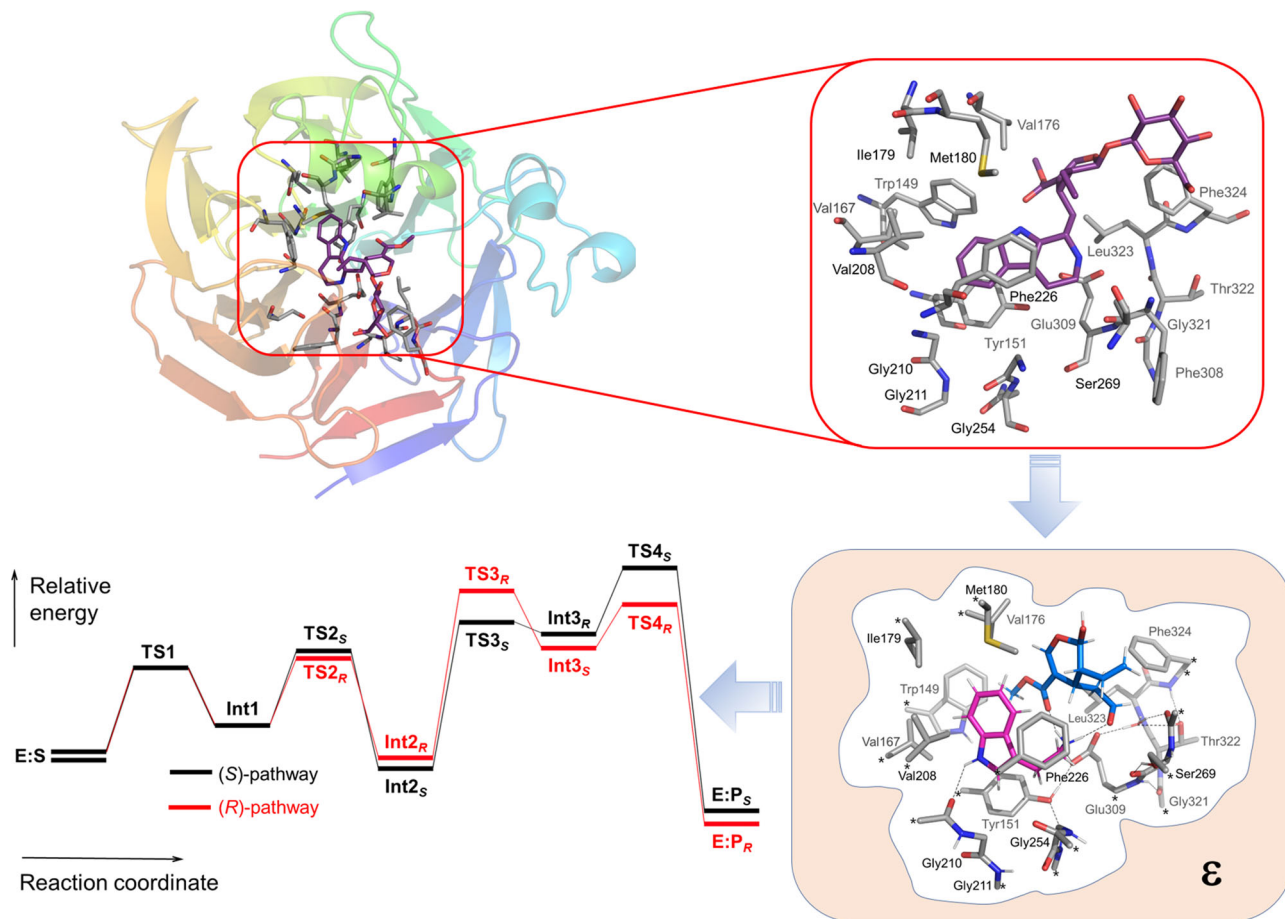
In brief, a limited model of the active site is carved out of the actual structure and an energy profile for a possible reaction mechanism is calculated, including all intermediates (local minima) and transition states leading to products and by-products through bifurcation pathways (Fig. 1). The feasibility of a certain mechanism can thereby be judged. Furthermore, the bifurcation pathways and transition states give structural and electronic explanations of what drives the selectivity of the reaction and how it could be manipulated. The rest of the enzyme is typically modeled assuming a homogeneous polarizable surrounding, and to keep the active site architecture reasonably intact, a number of centers at the periphery of the model are usually kept fixed to their crystallographic positions.

### Size of active-site models

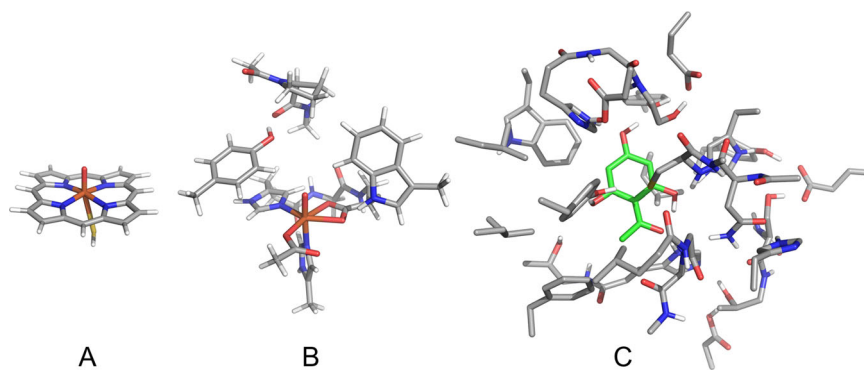
Some 20 years ago, active site models were typically 30–50 atoms in size and did usually not contain many atoms beyond the first-coordination sphere of the metal center<sup>4–6</sup>. With the

<sup>1</sup>Department of Organic Chemistry, Arrhenius Laboratory, Stockholm University, SE-10691 Stockholm, Sweden. <sup>2</sup>Manchester Institute of Biotechnology and Department of Chemical Engineering and Analytical Science, The University of Manchester, 131 Princess Street, Manchester M1 7DN, UK.

✉email: [fahmi.himo@su.se](mailto:fahmi.himo@su.se); [sam.devisser@manchester.ac.uk](mailto:sam.devisser@manchester.ac.uk)



**Fig. 1 Schematic drawing of the steps involved in the cluster approach.** A model of the active site is designed on the basis of a crystal structure. A number of positions at the edge of the model are kept fixed during the geometry optimizations, and the surrounding is modeled as a continuum solvent. Intermediates and transition states are located and their energies are used to judge reaction mechanisms and elucidate the origins of selectivity.



**Fig. 2 Examples of active site model sizes at different times.** **A** 40 atoms: Cytochrome P450 in 2000<sup>7</sup>. **B** 116 atoms: Prolyl-4-hydroxylase in 2010<sup>8</sup>. **C** 413 atoms: Acyltransferase in 2020<sup>9</sup>.

improvement of computer power, it soon became possible to treat larger models that include some part of the second-coordination sphere, which determines the substrate and cofactor positioning. These models were initially 100–150 atoms in size, while today's cluster models typically have >300 atoms (Fig. 2)<sup>7–9</sup>. The larger models make of course a more accurate representation of the real system and typically include all first- and second-coordination sphere effect perturbations, including those of charged groups and key hydrogen bonding and  $\pi$ -stacking interactions<sup>10</sup>.

### Some applications

The cluster approach has been instrumental in elucidating the catalytic cycles of a large number of enzymes and has been particularly helpful in gaining an understanding of the short-lived intermediates during the reaction<sup>1–3</sup>. As an early example of this, we can mention the work on cytochrome P450 enzymes, where experimental studies had proposed a second-oxidant active in the catalytic cycle, namely Compound 0 or the iron(III)-hydroperoxo species, rather than Compound I or the iron(IV)-oxo heme cation

radical species<sup>11</sup>. However, computational modeling established that Compound 0 cannot react with olefins or aromatic rings at room temperature and that Compound I was the sole oxidant in the P450s<sup>12</sup>.

Over the years, the cluster approach has been used to resolve experimental controversies and inconsistencies and has also guided further experimental studies. It has solved some of the most outstanding and complex problems in enzymology, such as, for example, the mechanism of dioxygen synthesis by Photosystem II and the catalytic reaction mechanisms of various metalloenzymes<sup>13</sup>. Overall, these cluster studies, provided they have been set up properly, yield relatively accurate activation energies that are consistent with experimental rate constants. The calculations are also able to reproduce product distributions well and give reliable geometries and electronic and spectroscopic properties that can be linked to experiments.

A consequence of the larger model size accessible today is that the application space now is wider. It has, for example, been demonstrated that it is possible to quite accurately reproduce enantioselectivity in enzymes and determine the factors controlling it<sup>14</sup>. The cluster approach has also been used as a predictive tool for the engineering of proteins to give them novel functions and selectivities. In an early example, a cluster model study on the nonheme iron enzyme *S*-mandalate synthase predicted how to engineer the substrate-binding pocket in order to obtain *R*-mandalate products instead, which was confirmed by experimental mutations<sup>15</sup>. Very recently, cluster calculations on the mechanism and origins of enantioselectivity of acyltransferase from *Mycobacterium smegmatis* assisted in the design of variants with very high selectivities for unnatural substrates<sup>16</sup>.

## Discussion

In general, the cluster method is highly robust and works particularly well for enzymes with small substrates. However, the cluster model is often dependent on the accuracy of the starting enzyme crystal structure, which needs to be correctly folded and should ideally have possible cofactors and substrates or substrate-analogs bound in the same conformation as its active form. If this structure is not in the active state or major folding changes happen upon substrate binding or during the catalytic cycle of the enzyme, it cannot be modeled. Similarly, large conformational movements and allosteric effects are challenging to incorporate in cluster models. One can in these cases usually resort to molecular dynamics simulations to obtain a better starting structure.

While the cluster approach produces relatively accurate reaction energies and barriers, it has difficulties reproducing absolute  $pK_a$  values and redox potentials, due to long-range effects related to the change of the overall charge of the cluster model. In these cases, one can use some experimental value to calibrate the cluster energies and obtain mechanistic insights<sup>13</sup>.

Despite these challenges, we believe that the cluster model approach is the method of the future, with ample opportunities for scientific research. The cluster models are usually easy to set up due to the use of 100 s of atoms as compared to 1000 s in the quantum mechanics/molecular mechanics (QM/MM) models. They suffer less from convergence problems during the calculations.

It is a robust scheme that minimizes the user-based errors that are often part of set-up procedures. The results are reproducible with different methods, software packages, starting conditions and give small deviations between the results. Importantly, the cluster approach is computationally less expensive than, e.g., QM/MM and related methods. This allows the user to examine many alternative reaction mechanisms in a relatively short time.

## Outlook

In what directions is the field likely to develop in the future? In our opinion, even if the computational resources will allow, it is not obvious that the models will grow much larger than today's size. Often the calculations converge with size, and expanding the model further does not necessarily give more insight. Large models will lead to problems with local minima, which would require sampling and thus many more calculations. The cluster approach will also be combined with other techniques, such as empirical valence bond and free energy perturbation techniques to ask new kinds of questions. Work in these directions has already started to appear<sup>17,18</sup>.

As can be understood from the discussion above, we are big proponents of the cluster approach. In our opinion, it should be the method of choice for studying enzymatic reaction mechanisms. It has many advantages, and we hope that this Comment will inspire more groups to use it.

Received: 21 December 2021; Accepted: 11 February 2022;

Published online: 08 March 2022

## References

1. Blomberg, M. R. A., Borowski, T., Himo, F., Liao, R.-Z. & Siegbahn, P. E. M. Quantum chemical studies of mechanisms for metalloenzymes. *Chem. Rev.* **114**, 3601–3658 (2014).
2. Shaik, S., Kumar, D., de Visser, S. P., Altun, A. & Thiel, W. Theoretical perspective on the structure and mechanism of cytochrome P450 enzymes. *Chem. Rev.* **105**, 2279–2328 (2005).
3. Himo, F. Recent trends in quantum chemical modeling of enzymatic reactions. *J. Am. Chem. Soc.* **139**, 6780–6786 (2017).
4. Basch, H., Mogi, K., Musaev, D. G. & Morokuma, K. Mechanism of the methane  $\rightarrow$  methanol conversion reaction catalyzed by methane monooxygenase: a density functional study. *J. Am. Chem. Soc.* **121**, 7249–7256 (1999).
5. Wirstam, M., Blomberg, M. R. A. & Siegbahn, P. E. M. Reaction mechanism of compound I formation in heme peroxidases: a density functional theory study. *J. Am. Chem. Soc.* **121**, 10178–10185 (1999).
6. Filatov, M., Harris, N. & Shaik, S. On the “rebound” mechanism of alkane hydroxylation by cytochrome P450: electronic structure of the intermediate and the electron transfer character in the rebound step. *Angew. Chem. Int. Ed.* **38**, 3510–3512 (1999).
7. Ogliaro, F. et al. A model “rebound” mechanism of hydroxylation by cytochrome P450: stepwise and effectively concerted pathways, and their reactivity patterns. *J. Am. Chem. Soc.* **122**, 8977–8989 (2000).
8. Karamzadeh, B., Kumar, D., Sastry, G. N. & de Visser, S. P. Steric factors override thermodynamic driving force in regioselectivity of proline hydroxylation by prolyl-4-hydroxylase enzymes. *J. Phys. Chem. A* **114**, 13234–13243 (2010).
9. Sheng, X., Kazemi, M., Źądło-Dobrowolska, A., Kroutil, W. ; & Himo, F. Mechanism of biocatalytic Friedel-Crafts acylation by acyltransferase from *Pseudomonas protegens*. *ACS Catal.* **10**, 570–577 (2020).
10. de Visser, S. P. Second-coordination sphere effects on selectivity and specificity of heme and nonheme iron enzymes. *Chem. Eur. J.* **26**, 5308–5327 (2020).
11. Vaz, A. D. N., McGinnity, D. F. & Coon, M. J. Epoxidation of olefins by cytochrome P450: Evidence from site-specific mutagenesis for hydroperoxo-iron as an electrophilic oxidant. *Proc. Natl Acad. Sci. USA* **95**, 3555–3560 (1998).
12. Ogliaro, F., de Visser, S. P., Cohen, S., Sharma, P. K. & Shaik, S. Searching for the second oxidant in the catalytic cycle of cytochrome P450: a theoretical investigation of the iron(III)-hydroperoxo species and its epoxidation pathways. *J. Am. Chem. Soc.* **124**, 2806–2817 (2002).
13. Siegbahn, P. E. M. A quantum chemical approach for the mechanisms of redox-active metalloenzymes. *RSC Adv.* **11**, 3495–3508 (2021).
14. Sheng, X., Kazemi, M., Planas, F. & Himo, F. Modeling enzymatic enantioselectivity using quantum chemical methodology. *ACS Catal.* **10**, 6430–6449 (2020).
15. Pratter, S. M. et al. Inversion of enantio-selectivity of a mononuclear non-heme iron(II)-dependent hydroxylase by tuning the interplay of metal-center geometry and protein structure. *Angew. Chem. Int. Ed.* **52**, 9677–9681 (2013).
16. Jost, E. et al. Variants of the acyltransferase from *Mycobacterium smegmatis* enable enantioselective acyl transfer in water. *ACS Catal.* **10**, 10500–10507 (2020).
17. Kazemi, M., Himo, F. & Åqvist, J. Enzyme catalysis by entropy without Circe effect. *Proc. Natl Acad. Sci. USA* **113**, 2406–2411 (2016).

18. Kazemi, M., Sheng, X., Kroutil, W. & Himo, F. Computational study of *Mycobacterium smegmatis* acyl transferase reaction mechanism and specificity. *ACS Catal.* **8**, 10698–10706 (2018).

### Acknowledgements

We thank Dr. Xiang Sheng for his help with preparing the Figures.

### Author contributions

F.H. and Sd.V. wrote the paper.

### Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to Fahmi Himo or Sam P. de Visser.

**Peer review information** *Communications Chemistry* thanks the anonymous reviewers for their contribution to the peer review of this work.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022