Chemical Science

EDGE ARTICLE

View Article Online
View Journal | View Issue

Check for updates

Cite this: *Chem. Sci.*, 2021, **12**, 12682

All publication charges for this article have been paid for by the Royal Society of Chemistry

Received 25th May 2021
Accepted 21st August 2021

DOI: 10.1039/d1sc02826j

rsc.li/chemical-science

Bifurcating reactions: distribution of products from energy distribution in a shared reactive mode†

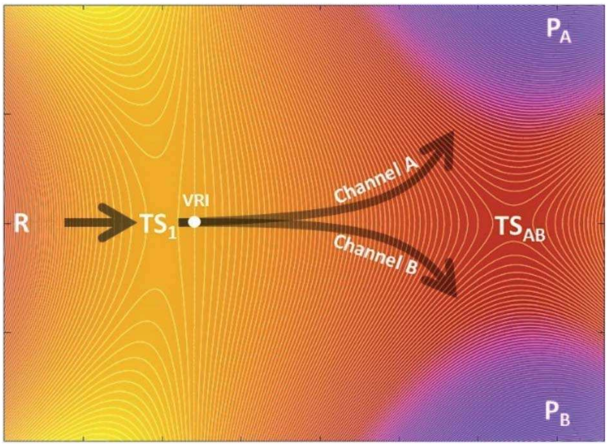
Priyam Bharadwaz,‡ Mauricio Maldonado-Domínguez ‡ and Martin Srnec *

Bifurcating reactions yield two different products emerging from one single transition state and are therefore archetypal examples of reactions that cannot be described within the framework of the traditional Eyring's transition state theory (TST). With the growing number and importance of these reactions in organic and biosynthetic chemistry, there is also an increasing demand for a theoretical tool that would allow for the accurate quantification of reaction outcome at low cost. Here, we introduce such an approach that fulfils these criteria, by evaluating bifurcation selectivity through the energy distribution within the reactive mode of the key transition state. The presented method yields an excellent agreement with experimentally reported product ratios and predicts the correct selectivity for 89% of nearly 50 various cases, covering pericyclic reactions, rearrangements, fragmentations and metal-catalyzed processes as well as a series of trifurcating reactions. With 71% of product ratios determined within the error of less than 20%, we also found that the methodology outperforms three other tested protocols introduced recently in the literature. Given its predictive power, the procedure makes reaction design feasible even in the presence of complex non-TST chemical steps.

Introduction

The kinetic ratio of competing reactions starting from the same reactants is governed by the difference in free energies of their associated transition states, as described by statistical transition-state theory (TST) so that the exclusive or dominant product arises from the lowest free-energy barrier.[1,2] However, there is a steadily growing number of reactions which exhibit a non-statistical (non-TST) behavior where the traditional TST breaks down.[3–5] Prototypical examples are bifurcating organic reactions, the channels of which diverge after passing a common (so-called ambimodal) transition state (TS_1) and lead to two different products as illustrated in the schematic potential energy surface (PES) depicted in Scheme 1.[6–11] Once the ambimodal TS is surmounted, this branching PES is characterized by two different product channels accessed without additional TSs. In such a scenario, selectivity is solely controlled by the atomic positions and momenta of the reactive system once the TS_1 configuration has been surpassed. The overlap of these structural and dynamic signatures with energetically downhill reactive channels determines the final branching ratio.

Cycloadditions are the archetypal and most widely explored reactions within this field.[2] Besides them, the palette of organic reactions known to possess branching PES features includes nucleophile substitution *vs.* addition in α-haloketones,[12] Beckmann and Schmidt rearrangements *vs.* fragmentation, and isomeric Pummerer rearrangements.[13] Metal-catalyzed



Scheme 1 Schematic plot of the potential energy surface (PES) characteristic for a bifurcating reaction. The PES includes three minima – one for the reactant (R) and two for products P_A and P_B, and one key (rate-determining) transition state TS_1 that is shared by two competing R → P_A and R → P_B channels. The PES topography also features the transition state that directly connects P_A with P_B (TS_{AB}) and a valley-ridge inflection (VRI).

J. Heyrovský Institute of Physical Chemistry, The Czech Academy of Sciences, Dolejškova 3, Prague 8, 18223, Czech Republic. E-mail: martin.srnec@jh-inst.cas.cz

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d1sc02826j

‡ These authors contributed equally.

reactions and biosynthetic routes are now recognized examples of bifurcating reactions, highlighting the ever-growing importance of non-equilibrium reactivity.[10,14–16] Importantly, this field has fruitfully evolved by the synergy of experiment and theory. However, prediction and quantification of the product outcome from computational models is still far from being routine. Successful models with proven simplicity and predictive power may thus find immediate application, making reaction design feasible even in the presence of complex non-TST chemical steps.

From the computational perspective, various methods have been used to determine or predict product ratios of such reactions. The most common approach employs *ab initio* molecular dynamics (MD) to evolve reaction systems starting from the rate-determining TS_1.[17,18] The atomic velocities along with their directions at TS_1 are set up randomly and produce trajectories that end up in one of the possible products. This approach requires a collection of a statistically significant number of trajectories, the ratio of which defines the product branching ratio. While *ab initio* MD can provide accurate predictions, its major drawback lies in its considerable computational/time cost. In contrast to this method, Carpenter *et al.* developed a much cheaper computational approach (denoted as a 'dynamic match')[19] that utilizes a projection of the TS_1 reactive mode (*i.e.* the eigenvector with an imaginary frequency) on two bifurcated post-TS_1 reaction coordinates, for which one needs three stationary points (TS_1, P_A and P_B) from the PES presented in Scheme 1. The method was demonstrated to be effective in qualitative/semi-quantitative predictions but the number of tested systems remains rather limited. An alternative approach put forward by Houk correlates bond order differences at TS_1 with the product ratio,[20] fitting a linear function to data from a set of **15** reactions. Although purely empirical, this model has been tested on several examples with positive results,[21–23] demonstrating that TS_1 contains information which, if decoded, allows to make predictions on bifurcating energy surfaces. Recently, Goodman *et al.* designed an approach for the quantitative prediction of product outcomes of bifurcating organic reactions.[24,25] Their ValleyRidge.py algorithm takes advantage of the topography of PES with a post-TS_1 bifurcation and returns the product ratio by combining three key atomic displacement gradients (in the $TS_1 \rightarrow TS_{AB}$, $TS_{AB} \rightarrow P_A$ and $TS_{AB} \rightarrow P_B$ directions; see Scheme 1) with a simplified model of the TS_1 valley. Thus, the method requires the structure of four key points of the PES – TS_1, TS_{AB}, P_A and P_B. Despite its simplicity, it is reported to be remarkably successful in the prediction of product branching ratios, as shown for a set of ~50 reactions. However, the method is originally tailored for reactions with a post-TS_1 bifurcation fulfilling all four above-mentioned points on the respective PES and its application requires the compulsory input of these structures. We note in passing that the protocols are referred in this work as *n*-point methods to express the number of points from the PES involved in the quantification of product distributions. The dynamic match approach and the ValleyRidge.py program are three-point- and four-point-methods, respectively, as their numerical solution depends on the specified number of points and variations in the character of any point will produce a different solution. Houk's bond-order method and the procedure herein described are one-point-based, as only qualitative information from other points is used to tailor the method but all numerical values are obtained exclusively from the TS_1 structure.

In this study, we present an approach that quantifies the branching ratio from one key point of the PES from Scheme 1 (that is the TS_1), which is also a good prerequisite for a broader applicability of the method with no limitation to reactions with 'four-point'-defined furcations and without bias towards predefined products. Specifically, the presented method relies on the kinetic energy distribution within the reactive mode at TS_1, as introduced by us and already applied to coupled electron-proton transfer (CEPT) reactivity and post-CEPT selectivity.[26,27] Here, we first concisely summarize the principles of the method (denoted as Reactive Mode Composition Factor – RMCF) and later assess its accuracy in the quantification of product ratios of bifurcating organic reactions and compare its performance with existing protocols. We also briefly discuss the chemical insight provided by the analysis and the limitations of the method.

Computational details

Density functional theory (DFT) calculations

Unless stated otherwise, geometry optimization and vibrational analysis of all the presented structures were performed using the B3LYP[28] functional combined with the def2-TZVP basis set,[29,30] applying the D3 dispersion correction[31] and the conductor-like polarizable continuum model (CPCM)[32] to mimic the solvent environment, if present (further denoted as B3LYP+D3/def2-TZVP(/CPCM)). The same method was applied to derive kinetic energy distribution (KED) factors defined later in the text. Reaction free energies, ΔG, were evaluated from equilibrium geometries using the following equation:

$$\Delta G = \Delta E_{\mathrm{el}} + \Delta[E_{\mathrm{ZPVE}} + pV - RT\ln Q]\ (+\ \Delta G_{\mathrm{solv}}) \qquad (1)$$

where ΔE_{el} is the change of potential energy, $\Delta[E_{\mathrm{ZPVE}} + pV - RT \times \ln Q]$ corresponds to the thermal enthalpic and entropic contributions to the change of the solute energy with E_{ZPVE} and Q being the zero-point vibrational energy and molecular partition function obtained from a frequency calculation (at 298 K, 1 atm; ideal-gas approximation) on top of optimized geometries; the ΔG_{solv} term, calculated using the CPCM method, was only included for those reactions where solvent was reported in the original reference. The functionals ωB97X-D[33] and mPW1k[34] combined with 6-31G(d) basis set[35] for TS optimization and frequency calculations were tested, as recommended by others.[36,37] Quasi-classical molecular dynamics was carried out using the Atom-centered Density Matrix Propagation (ADMP) formalism, with the B3LYP+D3/def2-TZVP(/CPCM) protocol.[38] Trajectories were initialized from the TS_1 structure. The initial total nuclear kinetic energy was set to the zero-point vibrational energy obtained during frequency calculations. Initial velocity vectors for all atoms were set random orientations. Velocities for the *j*-th atom were rescaled every five steps to ensure constant temperature, by the relation $\frac{3RT}{2} = \frac{1}{2}\sum_j m_j v_j^2$. MD was carried out with an integration time step of 0.5 fs and total simulation times of 200 fs in all cases. All calculations were used as implemented in the software Gaussian 16 version B.01.[39]

Results and discussion

Reactive mode composition factor (RMCF) analysis

To understand the usability of the analysis in the space of non-TST bifurcating reactions, we first introduce its physical background. Considering the harmonic approximation, the normal vibrational coordinates Q_α are related to the mass-weighted atomic displacements $\overrightarrow{r_j}\sqrt{m_j}$:

$$Q_\alpha = \sum_{j=1}^{n} \vec{e}_{j\alpha}\overrightarrow{r_j}\sqrt{m_j} \tag{2}$$

through a set of orthogonal unitary vectors $\vec{e}_{j\alpha}$ representing the motion of the j-th atom in the mode α. These vectors also allow to express the atomic kinetic energy $\langle T_j \rangle$ as a linear combination of kinetic energies of normal modes $\langle T_\alpha \rangle$:

$$\langle T_j \rangle = \sum_\alpha e_{j\alpha}{}^2 \langle T_\alpha \rangle = \sum_\alpha \mathrm{KED}_{j\alpha}\langle T_\alpha \rangle, \tag{3}$$

so that $e_{j\alpha}{}^2$ expresses the fraction of kinetic energy of the mode α associated with the motion of the j-the atom (denoted as atomic kinetic energy distribution factor $\mathrm{KED}_{j\alpha}$). The eqn (3) includes the normal modes with real and imaginary frequencies and it is thus applicable to transition states. Importantly, the KED factors are related to the cartesian atomic displacements $\overrightarrow{r_j}$ associated with mode α by

$$\mathrm{KED}_{j\alpha} = \left(\frac{m_j r_j{}^2}{\sum_k^n m_k r_k{}^2}\right)_\alpha, \tag{4}$$

which is readily computable by evaluating a standard vibrational analysis. The appealing feature of this method consists in its simplicity – optimization of one key transition state, TS_1, along with its vibrational analysis that provides $\vec{e}_{j,\alpha=\mathrm{RM}}$ and hence atomic KED factors of the reactive mode ($\equiv e_{j,\mathrm{RM}}{}^2$). Note that KED factors of this mode (as of any normal mode α) are normalized: $\sum_{k=1}^{n}\mathrm{KED}_{k,\mathrm{RM}} = 1$ (n – total number of atoms in the system). A more detailed description for the RMCF theory and its applications can be found in ref. 26.

A crucial part of the approach is to group atomic KED factors to N sets accounting for fractions of kinetic energy of the TS_1 reactive mode, which are differentially available for subsequent N reactive channels. In case of two competing post-TS_1 reactive channels A and B, it requires selection of two disjunctive groups of atoms and evaluate their $\mathrm{KED}_{\mathrm{RM}}$ factors at TS_1 –$\mathrm{KED}_\mathrm{A}(=\sum_{j=1}^{p}\mathrm{KED}_{j,\mathrm{RM}})$ and $\mathrm{KED}_\mathrm{B}(=\sum_{i=1}^{q}\mathrm{KED}_{i,\mathrm{RM}})$. A general prescription for such a selection is described in the following section. The percentage of the product $\mathrm{P_A}$ (further denoted as the product branching ratio χ_KED) resulting from the competition between two reactive channels is then defined as:

$$\chi_\mathrm{KED}\ (\%) = 100 \times \frac{\mathrm{KED_A}}{\mathrm{KED_A} + \mathrm{KED_B}} = 100 \times \frac{\sum_{j=1}^{p}\mathrm{KED}_{j,\mathrm{RM}}}{\sum_{j=1}^{p}\mathrm{KED}_{j,\mathrm{RM}} + \sum_{i=1}^{q}\mathrm{KED}_{i,\mathrm{RM}}} \tag{5}$$

where all terms are explained above. Eqn (5) can be readily adopted for systems with $N > 2$ reactive channels. We note that previous studies have demonstrated that the redistribution of excess vibrational kinetic energy among real modes has an impact on the selectivity of non-equilibrium processes.[40–42] In contrast to existing methods based on the analysis of real vibrational modes, the presented RMCF analysis is unique in its dissection of the imaginary reactive mode, translating the distribution of kinetic energy within this mode into a predictor of selectivity in complex reactions.

TS partition for the calculation of branching ratios

Regardless of the reaction type involved, the workflow for TS_1 partition that proved to be robust and reliable to all the reactions studied in this study is as follows:

(1) Identify the atom pairs directly involved in the formation/cleavage of the bonds associated with each of the N relevant products and ascribe these pairs to N different partitions. Exclude the cleaved/formed bonds that are common to all products.

(2) Include to each partition all directly bonded H atoms.

(3) Include to each partition only the directly bonded heavy atoms (along with their H-atom caps), which are not covalently bridged to the opposite partition.

(4) Ignore all further atoms.

For most reactions, at least one bond exclusive to each product is discernible at TS_1 and, thus, fragment definition is unambiguous. In cases where only one bond is formed and one is cleaved, as in reactions **13–16**, **24** and **52**, the modified instructions are to be followed:

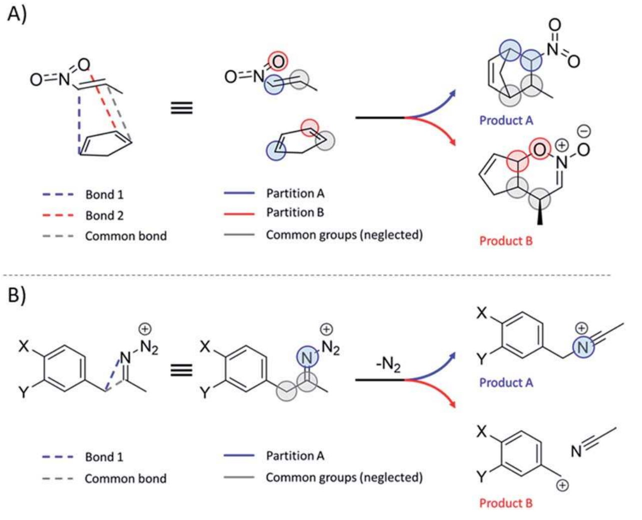
(1) Identify the atom pair directly involved in the formation of the unique bond. Include in this partition all directly bonded H-atoms. This fragment corresponds to channel A.

(2) Identify the atom pair ascribed to the bond scission or formation that is common to all products. Include in this partition all directly bonded H atoms.

(3) The fragment B is composed of all atoms that are not ascribed to A nor to the common fragment.

Scheme 2 illustrates the above-described workflow for TS partition in terms of atomic KED factors that are grouped into two relevant partitions for each upcoming reaction channel. The selection of the relevant atomic pairs can be inferred from the TS_1 structure, and we provide as ESI† a python program (rmcf.py) to expedite this process by performing the following tasks:

(1) Calculation of the kinetic energy distribution within the reactive mode.



Scheme 2 (A) Partition of the reactive system at TS_1 into two groups of atoms ascribed to two competing channels in bifurcating reactions presented in this study. Note that these groups also include hydrogen atoms that are not explicitly visualized. Each of two groups is associated with a fraction of kinetic energy of the reactive mode at TS_1 (see the main text), which is utilized for evaluation of the product branching ratio χ_{KED} using eqn (5). (B) Partitioning scheme for reactions where one bond is cleaved and only one bond is formed.

(2) Displacement of the TS geometry along the reactive mode.

(3) Gauge of all interatomic displacements and their ranking as a list of potential bond formation/breaking events.

By inspection of the provided list of potential bond formation/cleavage events, the user can identify chemically relevant bonds and follow the provided workflow for TS partitioning. The selection of relevant atomic pairs is not completely automatic, but this provides the flexibility needed to screen numerous potential products from a single TS structure. We refer the reader to Table S1† where detailed partitions for all reactions are included. In addition, we compare alternative partitioning schemes for all reactions in ESI (Tables S2 and S3†), corroborating the outlined selection strategy. As demonstrated in ESI (Table S4†), the use of alternative functionals ωB97X-D and mPW1k combined with the 6-31G(d) basis set has a minor impact on the results from the reactive mode composition factor analysis.

Studied bifurcating reactions

In the first part of the presented study, we use a collection of 48 bifurcating reactions, which were computationally investigated previously by Lee and Goodman in ref. 24 and for which the referential data (*i.e.*, experimental or MD-based product branching ratios) were reported in the literature. All these reactions are given in Fig. 1 and all associated TSs are visualized in Fig. S1† (for more detailed information on ratios and solvent conditions, see Table S5†). The studied reactions cover a broad spectrum of reactions, featuring a rich set of bifurcating pericyclic processes (**1–12, 23–34, 47–48**),[14,43–56] nitrene insertions (**35–38**),[57] rearrangement and fragmentation reactions (**13–16, 39**),[13,58–60] the branching in nucleophilic additions and substitutions (**17–22, 46**),[11,61] and solvent-dependent isomeric Pummerer rearrangements (**40–44**).[12]

RMCF analysis of transition states from bifurcating reactions

The product ratios calculated from eqn (5) and their comparison with experimental and MD data for all reactions depicted in Fig. 1 are presented in Fig. 2. From that, we found that RMCF correctly determines the major product in 89% of the studied reactions (*i.e.*, **41** from the **46** bifurcating reactions) and, for **21** and **45**, it yields 50% : 50% distributions (with <10% error) over products A and B, in excellent agreement with the referential data. This is apparent from Fig. 2A, where the top-right quadrant of the plot is most populated. The exceptions are reactions **5, 12, 15, 16** and **22**, for which the referential major products are predicted by RMCF to be formed in minority. In addition, Fig. 2B shows that an unsigned deviation from the referential data (KED_{error}) is only ≤10% for 25 and ≤20% for 34 of the total 48 cases.

Considering the computational subset of references, the RMCF shows a very good performance. Namely, KED_{error} falls in the range of ≤10% for 20 and ≤20% for 27 of the 38 reactions (Fig. 2, the orange bars). In comparison with the available experimental data, RMCF is found to perform even better: KED_{error} of ≤10% for 5 and ≤20% for 8 of the 10 cases is observed (Fig. 2, blue bars). We consider of utmost importance the comparison with available experimental data.§ In this context, the RMCF analysis is capable of correct determination of the major product in 89% of studied cases, with a correct quantification (with a tolerance of <20%) in 80% of them. The results of the analysis of all reactions in the present work with the rmcf.py program are summarized in Table S6.†

Selectivity and thermodynamic driving force in bifurcating reactions

According to a traditional linear free energy relationship (LFER), a more exergonic reaction within a set of closely related reactions tends to have a lower barrier and hence to proceed preferably. Namely, if reaction A is more exergonic than B (*i.e.*, $\Delta G_{AB} < 0$), the barrier for A tends to be lower than that for B, which leads to a larger ratio in favor of P_A. In the case of bifurcating reactions, the situation is slightly different since both reactions A and B share a common barrier, and only a frugal number of discussions relate the selectivity of bifurcating reactions with their corresponding driving forces.[4,62] Some selected examples have been pointed out in the literature,[63] where the effect of ΔG_{AB} is overridden by the dynamic match between atomic momenta at TS_1 and an upcoming reaction channel. Whether the occurrence of these examples is common or only a minority has, to our knowledge, not been addressed in a broad chemical space. The present set of ∼50 diverse reactions enriches the pool of data correlating ΔG_{AB} and the excess of one of the products, $\Delta\chi_{AB}$ and these results are shown in Fig. 3. All ΔG values are condensed in Table S7.†

As seen in Fig. 3, it is clear that the effect of thermodynamic driving force on A *vs.* B selectivity in bifurcating reactions

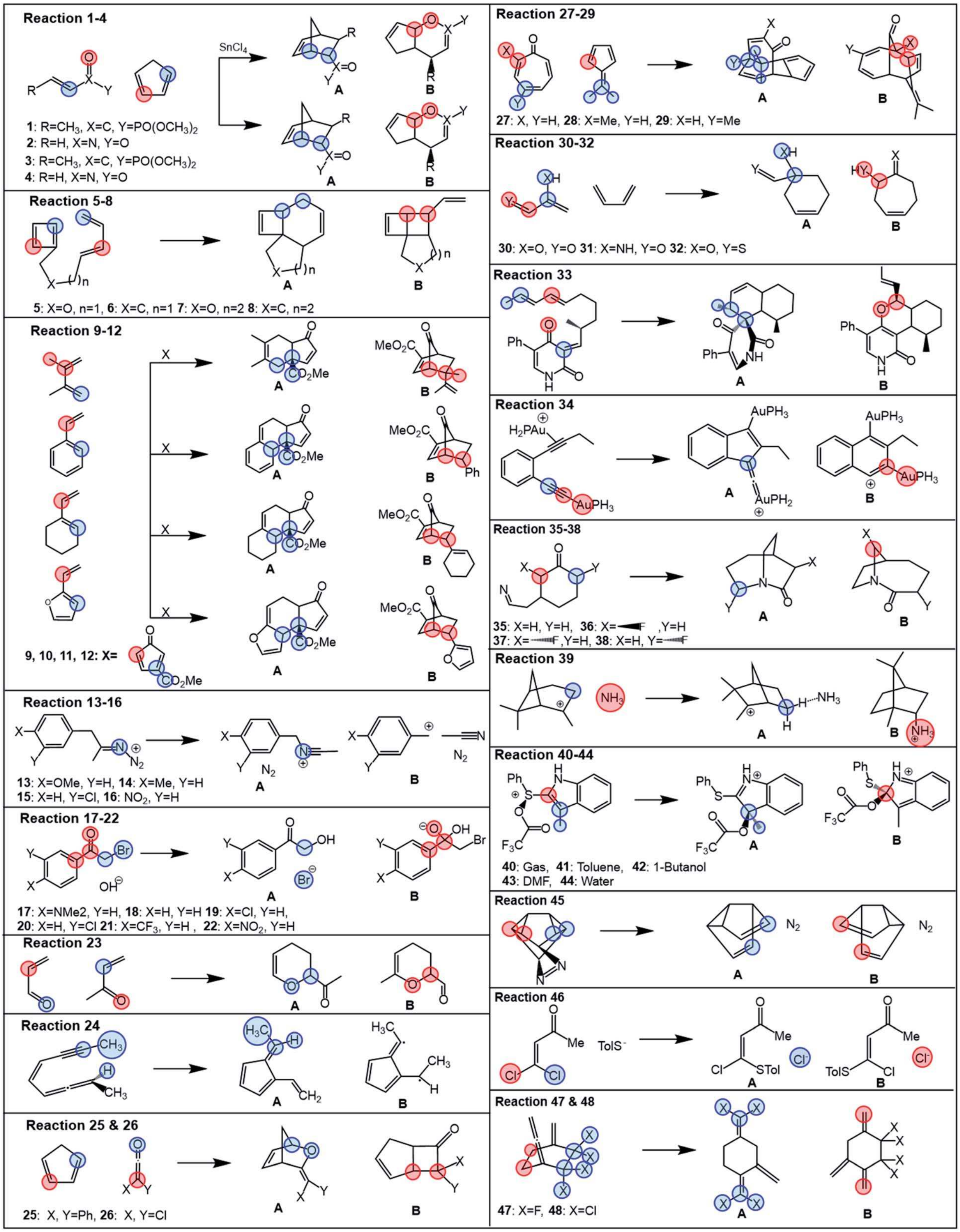


Fig. 1 Bifurcating reactions considered in this study. Referential data for product ratios are taken from the literature (references in the main text). Partition of each reactive system for evaluation of branching ratio (χ) is carried out according to the description in Scheme 2. Blue and red colors correspond to partitions A and B, specified in more detail in Table S1.†

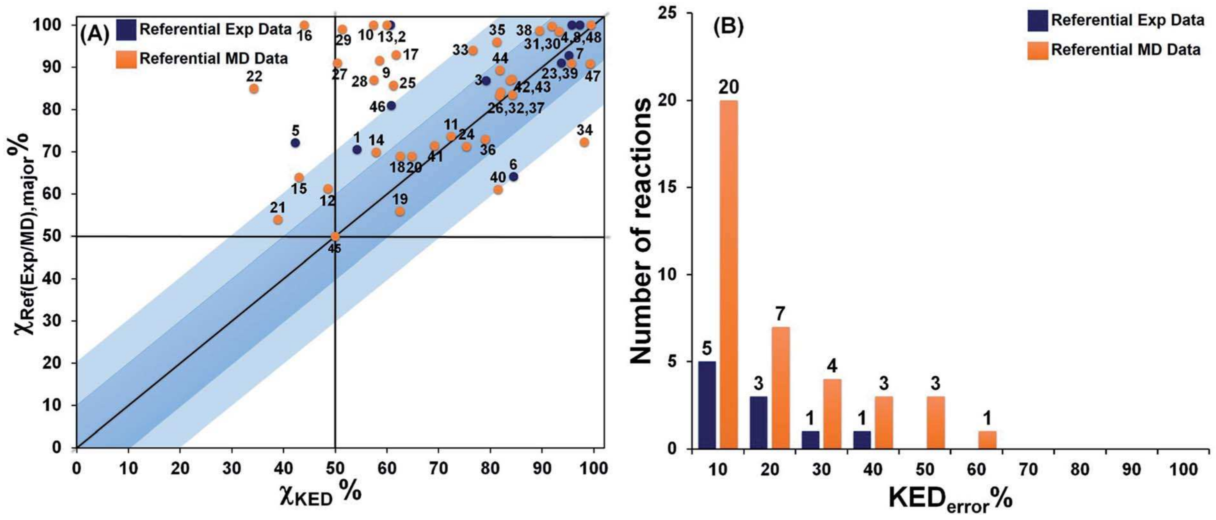


Fig. 2 The correlation between χ_{KED} from eqn (5) and the referential $\chi_{ref(exp/MD),major}$ values, as obtained from experiment or MD simulations (blue or orange points) for reactions **1–48** from Fig. 1 are presented (panel A). Note that χ_{KED} refers to the product that is determined as a major product in the reference. The distribution of the number of reactions according to the unsigned deviation $KED_{error} \equiv |\chi_{KED} - \chi_{ref(exp/MD),major}|$) are shown in panel B. Note that more details are provided in Table S5.†

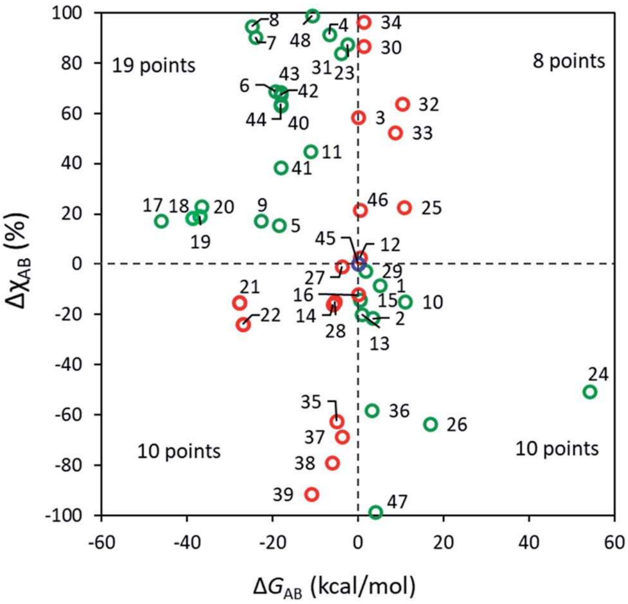


Fig. 3 Correlation between the relative stability of P_A *vs.* P_B (ΔG_{AB}, B3LYP+D3/def2TZVP) and the excess of product A ($\Delta\chi_{AB}$, from RCMF analysis) for reactions **1–48**. For reactions in green (29 cases, 60%) the major product is the thermodynamically favored, whereas those in red (18 cases, 38%) favor a less exergonic product. Reaction **45** is colored blue and not ascribed to any quadrant due to its exact 50 : 50 product ratio and ΔG_{AB} = 0 kcal mol^{-1}.

should only be considered with high caution. For 29 out of 48 reactions (60% of cases), the major product is indeed more stable, although no clear correlation can be discerned between ΔG_{AB} and the excess of P_A, even for apparently similar processes. For example, reactions **5** (with ΔG_{AB} = 18.4 kcal mol^{-1} and $\Delta\chi_{AB}$ = 15%) and **6** (ΔG_{AB} = 19.4 kcal mol^{-1} and sizeable $\Delta\chi_{AB}$ = 69%) share the same mechanism yet their outcome does not follow an intuitive LFER-like trend. For 18 reactions the major predicted product is less stable demonstrating that, in a substantial number of cases, the local curvature of the PES at TS_1 overcomes the effect of having product basins of different depths. The RMCF method correctly predicts the major product for 89% of the reactions (for 71% within a 20% error) implying that the post-TS dynamics for most of the studied reactions is encoded in TS_1. Our present results demonstrate that the dissection of TS structures using the RMCF approach is semiquantitatively predictive even in complex bifurcating PES where LFER analyses might be inconclusive or lead to incorrect predictions.

Comparison of RMCF with existing approaches designed for product ratio evaluation

Here, we assess the performance of three alternative computational PES-topography based procedures relative to the RMCF analysis: (i) Goodman's four-point algorithm,[17] which involves the ambimodal TS_1 along with TS_{AB}, P_A and P_B points of the PES as illustrated in Scheme 1; (ii) Carpenter's 'dynamic match' based on TS_1, P_A and P_B,[19] and (iii) an approach put forward by Houk,[20] which correlates bond order ratios at the ambimodal TS_1 with the product ratio. For the sake of comparison with RMCF, we use the B3LYP+D3/def2-TZVP(/CPCM) level of theory for all these protocols.

In our hands, Goodman's 4-point algorithm could be applied to 45 out of the 48 reactions when the B3LYP + D3/def2-TZVP method is used (all attempts to calculate **15**, **16** and **34** were unsuccessful, as described in ESI†). For this reason, reactions **15**, **16** and **34** will be excluded in all forthcoming comparisons between methods, to treat all protocols on an equal footing. In addition, reactions **21** and **45** are not considered for the ranking of selectivity predictions, as their ratios are within 45–55% and no major product can be discerned neither experimentally nor computationally in such cases.

Goodman's method correctly predicts the major product in 81% of 43 reactions from Fig. 1, with the error below 10% for 53% of them (and below 20% for 71% of cases) as shown in Fig. 4A and S3.† As evidenced by reactions **15** and **16**, where TS_{AB} could not be located, the strict requirement of four optimized structures may render it less applicable to broader types of reactions. Further, the explicit tailoring of Goodman's method towards bifurcating reactions turns it increasingly demanding for higher order furcations, as in the case of trifurcations, where a division into three competing bifurcations with a total input of 7 stationary points was necessary in the original ref. 24. An important remark is that Goodman's 4-point approach was tested on reactions calculated with different methods, some of them selected by the original authors to maximize the agreement with experiments.[28,29,40] Under such conditions, both Goodman's and RMCF methods perform even better (Fig. S4 and S5†), yet this heterogeneity precludes the selection of a generally reliable method. For the same set of reactions, we herein show that with qualitatively correct selectivity in 93% (40 out of 43) and with the correct quantification (with a tolerance of 20%) in 71% of cases, the RMCF analysis with the B3LYP functional proves to be a general and balanced prescription requiring a single point from the PES complemented by qualitative information of the suspected products or chemical knowledge from the user.

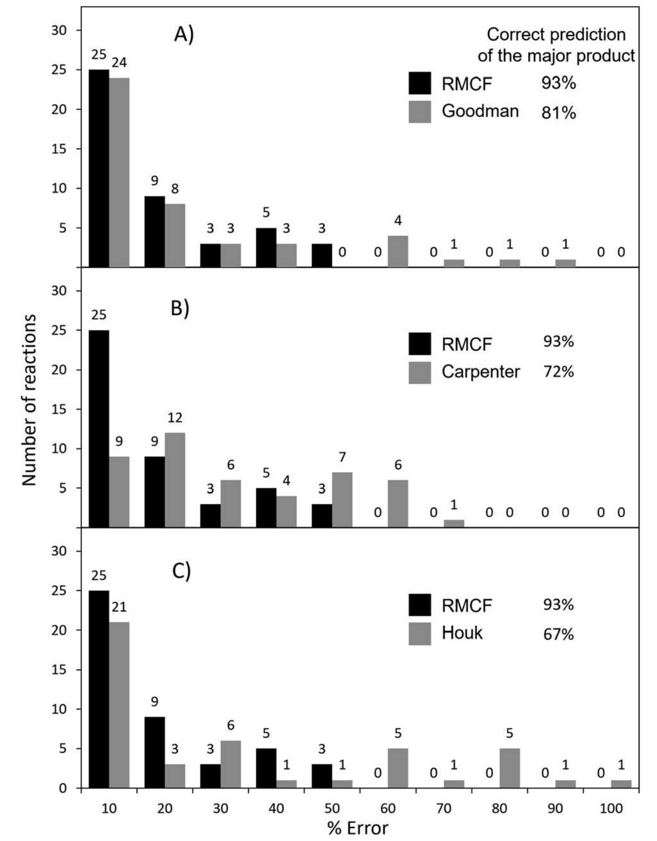


Fig. 4 The performance of the RMCF approach *versus* the performance of three other PES-topography based approaches: the Goodman's algorithm from ref. 24 (panel A), Carpenter's dynamic match from ref. 19 (panel B) and Houk's bond-order based fitting function from ref. 20 (panel C). The performance of all four methods is assessed with respect to referential (experimental or MD) data for reactions presented in Fig. 1. Reactions **21** and **45**, with referential ratios of 45–55% were not considered for evaluating major product predictions.

In case of Carpenter's dynamic match, it correctly predicts the major product for 72% of reactions from Fig. 1 but the quantification of the product ratio is considerably less successful: only 20% (and 47%) of reactions fall within the deviation of 10% (and 20%) from referential data (Fig. 4B, S6 and S7†).

Finally, Houk's method of bond order ratios (BOR) at the ambimodal TS_1 relies on a linear fit to a training set of reactions and is expected to work adequately for processes which are closely related to those used in the fit. It determines the correct major product for 67% of the reactions in Fig. 1, with 47% (and 54%) of all reactions predicted within error of 10% (and 20%) from the reference data (Fig. 4C, S8 and S9†). The BOR and RMCF methods were also applied to the training set of 15 reactions used the original ref. 20 and the results are given in ESI (Table S8, Fig. S10 and S11†).

Robustness of the DFT protocol

Since the existing pool of computational studies have employed a broad gamut of density functionals to tackle each specific problem, we compared the accuracy of B3LYP+D3/def2-TZVP(/CPCM) with that of other DFT-based levels of theory employed in the original references (see ESI for more details, Table S9 and Fig. S4†). For 43 out of 48 bifurcations, we observe that B3LYP yields results comparable to those obtained with the approaches used in the original references, *i.e.*, with a difference of <10% (Tables S4 and S10†). For two of the five remaining reactions, B3LYP improves product ratios by >10% (**25** and **42**). Contrarily, the B3LYP results for **1**, **10** and **40** are worsened by >10%. Despite this robustness, the tendency of the B3LYP-based approach to overestimate asynchronicity of highly asynchronous TSs has been recently pointed out.[28,29] In such extreme cases, the ωB97XD/6-31G(d)(/CPCM) and mPW1k/6-31G(d)(/CPCM) levels of theory were proposed as reliable alternatives. Thus, we recalculated with them all organic reactions from Fig. 1. The results are summarized in Table S5.† With mean unsigned errors of ∼16% for B3LYP, and ∼15% for ωB97X-D and mPW1k functional based calculations, there is no significant advantage for any of the alternative functionals over B3LYP (see Fig. S12†). Overall, we conclude this section stating that the B3LYP-based methodology is sufficiently robust for the calculation of reactive mode composition factors and hence distributions of bifurcating reaction products.

Performance of the RMCF method with unseen ambimodal reactions

The partition scheme for transition states was selected to guarantee maximum robustness and applicability of the RMCF method to the broadest possible chemical space. As such, reactions **1–48** served as a training set for the model. Next, we proceed to evaluate its performance on a test set of 13 additional reactions, **49–61**,[64–70] shown in Fig. 5.

The RMCF method retains well its robustness on these reactions, predicting product distributions with errors of <20% for 77% of the tested reactions, with mean unsigned error of 14% (Fig. S13, Tables S10 and S11†). This demonstrates the adequate performance of the method on general organic reactions beyond the set employed during its development.

Among the test set, reaction **55** is striking as 11 different stationary points were obtained from MD simulations[68] (the structure of all the products accessible from TS_{55} and their yields from MD, as reported in ref. 68, are condensed in Fig. S14†). The discrimination of all 11 products is admittedly out of the scope of the presented method, as explained in detail in ESI (Table S14† and accompanying discussion). Nonetheless, as the authors note, this dauntingly complex reactive system lands predominantly on species **55A** and **55B**, each as a mixture of two conformers. On the basis of this observation, we treated this reaction as a bifurcating system and applied the RMCF analysis predicting a 57% : 43% ratio, in good agreement with the MD result of 59% : 32%. The remaining 9% (products **55C–55H**, Fig. S14†) cannot be accounted for by the RMCF method.

Also of interest is the set of reactions **59–61**, involving different tropones and cycloheptatriene. It has been suggested that these processes involve trifurcating PES, where a single TS_1 leads to three different products. In all these cases, the rmcf.py program does not predict any propensity for formation of the bond leading to product C in line with MD carried out by Houk

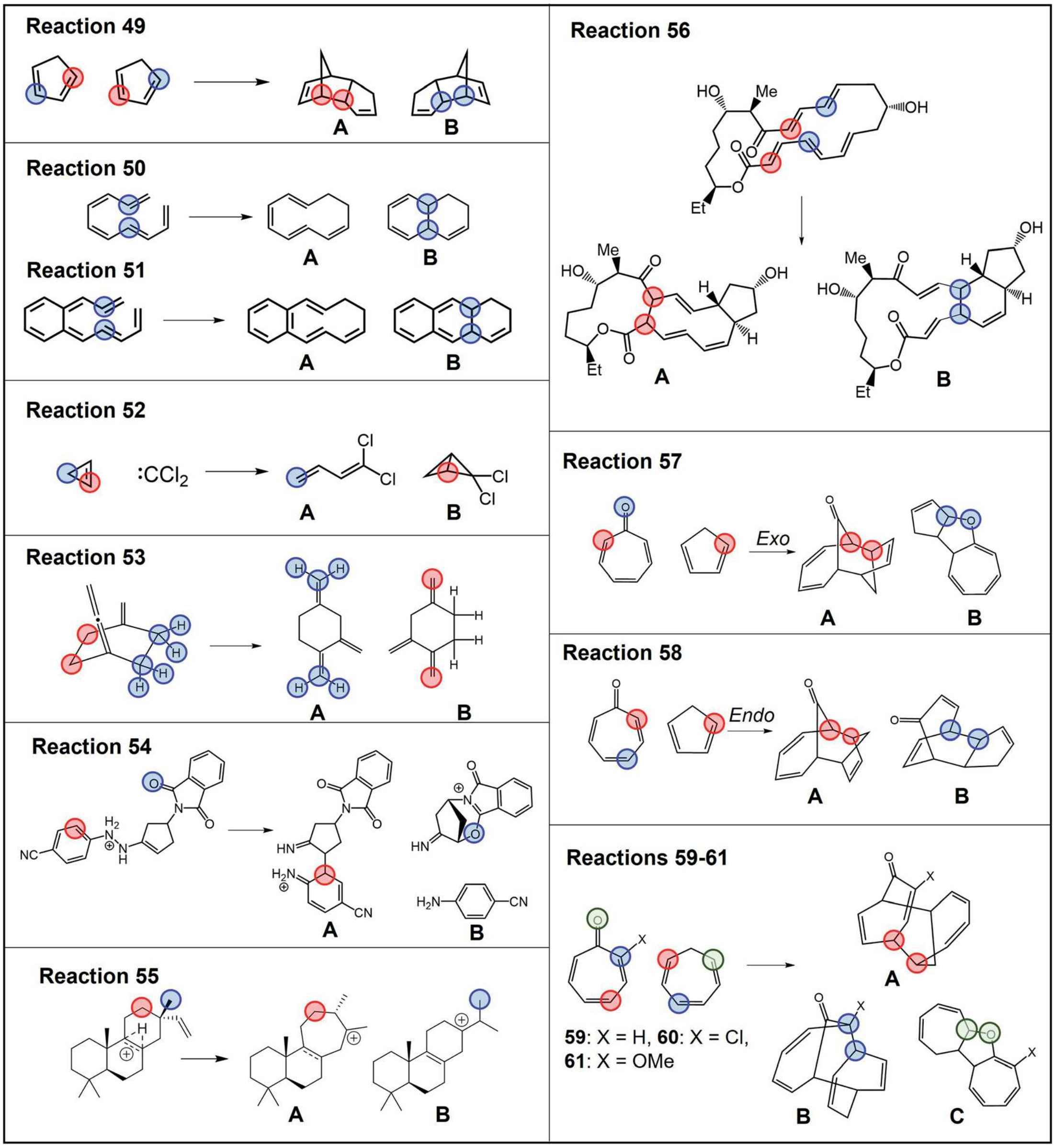


Fig. 5 Test set of ambimodal reactions **49–61** (ref. 64–70) not included in the selection of the partitioning scheme. Partition of each reactive system for evaluation of branching ratio (χ) is carried out according to the description in Scheme 2. Blue, red and green colors correspond to partitions A–C, specified in more detail in Tables S10 and S11.†

and coworkers, which yielded 1% of **59C**, 1% of **60C** and 0% of **61C**.[65] Hence, we treated these reactions as bifurcations, obtaining ratios of 39% : 61% (*vs.* Houk's MD ratio of 36% : 58% for **59A : 59B**), 42% : 58% (*vs.* 63% : 30% for **60A : 60B**) and 36% : 64% (*vs.* 55% : 37% for **61A : 61B**). The origin of the discrepancies for **60** and **61** is unclear but it is likely is a consequence of steric interactions, as the authors mention in the original work, and which would be consistent with the erroneous prediction of the major product in reaction **5** (see the section Advantages and limitations of RMCF).

Application to reactions featuring statistical and nonstatistical contributions

There exist cases, where a reaction might present statistical steps (*i.e.*, the selectivity between two or more transition states can be estimated using Eyring's TST) and nonstatistical steps (where the RMCF analysis can predict product distributions). We will exemplify this application as an additional test of the versatility of the RMCF protocol. The first example is the reaction between dichlorocarbene and 2-methylbenzocyclopropene (reaction **62**, Fig. 6). In agreement with the original report,[63] we predict two isomeric and nearly degenerate transition states ($\Delta\Delta G^{\neq} = 0.5$ kcal mol^{-1}), which can both bifurcate to yield the isomeric products **62A** and **62B**. By applying TST, we estimate a statistical 70% : 30% partitioning between both transition states. RMCF analysis of them yields nonstatistical ratios of 95% : 5% and 9% : 91% to products **62A** and **62B** (Table S12†). Combining these results we predict a 69% : 31% ratio, in excellent agreement with the experimental quantification that is 68% : 32%.

A final test to our protocol is the challenging tripericyclic reaction between 8,8-dicyanoheptafulvene and 6,6-dimethylfulvene. Houk *et al.* reported an ambimodal TS leading to the formation of [4 + 6], [6 + 4] and [8 + 2] cycloproducts (**63A**, **63B** and **63C** in Fig. 7).[71] From MD simulations, all three cycloadducts are formed in the ratio of 87% : 3% : 3%. However, according to experiments of Liu and Ding, there are only two detectable cycloadducts **63B** and **63C**, with the ratio of 54% : 46%.[72,73] To reconcile this discrepancy, Houk proposed that the kinetically favored major product **63A** undergoes rapid conversion to the thermodynamically more stable cycloadducts **63B** and **63C**, although no estimation of the final **63B : 63C** ratio was possible from MD studies and thus the link between the trifurcation outcome and the experimentally observed selectivity in this process was not addressed.

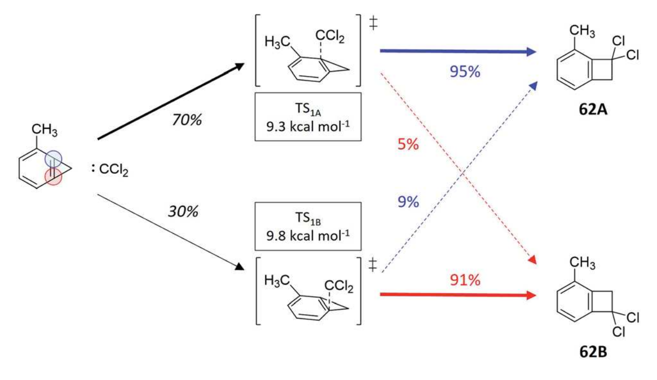


Fig. 6 Reaction between dichlorocarbene and 2-methylbenzocyclopropene (ref. 63), featuring two energetically close transition state, each of them showcasing a different branching ratio to the experimentally observed products. Ratios in black were obtained from TST and colored percentages from RMCF analysis.

To address it, we first calculated KED values for ambimodal **63-TS$_1$** to determine a product distribution of 22% : 26% : 52% for **63A**, **63B** and **63C**, respectively (Table S13†). In agreement with ref. 72, we observe that **63A** must be redistributed over **63B** and **63C** since it undergoes two bifurcating Cope rearrangements *via* two energetically comparable ambimodal transition states, TS$_{2A}$ and TS$_{2B}$ (Fig. 8). Since **63A** originates from an 8.2 kcal mol^{-1} descent from TS$_1$ and the upcoming barriers to **63B** and **63C** are 10.0 and 11.4 kcal mol^{-1}, we expect a randomization of momenta in the **63A** basin. This renders TST applicable to estimate the selectivity between these two channels. The calculated preference of TS$_{2A}$ over TS$_{2B}$ by 1.4 kcal mol^{-1} yields a ratio of 91 : 9 for the transformation of **63A** to **63B** and **63C**. Applying now the RMCF protocol, we foresee that TS$_{2A}$ and TS$_{2B}$ favor 99% of **63B** and 98% of **63C**, respectively. Thus, the initial fraction of 22% for **63A** is eventually partitioned ∼9 : 1 between **63B** and **63C**, leading to the final ratio of 46% : 54%, with a deviation of only 8% from experiment. The ωB97X-D- and mPW1k-calculated KED values obtained for TS$_1$ predict relative abundance of **63A**, **63B** and **63C** to be 42% : 16% : 42% and 44% : 16% : 40%, respectively (see Fig. S15†). The subsequent post-TS$_{2A}$ and post-TS$_{2B}$ bifurcations towards the final products **63B** and **63C** yield a 49% : 51% ratio for ωB97XD and 43% : 57% for mPW1k, in agreement with B3LYP.

These results show that the RMCF analysis can be combined with TST for the quantitative analysis of complex reactions involving both statistical and nonstatistical contributions, including reactions with $N > 2$ post-TS channels.

Advantages and limitations of RMCF

In light of other methods, RMCF is the best balanced with respect to simplicity and accuracy. It is comparably simple to Houk's BOR approach since it requires a minimal input, which is only one point from the PES. When the kinetic energy distribution at the reactive mode is complemented with qualitative information about the expected products, it reaches the accuracy of (or even surpasses) the four-point algorithm of Goodman *et al.* (Scheme 3).

One-point methods will prove especially advantageous when additional points of the PES are inaccessible or computationally expensive to optimize, and when more than two products emerge from TS$_1$. However, it is worth noting that reactions **5** and **13–16** are mostly out of reach for RMCF and the three other methods tested here, demonstrating that complex trajectories are still a challenge for simplified models. In the case of **5**, the motion at the TS$_1$ structure points towards (4 + 2) cycloaddition, which in fact should be a less favored pathway. Thus, it seems that a tight –CH_2OCH_2– tether outweighs this kinetic-energy

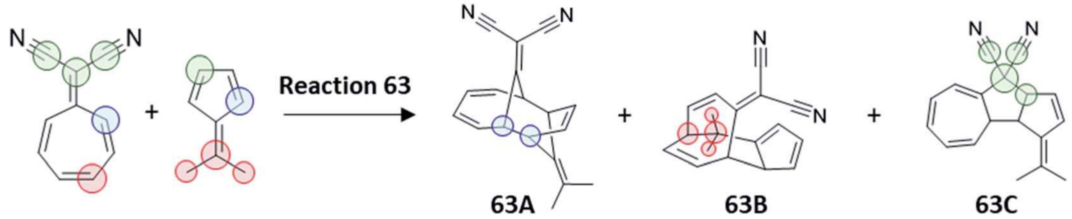


Fig. 7 Trifurcating cycloaddition reaction **63**, between 8,8-dicyanoheptafulvene and 6,6-dimethylfulvene. The partition of TS_1 required for calculation of χ_{KED} of products is highlighted by colored circles following the prescription in Scheme 2.

propensity at TS_1, which eventually leads to the (2 + 2) product. An analogous example was pointed out by us in the past in the context of 'rebound' hydroxylation *vs.* dissociation in post H-atom abstraction reactivity,[27] where a reaction poised for dissociation in terms of KED follows the hydroxylation channel due to tethering. Such examples demonstrate the possibility to tilt the selectivity against the kinetic energy distribution at TS_1.

One noticeable limitation of the method stems from the lack of a temperature dependence in the RMCF analysis. While a thermostat can be routinely applied in MD simulations so that product ratios can vary as a function of temperature, the eigenvalues of the diagonalized hessian matrix (and, consequently, kinetic energy distributions) are independent of the temperature. However, there is only a little evidence of a change in branching ratios emanating from a change in temperature. For example, experimentally determined product ratio for thermolysis of enyne-allenes (a reaction similar to **24**) changes in the range of 2–4% as T is increased from 30° to 100 °C.[74] Another relevant example is the activation of CH_4 by MgO^+, which has been studied experimentally and by means of MD.[75] Since the calculated PES is expected to display a shallow intermediate, this process cannot be considered a canonical bifurcating reaction. However, MD trajectories bypassed such intermediate, leading directly to either of the two accessible reactive channels akin to a bifurcating reaction. Remarkably, only a small product redistribution (3%) was observed in going from 300 to 600 K, suggesting that dynamically controlled reactions might be relatively insensitive to changes in temperature and thus amenable to RMCF analysis. The influence of temperature on the outcome of branching reactions remains an underdeveloped area.

Regarding the presence of shallow intermediates in branching PES and their influence on selectivity, we demonstrated in the past the successful application of the RMCF methodology to the rebound *vs.* dissociation dichotomous mechanism in C–H activation reactions by iron-oxo species.[27] In this context, a nascent carbon-centered radical can either (1) dissociate out of the solvent cage and become susceptible to trapping and further transformations, or (2) 'ballistically' recombine with the Fe–OH species in a nonequilibrium

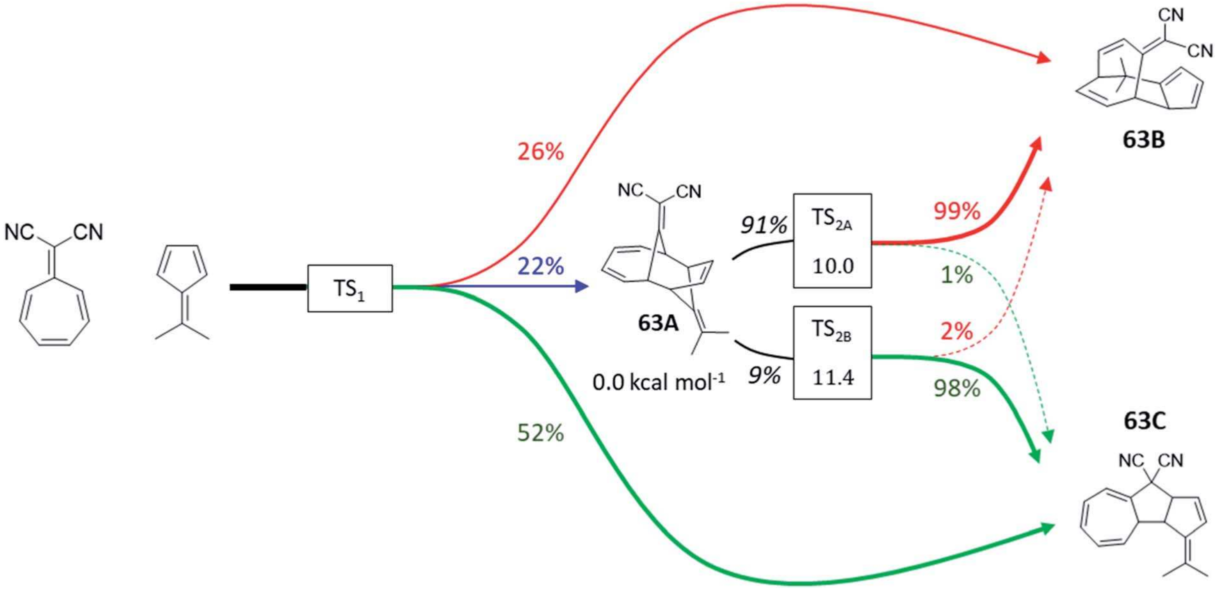
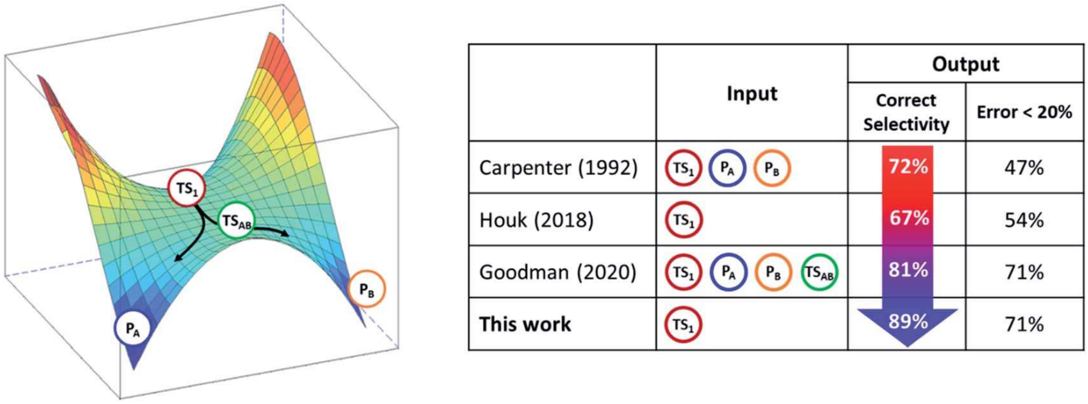


Fig. 8 Tripericyclic reaction between 8,8-dicyanoheptafulvene and 6,6-dimethylfulvene in chloroform. Only **63B** and **63C** were observed experimentally (ref. 54). Two bifurcating Cope rearrangements occurring *via* two ambimodal transition states TS_{2A} and TS_{2B} act as exit channels from the unstable **63A** species. The product distributions are calculated at the B3LYP+D3/def2-TZVP/CPCM level of theory. Ratios in black were obtained from TST and colored percentages from RMCF analysis.



	Input	Output	
		Correct Selectivity	Error < 20%
Carpenter (1992)	TS$_1$ P$_A$ P$_B$	72%	47%
Houk (2018)	TS$_1$	67%	54%
Goodman (2020)	TS$_1$ P$_A$ P$_B$ TS$_{AB}$	81%	71%
This work	TS$_1$	89%	71%

Scheme 3 The transition state TS$_1$ in bifurcating reactions encodes enough information to forecast the product ratio. The analysis of reactive mode composition factors (RMCF) provides an intuitive means to achieve this. Correct selectivity implies the prediction of the correct major product. The measures reflect method performance on the reaction set from Fig. 1.

process. The kinetic energy distributions calculated using the RMCF protocol provided a clear and predictive rationalization of the selectivity observed in such cases. Also relevant are the MD studies on the cycloaddition between cyclopentadiene and dichloroketene (reaction **26**) by Singleton and coworkers,[37] where they observed that the B3LYP PES features a shallow intermediate, yet most trajectories bypass it suggesting that the influence of such intermediates might be only marginal, in agreement with the MD studies on CH_4 activation by MgO^+ in ref. 75. Recently, Goodman's 4-point method was extended to also account for shallow intermediates by the inclusion of the intermediate as a fifth stationary point.[25]

Conclusions

The Reactive Mode Composition Factor (RMCF) analysis and its application to bi- and multifurcating reactions and their product distribution is presented. This significantly extends the portfolio of non-TST behaving reactions whose selectivity is reliably predicted by RMCF. As we demonstrate in the present work, the protocol allows the quantification of product ratios for bifurcating reactions by partitioning the kinetic energy distribution (KED) of the reactive mode of the shared transition state into chemically meaningful and well-defined fragments. A theoretical connection between KED with branching ratios was postulated and its validity was tested on a set of >60 bifurcating reactions. To expedite the application of the method, a program (rmcf.py) to perform the RMCF analysis on transition states calculated with the Gaussian software is provided and complemented by a robust workflow to aid the partition any arbitrary (ambimodal) TS structure.

Regarding its power, the RMCF analysis compares favorably with existing computational protocols designed to predict branching ratios, outperforming all tested alternatives in predicting major products and yielding comparable results to the best performing method reported so far, while requiring as input a single transition state connecting the reactant complex with the available product channels, complemented by qualitative information about the products of the bifurcation. The method proved to be capable of qualitatively correct predictions for 89% from a set of *ca.* 50 branching reactions presented in Fig. 1, yielding quantitative and semiquantitative (<10% and <20% deviations) for 52% and 71% of reactions, while providing an intuitive picture of the motion signatures responsible for the predicted ratios. The method can be applied in tandem with traditional transition state theory, to tackle reactions where statistical and nonstatistical steps occur sequentially, providing excellent agreement with experimental and MD results.

Our results also demonstrate that, despite the thermodynamic bias towards a given reactive channel, a linear-free energy relationship rationale is generally not applicable to anticipate the selectivity in bifurcating reactions, where atomic motion along the PES can override the post-TS curvature imprinted by different reaction driving forces. The success of the RMCF method under such circumstances pinpoints that the reactive mode of TS structures encodes sufficient information to reliably predict branching ratios at a fraction of the cost of MD simulations.

Overall, the RMCF protocol is a versatile tool for the prediction and understanding of bi- and multifurcating reactions. Considering its intuitive application to the analysis of chemical reactions, coupled to the *ad libitum* partition of transition states and the quantification of kinetic energy distributions within reactive modes that the method allows, we foresee that this study will foster the update of known reactions with yet unexplained facts and also aid in the understanding of novel and puzzling reaction mechanisms.

Data availability

The datasets supporting this article have been uploaded as part of the ESI.†

Author contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The project was supported by the Grant Agency of the Czech Republic (Grant No. 21-10383S) and by The Ministry of Education, Youth and Sports from the Large Infrastructures for Research, Experimental Development and Innovations project 'e-Infrastructure CZ – LM2018140'. We thank Barry Carpenter for his generosity in providing us the code for calculation of bifurcation ratios with the dynamic match approach. We also thank Jonathan Goodman and Sanha Lee for their helpful suggestions to calculate bifurcation product ratios using their ValleyRidge.py program. We also thank Santiago Alonso-Gil for his help in writing the Fortran subroutine included in the rmcf.py program.

Notes and references

§ Experiments and simulations complement each other in this context. In case of reactions **1** and **2**, where low free energy barriers might facilitate product interconversion, we carried out MD to ascertain the reported experimental values. These results can be found in Fig. S2.† A relevant example of the limitation of MD simulations and the importance of experiments is reaction **34**, where only product P_A is obtained experimentally yet the reported ratio from MD is 72%. The authors admit (ref. 44) that the ratio is preliminary since only 29 trajectories were productive. The RMCF method predicts a 98% ratio of P_A, in good agreement with the experiment, yet we used the reported MD result for comparison with Goodman's protocol.

1 *Modern Physical Organic Chemistry*, ed. E. V. Anslyn and D. A. Dougherty, University Science Books, Sausalito, CA, 2006.
2 J. L. Bao and D. G. Truhlar, *Chem. Soc. Rev.*, 2017, **46**, 7548–7596.
3 D. H. Ess, S. E. Wheeler, R. G. Iafe, L. Xu, N. Çelebi-Ölçüm and K. N. Houk, *Angew. Chem., Int. Ed.*, 2008, **47**, 7592–7601.
4 J. Rehbein and B. K. Carpenter, *Phys. Chem. Chem. Phys.*, 2011, **13**, 20906–20922.
5 S. R. Hare and D. J. Tantillo, *Pure Appl. Chem.*, 2017, **89**, 679–698.
6 D. A. Singleton, C. Hang, M. J. Szymanski, M. P. Meyer, A. G. Leach, K. T. Kuwata, J. S. Chen, A. Greer, C. S. Foote and K. N. Houk, *J. Am. Chem. Soc.*, 2003, **125**, 1319–1328.
7 T. Bekele, C. F. Christian, M. A. Lipton and D. A. Singleton, *J. Am. Chem. Soc.*, 2005, **127**, 9216–9223.
8 A. E. Litovitz, I. Keresztes and B. K. Carpenter, *J. Am. Chem. Soc.*, 2008, **130**, 12085–12094.
9 D. R. Glowacki, S. P. Marsden and M. J. Pilling, *J. Am. Chem. Soc.*, 2009, **131**, 13896–13897.
10 Z. Wang, J. S. Hirschi and D. A. Singleton, *Angew. Chem., Int. Ed.*, 2009, **48**, 9156–9159.
11 A. Patel, Z. Chen, Z. Yang, O. Gutierrez, H. Liu, K. N. Houk and D. A. Singleton, *J. Am. Chem. Soc.*, 2016, **138**, 3631–3634.
12 D. J. Pasto, K. Garves and M. P. Serve, *J. Org. Chem.*, 1967, **32**, 774–778.
13 S. R. Hare, A. Li and D. J. Tantillo, *Chem. Sci.*, 2018, **9**, 8937–8945.
14 Y. J. Hong and D. J. Tantillo, *Org. Biomol. Chem.*, 2010, **8**, 4589–4600.
15 E. L. Noey, X. Wang and K. N. Houk, *J. Org. Chem.*, 2011, **76**, 3477–3483.
16 Y. J. Hong and D. J. Tantillo, *Nat. Chem.*, 2014, **6**, 104–111.
17 S. J. Ang, W. Wang, D. Schwalbe-Koda, S. Axelrod and R. Gómez-Bombarelli, *Chem*, 2021, **7**, 738–751.
18 S. R. Hare, R. P. Pemberton and D. J. Tantillo, *J. Am. Chem. Soc.*, 2017, **139**, 7485–7493.
19 T. H. Peterson and B. K. Carpenter, *J. Am. Chem. Soc.*, 1992, **114**, 766–767.
20 Z. Yang, X. Dong, Y. Yu, P. Yu, Y. Li, C. Jamieson and K. N. Houk, *J. Am. Chem. Soc.*, 2018, **140**, 3061–3067.
21 B. Li, Y. Li, Y. Dang and K. N. Houk, *ACS Catal.*, 2021, **11**, 6816–6824.
22 H. Zhang, A. J. E. Novak, C. S. Jamieson, X.-S. Xue, S. Chen, D. Trauner and K. N. Houk, *J. Am. Chem. Soc.*, 2021, **143**, 6601–6608.
23 C. S. Jamieson, A. Sengupta and K. N. Houk, *J. Am. Chem. Soc.*, 2021, **143**, 3918–3926.
24 S. Lee and J. M. Goodman, *J. Am. Chem. Soc.*, 2020, **142**, 9210–9219.
25 S. Lee and J. M. Goodman, *Org. Biomol. Chem.*, 2021, **19**, 3940–3947.
26 M. Maldonado-Domínguez, D. Bím, R. Fučík, R. Čurík and M. Srnec, *Phys. Chem. Chem. Phys.*, 2019, **21**, 24912–24918.
27 M. Maldonado-Domínguez and M. Srnec, *J. Am. Chem. Soc.*, 2020, **142**, 3947–3958.
28 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.
29 F. Weigend and R. Alhrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297–3305.
30 F. Weigend, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1057–1065.
31 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
32 M. Cossi, N. Rega, G. Scalmani and V. Barone, *J. Comput. Chem.*, 2003, **24**, 669–681.
33 J. D. Chai and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2008, **10**, 6615–6620.
34 C. Adamo and V. Barone, *J. Chem. Phys.*, 1998, **108**, 664–675.
35 M. M. Francl, W. J. Pietro and W. J. Hehre, *J. Chem. Phys.*, 1982, **77**, 3654–3665.
36 M. Linder and T. Brinck, *Phys. Chem. Chem. Phys.*, 2013, **15**, 5108–5114.
37 B. R. Ussing, C. Hang and D. A. Singleton, *J. Am. Chem. Soc.*, 2006, **128**, 7594–7607.
38 H. B. Schlegel, J. M. Millam, S. S. Iyengar, G. A. Voth, G. E. Scuseria, A. D. Daniels and M. J. Frisch, *J. Chem. Phys.*, 2001, **114**, 9758–9763.

39 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. V. 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. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. 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 16*, Revision B.01, Gaussian, Inc., Wallingford CT, 2016.

40 H. Kikouchi and D. A. Singleton, *Nat. Chem.*, 2018, **10**, 237–241.

41 L. M. M. Quijano and D. A. Singleton, *J. Am. Chem. Soc.*, 2011, **133**, 13824–13827.

42 C. Doubleday and W. L. Hase, *J. Phys. Chem. A*, 1998, **102**, 3648–3658.

43 S. Hanessian and P. Compain, *Tetrahedron*, 2002, **58**, 6521–6529.

44 S. E. Denmark, B. S. Kesler and Y. C. Moon, *J. Org. Chem.*, 1992, **57**, 4912–4924.

45 N. Çelebi-Ölçüm, D. H. Ess, V. Aviyente and K. N. Houk, *J. Am. Chem. Soc.*, 2007, **129**, 4528–4529.

46 J. Limanto, K. S. Khuong and K. N. Houk, *J. Am. Chem. Soc.*, 2003, **125**, 16310–16321.

47 M. Harmata and M. G. Gomes, *Eur. J. Org. Chem.*, 2006, 2273–2277.

48 J. B. Thomas, J. R. Waas, M. Harmata and D. A. Singleton, *J. Am. Chem. Soc.*, 2008, **130**, 14544–14555.

49 Z. Wang, J. S. Hirschi and D. A. Singleton, *Angew. Chem., Int. Ed.*, 2009, **48**, 9156–9159.

50 M. Schmittel, M. Keller, S. Kiau and M. Strittmatter, *Chem.–Eur. J.*, 1997, **3**, 807–816.

51 S. Yamabe, T. Dai, T. Minato, T. Machiguchi and T. Hasegawa, *J. Am. Chem. Soc.*, 1996, **118**, 6518–6519.

52 P. Yu, T. Q. Chen, Z. Yang, C. Q. He, A. Patel, Y. H. Lam, C. Y. Liu and K. N. Houk, *J. Am. Chem. Soc.*, 2017, **139**, 8251–8258.

53 S. Chen, P. Yu and K. N. Houk, *J. Am. Chem. Soc.*, 2018, **140**, 18124–18131.

54 M. Ohashi, F. Liu, Y. Hai, M. Chen, M. C. Tang, Z. Yang, M. Sato, K. Watanabe, K. N. Houk and Y. Tang, *Nature*, 2017, **549**, 502–506.

55 R. Villar López, O. N. Faza and C. Silva López, *J. Org. Chem.*, 2017, **82**, 4758–4765.

56 L. Ye, Y. Wang, D. H. Aue and L. Zhang, *J. Am. Chem. Soc.*, 2012, **134**, 31–34.

57 R. B. Campos and D. J. Tantillo, *Chem*, 2019, **5**, 227–236.

58 T. Katori, S. Itoh, M. Sato and H. Yamataka, *J. Am. Chem. Soc.*, 2010, **132**, 3413–3422.

59 N. Mandal and A. Datta, *J. Phys. Chem. B*, 2018, **122**, 1239–1244.

60 D. T. Major and M. Weitman, *J. Am. Chem. Soc.*, 2012, **134**, 19454–19462.

61 X. S. Bogle and D. A. Singleton, *Org. Lett.*, 2012, **14**, 2528–2531.

62 B. K. Carpenter, *Angew. Chem., Int. Ed.*, 1998, **37**, 3340–3350.

63 M. Khrapunovich, E. Zelenova, L. Seu, A. N. Sabo, A. Flatherty and D. C. Merrer, *J. Org. Chem.*, 2007, **72**, 7574–7580.

64 S. Itoh, N. Yoshimura, M. Sato and H. Yamataka, *J. Org. Chem.*, 2011, **76**, 8294–8299.

65 C. S. Jamieson, A. Sengupta and K. N. Houk, *J. Am. Chem. Soc.*, 2021, **143**, 3918–3926.

66 D. C. Merrer and P. R. Rablen, *J. Org. Chem.*, 2005, **70**, 1630–1635.

67 M. Khrapunovich, E. Zelenova, L. Seu, A. N. Sabo, A. Flaherty and D. C. Merrer, *J. Org. Chem.*, 2007, **72**, 7574–7580.

68 Y. J. Hong and D. J. Tantillo, *Nat. Chem.*, 2014, **6**, 104–111.

69 E. L. Noey, Z. Yang, Y. Li, H. Yu, R. N. Richey, J. M. Merritt, D. P. Kjell and K. N. Houk, *J. Org. Chem.*, 2017, **82**, 5904–5909.

70 H. J. Kim, M. W. Ruszczycky, S. H. Choi, Y. N. Liu and H. W. Liu, *Nature*, 2011, **473**, 109–112.

71 X. S. Xue, C. S. Jamieson, M. García-Borrás, X. Dong, Z. Yang and K. N. Houk, *J. Am. Chem. Soc.*, 2019, **141**, 1217–1221.

72 C. Y. Liu and S. T. Ding, *J. Org. Chem.*, 1992, **57**, 4539–4544.

73 C. Y. Liu, S. T. Ding, S. Y. Chen, C. Y. You and H. Y. Shie, *J. Org. Chem.*, 1993, **58**, 1628–1630.

74 S. Itoh, N. Yoshimura, M. Sato and H. Yamataka, *J. Org. Chem.*, 2011, **76**, 8294–8299.

75 B. C. Sweeny, H. Pan, A. Kassem, J. C. Sawyer, S. G. Ard, N. S. Shuman, A. A. Viggiano, S. Brickel, O. T. Unke, M. Upadhyay and M. Meuwly, *Phys. Chem. Chem. Phys.*, 2020, **22**, 8913–8923.

JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JACS

Article

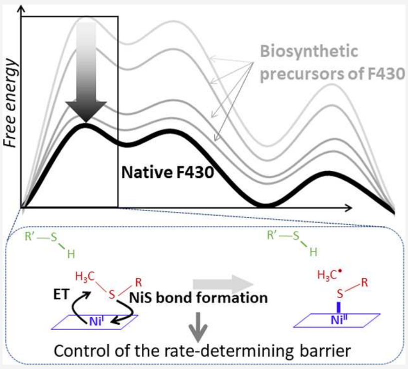
Reactivity Factors in Catalytic Methanogenesis and Their Tuning upon Coenzyme F430 Biosynthesis

Priyam Bharadwaz, Mauricio Maldonado-Domínguez, Jakub Chalupský, and Martin Srnec*

Cite This: *J. Am. Chem. Soc.* 2023, 145, 9039−9051

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: Methyl-coenzyme M reductase, responsible for the biological production of methane by catalyzing the reaction between coenzymes B (CoBS-H) and M (H_3C-SCoM), hosts in its core an F430 cofactor with the low-valent Ni^I ion. The critical methanogenic step involves F430-assisted reductive cleavage of the H_3C−S bond in coenzyme M, yielding the transient CH_3 radical capable of hydrogen atom abstraction from the S−H bond in coenzyme B. Here, we computationally explored whether and why F430 is unique for methanogenesis in comparison to four identified precursors formed consecutively during its biosynthesis. Indeed, all precursors are less proficient than the native F430, and catalytic competence improves at each biosynthetic step toward F430. Against the expectation that F430 is tuned to be the strongest possible reductant to expedite the rate-determining reductive cleavage of H_3C−S by Ni^I, we discovered the opposite. The unfavorable increase in reduction potential along the F430 biosynthetic pathway is outweighed by strengthening of the Ni−S bond formed upon reductive cleavage of the H_3C−S bond. We found that F430 is the weakest electron donor, compared to its precursors, giving rise to the most covalent Ni−S bond, which stabilizes the transition state and hence reduces the rate-determining barrier. In addition, the transition state displays high pro-reactive motion of the transient CH_3 fragment toward the H−S bond, superior to its biosynthetic ancestors and likely preventing the formation of a deleterious radical intermediate. Thus, we show a plausible view of how the evolutionary driving force shaped the biocatalytic proficiency of F430 toward CH_4 formation.

■ INTRODUCTION

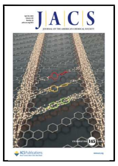
Methane is an important source of energy due to its heat of combustion, the highest among carbon-based fuels.[1,2] Globally, 90−95% of methane is formed by methanogenic archaea in anoxic environments from CO_2 and H_2, acetate, methylamines, and methanol.[3,4] In such way, nearly 1 billion tons of methane are produced every year.[5]

Methyl-coenzyme M reductase (MCR) is essential for catalysis of the terminal and rate-determining step in biological methanogenesis. Each of its two identical active sites hosts a nickel-containing cofactor F430, structurally similar to porphyrin, chlorophyll, and vitamin B12.[6−9] This cofactor promotes the reaction between two co-substrates—methyl donor [methyl-coenzyme M: H_3C−SCoM] and hydrogen donor [coenzyme B: CoB-SH], yielding methane and heterodisulfide CoMS−SCoB (Scheme 1).[10,11] Note that MCR is also capable of catalyzing the reverse process as reported for anaerobic methanotrophic sulfate, nitrate, or Fe^{III}-reducing bacteria.[8,10−12]

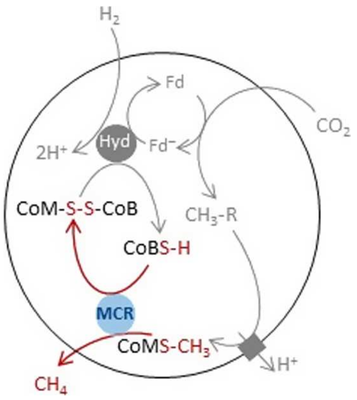
To date, a number of studies have been carried out toward understanding the mechanistic details for methane production by MCR.[14−22] The current consensus on the reaction mechanism of MCR, which was first computationally proposed by Siegbahn and co-workers[21,23] and later supported experimentally by Ragsdale and co-workers,[24] is summarized in Figure 1A. Importantly, the main part of the catalytic cycle consists of three subsequent steps: (i) homolytic cleavage of the H_3C−S_{CoM} bond with the concomitant oxidation of the nickel center from +I to +II and the formation of a bond between Ni^{II} and SCoM, followed by (ii) attack of the transient methyl radical to the H−S_{CoB} bond, which releases CH_4 and the substrate $^{•}S_{CoB}$ radical, the latter of which subsequently (iii) recombines with the Ni^{II}-bound SCoM to form a disulfide product. These three steps are labeled in Figure 1A as **step 1**, **step 2**, and **step 3**, and this labeling will be used further in the text.

Recently, Warren and co-workers elucidated the biosynthetic pathway of F430,[25] where the late stage comprises four enzymatically controlled steps in which the porphyrin-like skeleton is gradually modified including chelation, amidation, reduction by six electrons with addition of seven protons, lactamization, and closure of a propionate side chain coupled to water extrusion. All experimentally determined Ni-

Received: January 13, 2023
Published: April 12, 2023



Scheme 1. MCR Catalyzes the Reaction between Coenzymes M and B[a]



[a]The reaction is highlighted in red and placed in the context of a simplified respiratory chain in archea, in which CO_2 is reduced by H_2 to propel the pumping of protons through the membrane and drive the synthesis of energy-rich ATP molecules. The acronyms Hyd, Fd, and Fd^- stand for hydrogenase and oxidized and reduced ferredoxin, respectively. The scheme is adapted from ref 13.

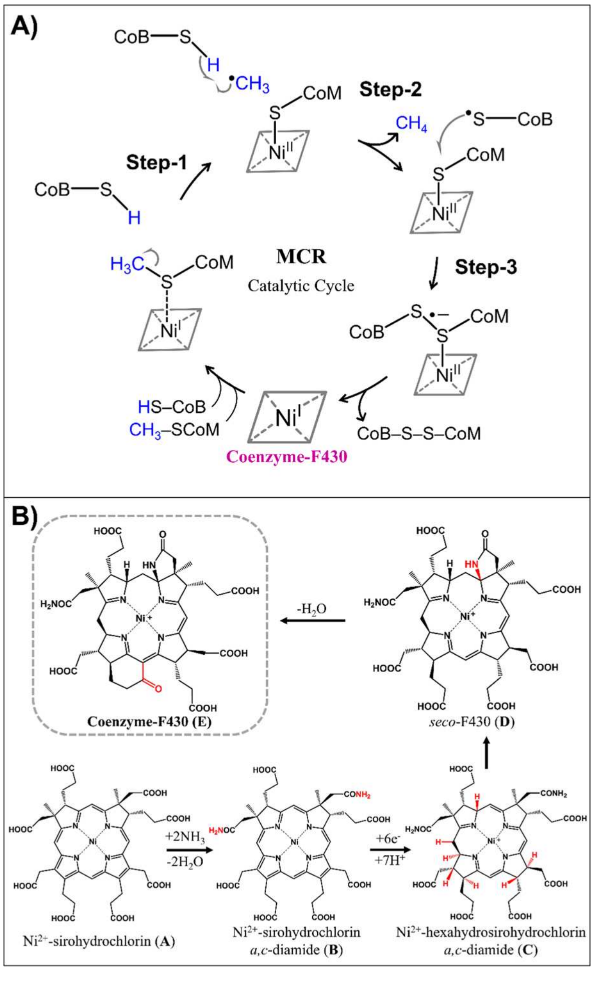


Figure 1. (A) Canonical catalytic cycle of the MCR (refs 21 and 24). (B) Biosynthesis of the native cofactor F430 comprises four consecutive Ni-containing intermediates, as reported in ref 25. Labels **A**, **B**, **C**, **D**, and **E** are used for the cluster models from Figure 2 anchoring these biosynthetic precursors.

anchoring biosynthetic precursors of F430 are shown in Figure 1B.

Here, driven by curiosity about whether and how the evolutionary force forming such a complex and energy-demanding pathway lies in optimizing reactivity toward the CH_4 formation (and its oxidation in the reverse process), we computationally investigated energetics and dissected key reactivity factors contributing to the consensual mechanism of methane production by the native F430 cofactor and its four biosynthetic precursors from Figure 1B.

■ METHODOLOGY AND COMPUTATIONAL DETAILS

Structural Models. The cluster model of the active site of MCR, constructed from the crystal structure with PDB code 1HBN,[17] consists of 190 atoms including (i) the Ni-containing F430 coenzyme; (ii) the two co-substrates CH_3S-CoM and CoB-SH; and (iii) three residues, Tyr333, Tyr367, and Gln147 (Figure 2). If not stated otherwise, the carboxylate groups of the coenzyme are protonated, giving the MCR model the total charge of −1. In analogy, the cluster models of the fictitious MCR-like active sites anchoring various F430-related complexes (non-native coenzymes) were derived from the MCR native model. Such *non-native* coenzymes were identified as intermediates of the biosynthetic pathway of the native F430 (Figure 1B).[25] The corresponding models are further labeled as **A**, **B**, **C**, **D**, and **E**, where the latest stands for the F430-containing MCR model, while **A**, **B**, **C**, and **D** are the active-site models anchoring biosynthetic precursors of F430: Ni^{II}-sirohydrochlorin (**A**), Ni^{II}-sirohydrochlorin *a,c*-diamide (**B**), Ni^{II}-hexahydrosirohydrochlorin *a,c*-diamide (**C**), and *seco*-F430 (**D**). **A**/**B** and **C**/**D** models have the charge of −2 and −1, respectively. All the cluster models and their overlay are shown in Figure S1A,B. Comparison of the referential PDB geometry

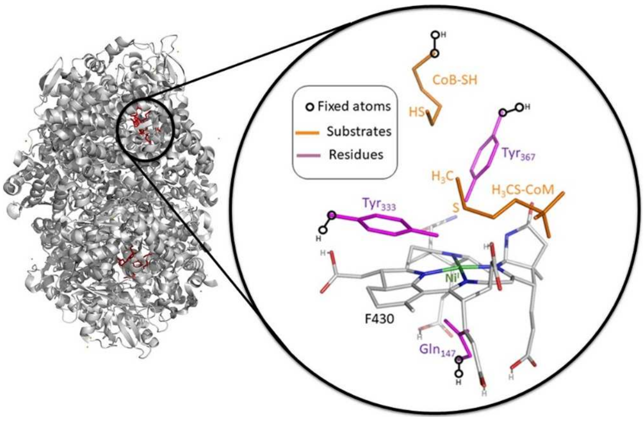


Figure 2. Cluster model of the MCR active site constructed from the X-ray crystal structure 1HBN. Hydrogens are omitted for clarity. Carboxylates in F430 are protonated (as schematically depicted), and the residues and the CoB-SH are truncated and capped by H atoms. In the preparatory stage, the capped C−H bonds were aligned and optimized along the original (crystallographic) C−C axes, while all non-hydrogen atoms were kept fixed. In all subsequent optimizations, only these capping H atoms along with the adjacent C atoms were kept fixed as indicated in the figure. The depicted model with the native F430 is labeled in the text as model **E**. Other models with the non-native cofactors corresponding to biosynthetic precursors from Figure 1B are labeled **A**, **B**, **C**, and **D**. All the cluster models (**A**−**E**) and their overlay are shown in Figure S1.

and its optimized form is presented in Figure S1C, supporting the truncation scheme employed throughout this work.

Density Functional Theory Calculations. All of the structures were optimized at the B3LYP(D3)/BS1/CPCM($\varepsilon_r = 4$) level of theory, i.e., we employed the B3LYP[26] functional with the empirical correction to the dispersion effects using the original Grimme's D3 damping function[27] and in combination with the hybrid basis set BS1, which includes the LANL2TZ ECP basis set for Ni,[28] 6-311G* for key atoms involved in reactivity (the SCH_3 and SH groups of the H_3CS-CoM and truncated CoB-SH substrates, respectively), and 6-31G* for the rest,[29] while the conductor-like polarizable continuum model (CPCM)[30] with a dielectric constant $\varepsilon_r = 4$ (explicit parameters for the CPCM solvent model are provided in Table S1) was used to represent the protein environment. Gibbs free energies were evaluated:

$$G = E_{\mathrm{el}} + [E_{\mathrm{ZPVE}} + pV - RT\ln Q] \tag{1}$$

where E_{el} are the B3LYP(D3)/BS2/CPCM single-point energies on top of the optimized geometries, calculated using the larger basis set BS2: LANL2TZ+ ECP for Ni and 6-311++G** for the rest; the $[E_{\mathrm{ZPVE}} + pV - RT \ln Q]$ term includes the thermal enthalpic and entropic contributions of the solute energy with E_{ZPVE} and Q corresponding to zero-point vibrational energy and the molecular partition function, respectively, obtained from B3LYP(D3)/BS1/CPCM. Frequency calculations were performed with the rigid rotor/harmonic oscillator approximation (for $p = 1$ bar, $T = 298$ K). In all cases, the vibrational frequencies associated to the eight frozen atoms (Figure 2) were projected out from the hessian, yielding the consistent number of degrees of freedom for minima ($3n - 24$) and for transition states $3n - 25$, where n is the total number of atoms in the cluster model. Only these degrees of freedom were included when calculating entropic and enthalpic contributions. All calculations were carried out using Gaussian 16.[31]

Atomic charges and delocalization indices were obtained by integration of the DFT-optimized electron density using the atoms-in-molecules (AIM) theory as implemented in the AIMAll program.[32] Atomic basins were integrated using the Proaim method,[32] where a "Fine" interatomic surface mesh, an outer angular integration quadrature of 7200 grid points, and a maximum integration radius of 13.0 Bohr were used for all atoms. Electrostatic potential (ESP) maps were obtained by probing the 0.001 isodensity surface with a point charge. All AIM atomic charges are included in the Supporting Information.

Redox Potentials and Validation of the Methodology. The critical step in the MCR catalytic cycle includes the reductive cleavage of the $H_3C{-}SCoM$ bond with the concomitant oxidation of Ni^{I} to Ni^{II} (Figure 1). Thus, the reduction potential of the metal center in the F430 cofactor must be key for successful C−S cleavage, and its accurate evaluation is a prerequisite for a reliable description of reaction energetics of such a critical catalytic step. The reduction (redox) potentials ($E°$, in V) are calculated as follows:

$$E° = \frac{G_{\mathrm{ox}} - G_{\mathrm{red}} - FE^{\circ}_{\mathrm{abs}}\,(\mathrm{reference})}{F} \tag{2}$$

where G_{ox} and G_{red} are the Gibbs free energies of the oxidized and reduced forms of the structural models containing Ni^{II} and Ni^{I}, respectively, and E°_{abs} (reference) is the absolute potential of a reference electrode. F is the Faraday constant. If not stated otherwise, the absolute potential of a normal hydrogen electrode (NHE) in water with the value of $E^{\circ}_{\mathrm{NHE}} = 4.28$ eV was used.[33]

To validate the computational methodology for calculation of Gibbs free energies from eq 2, we calculated reduction potentials of three structurally well-defined synthetic complexes, for which the experimental data are available.[34−37] The calculated potentials presented in Figure S2 are in all cases in good agreement with the experiments (deviation of calculated reduction potentials from experimental data are provided in Figure S2). Note that the redox-active molecular orbital in $Ni^{I/II}$ is the in-plane $d_{x^2-y^2}$-based orbital. Orbital rendering and visualization were achieved with the aid of Charmol.[38]

Multiconfigurational Calculations. The presented results were calculated using the ANO-RCC-DZP[39] basis set, where Coulomb and exchange two-electron integrals were approximated by the resolution of identity (RIJK) with our own auxiliary basis set generated for ANO-RCC-DZP by a procedure similar to ref 40. Scalar relativistic effects were accounted for via the second-order Douglas−Kroll−Hess (DKH2) approximation.[41]

The complete active space self-consistent field (CASSCF)[42] calculations were based on an active space composed of two Ni 3d orbitals (d_{z^2} and $d_{x^2-y^2}$), the S $3p_z$ orbital of the Ni−S bond, the N $2p_x$ and $2p_y$ orbitals of the Ni−N bonds, and the methyl singly occupied orbital. State-specific CASSCF calculations were performed for the doublet ground state on top of DFT-optimized geometries. All calculations were carried out with the in-house ORZ program developed in the group of Yanai et al.[43]

Kinetic Energy Distribution (KED) Analysis. Kinetic energy of the ith atom $\langle T_i \rangle$ in the molecular systems of n atoms is derived from kinetic energies of the $3n$ normal modes $\langle T_\alpha \rangle$:

$$\langle T_i \rangle = \sum_{\alpha}^{3n} \mathrm{KED}_{i\alpha} \langle T_\alpha \rangle \tag{3}$$

where $\mathrm{KED}_{i\alpha}$ is the fraction of kinetic energy associated with the ith atom in the mode α with real (or imaginary) frequency. Note the number of $3n$ modes also includes translational and rotational degrees of freedom. KED values are readily obtainable from cartesian atomic displacements; in this study, these are obtained from the B3LYP(D3)/BS1/CPCM frequency calculations as

$$\mathrm{KED}_{i\alpha} = \left(\frac{m_i r_i^2}{\sum_k^n m_k r_k^2} \right)_{\alpha} \tag{4}$$

ranging between 0 and 1; r_i and m_i are the displacement and the mass of the ith atom, respectively (k runs over all n atoms). The complete theory is provided in ref 44, while some of its applications are refs 45 and 46. For purposes of the presented study, only KED within the reactive mode with one imaginary frequency was analyzed and correlated with other properties. Mapping of KED values on molecular structures was carried out with the aid of Charmol.[38]

RESULTS AND DISCUSSION

Redox Potentials: Models with the Native F430 vs Biosynthetic Precursors of F430. The Ni center present in F430 switches between the +I and +II oxidation states during the catalytic process (Figure 1A). This raises the question of whether the redox activity of F430, controlled by its reduction potential $E°$, is the key variable that makes F430 a unique catalyst for methane production compared to all four F430 biosynthetic precursors presented in Figure 1B. Naturally, such hypothesis is only viable if the chemical modifications made on the macrocyclic ring during the biosynthetic pathway of F430 translate into substantial changes of $E°$.

Table 1 summarizes the results for enzymatic active-site models **A**−**E**, which host the F430 cofactor (model **E**) and its biosynthetic precursors (models **A**, **B**, **C**, and **D**), respectively. As seen, $E°$ varies significantly across the series. The active-site models containing early-stage biosynthetic precursors are very strong reductants, while the native active site with F430 is the strongest oxidant across the series (the span of $E°$ is almost ∼1.5 V). We also note in passing that the oxidized state Ni^{II} in **A**−**D** resides in the triplet state with two unpaired electrons in $d_{x^2-y^2}$ and d_{z^2}, while model **E** adopts the singlet state (where the $d_{x^2-y^2}$ orbital remains unoccupied) with the $S_{ox} = 1$ state lying only ∼1 kcal mol^{-1} above the ground state. Interestingly, the splitting of the singlet/triplet spin states of the Ni^{II} center follows the same trend as the reduction potential (cf. Table 1), both indicating that the stability of the triplet state of Ni^{II} decreases in the order **B** > **A** > **C** > **D** > **E**. The reduced Ni^{I} form displayed a doublet ground spin state in all cases, with an unpaired electron in the $d_{x^2-y^2}$ orbital.

Table 1. Redox Potentials for the Couple Ni^{II}/Ni^{I} in All Cluster Models A−E from Figure 2 (and Figure S1) Calculated Using eq 2 and Referenced to the NHE[a]

cluster model	S_{ox}/S_{red}	$E°_{calc}$ (V)	oxidized form: singlet/triplet gap (kcal mol^{-1})	reduced form: doublet/quartet gap (kcal mol^{-1})
A	$1/\frac{1}{2}$	−1.77	5.8	12.8
B	$1/\frac{1}{2}$	−1.95	6.7	12.9
C	$1/\frac{1}{2}$	−1.11	4.9	17.2
D	$1/\frac{1}{2}$	−0.88	1.8	26.9
E	$0/\frac{1}{2}$	−0.53	0.9	20.0

[a] The ground spin states for both oxidized and reduced forms along with the spin-state energetics are also shown. Note that the oxidized form of the Ni center in the triplet state in **A**−**E** is axially coordinated by the glutamine residue, Gln147 (the reduced form and the oxidized form in the singlet state have no axial ligation to Ni).

To sum this section, we conclude that biosynthetic modifications of the macrocyclic ring of the Ni complex have a sizable impact on the redox activity of Ni, and hence, its contribution to reactivity will be addressed in the following sections.

Reaction Energetics: Models with the Native F430 vs Biosynthetic Precursors of F430. As a referential system, we first calculated the energetics of the uncatalyzed reaction between the CH_3S-CoM and CoB-SH substrates, i.e., in the absence of a Ni-macrocycle complex, while imposing a homogeneous water-like surrounding through the continuum solvation model with $\varepsilon_r = 80.0$. The corresponding reaction mechanism and its energetics are described in Figure 3A and in detail in Figure S3, where the intrinsic reaction coordinate (IRC) with the key structures is shown. From this IRC, the uncatalyzed process begins with the cleavage of the S−CH_3 bond in coenzyme M, which further initiates the cleavage of the S−H bond in coenzyme B with the concomitant formation of CH_4. All these three events occur in one single-step process with a free-energy barrier ($\Delta G^{\neq}$) of ∼89 kcal mol^{-1}, with S−S bond formation between coenzymes M and B occurring in a second step characterized by essentially no barrier. The free energy of reaction is $\Delta G_0 = -4.6$ kcal mol^{-1}. These results clearly show that such a direct reaction is unfeasible. *Why and how does this change when MCR participates in the reaction?* Here, we take advantage of the consensual mechanism for the catalytic CH_4 formation from Figure 1A and evaluate its energetics for five active-site models **A**−**E**, one with the native F430 and four with its biosynthetic precursors from Figure 1B. In all cases, the reaction mechanism consists of three consecutive steps, each associated with its own barrier (Figure 3B).

The first barrier $\Delta G_1^{\neq}$ (associated with **step 1** from Figure 4, vide infra) involves formation of the transient methyl radical and must be critical for catalysis due to the direct participation of the redox-active Ni center of the cofactor via Ni^{I} oxidation and Ni^{II}-thiolate bond formation. Such chemical events strongly suggest that $\Delta G_1^{\neq}$ should be sensitive to modifications of the macrocyclic ring. Indeed, the calculations show that $\Delta G_1^{\neq}$ varies significantly in going from model **A** to **E** with the lowest barrier for **E** containing the native F430 cofactor (∼20 kcal mol^{-1}; structures of all minima and transition states appearing along the pathway in Figure 3B are depicted in Figure S4). In addition, $\Delta G_1^{\neq}$ appears to decrease almost monotonously in the transition from **A** to **E** (**B** > **A** > **C** > **D** > **E**), suggesting that all chemical steps along biosynthesis of F430 are evolutionary designed to make the CH_4 formation feasible. Importantly, the change of $\Delta G_1^{\neq}$ correlates with the change of free energy of reaction of **step 1** ($\Delta G_{0,1}$) in the ratio ∼1:1 (Figure S5). Such one-to-one correlation is the consequence of geometric- and electronic-structure similarity of the transition state TS_1 with the first methyl radical-containing intermediate Int_1 (i.e., TS_1 is late along the reaction coordinate), which is further reflected by a small energetic difference between TS_1 and Int_1 ($G_{TS1} - G_{Int1} < 5$ kcal mol^{-1}; reaction **step 1** is strongly endergonic in all **A**−**E**; Figure 3B). This implies that $\Delta G_1^{\neq}$ and its change across **A**−**E** must be controlled thermodynamically. In all cases, along the reaction coordinate of **step 1**, the oxidized Ni adopts the local triplet state ($S_{Ni^{II}} = 1$) that couples antiferromagnetically to the transient CH_3 radical (cf., spin densities in Table S3).

We note that the carboxylate groups located on the periphery of the macrocyclic cofactors in the set **A**−**E** are protonated in our study (Figure 2), while these groups are likely charged in the native MCR enzyme. Thus, we evaluated the electrostatic effect of these charged groups on $\Delta G_1^{\neq}$ and $\Delta G_{0,1}$ and their evolution across **A**−**E** and we found that both $\Delta G_1^{\neq}$ and $\Delta G_{0,1}$ are systematically downshifted by ∼5 kcal mol^{-1} (Figure S6); $\Delta G_1^{\neq}$ for the MCR model **E** reaches 14 kcal mol^{-1}, which is close to the experimental value of 13.2 kcal mol^{-1}.[24] Thus, it seems that the charged carboxylate groups on the corphinoid periphery in F430 play an important role in reducing the rate-determining barrier and possibly in anchoring the macrocycle in MCR. Nevertheless, our results suggest that they do not contribute to differences in reactivity between F430 and its precursors. Also, we note that the distortion of tetrapyrrole macrocycles is found in nature, as is the case of *cytochrome c*, to maximize the efficiency of this kind of cofactors through the fine-tuning of their electronic properties by protein-imposed geometry constraints.[47] In the case of cofactor F430, the native geometry as obtained by XRD in ref 17 deviates minimally (RMSD = 1.0 Å) from the DFT-optimized cluster model **E**, suggesting that the maturation

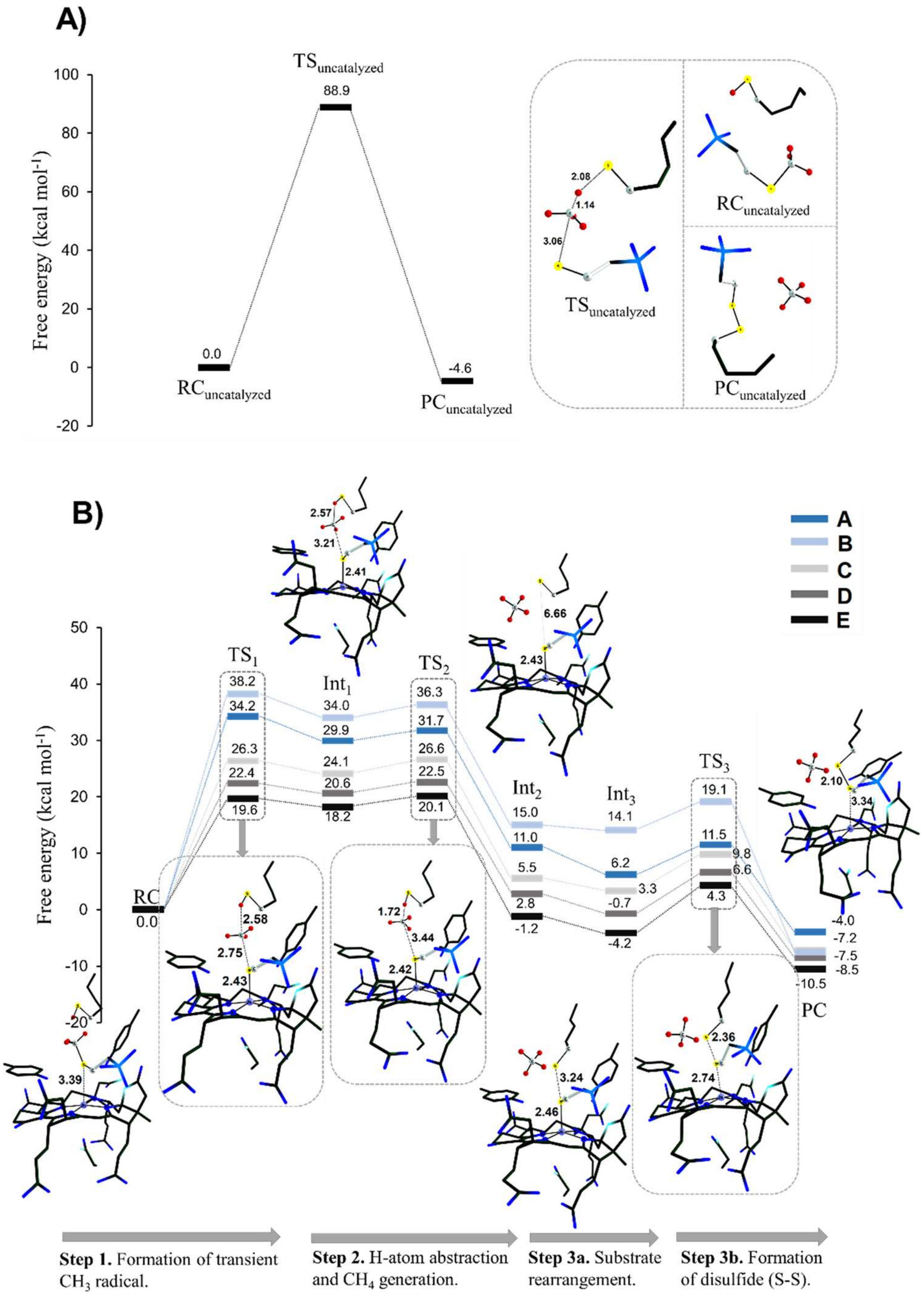


Figure 3. (A) Free-energy profile of uncatalyzed one-step reaction summarized in Scheme 1 along with the structures and critical distances (in Å). (B) Free-energy profile for catalyzed methanogenesis within enzyme-like models (**E** representing MCR with the native F430; **A**–**D** hosting biosynthetic precursors of F430). For the sake of clarity, all the critical structures along the catalytic reaction coordinate are displayed for model **E**, and only key hydrogen atoms are visualized. The potential energies and enthalpies of all key geometries are given in Table S2.

process suffices to reach catalytic efficiency and cofactor distortion was not necessary during the evolutionary process, leading to F430.

In the case of **step 2** from Figure 3B, the transient methyl radical abstracts a hydrogen atom from the coenzyme HS-CoB. The energetics of this step is practically independent of the chemical character of the cofactor: the barrier $\Delta G_2^{\neq}$ in all cases is only ~2 kcal mol^{-1} and the reaction free energy associated with this step $\Delta G_{0,2}$ is exergonic (≈−19 kcal mol^{-1}). Regarding the possible formation and fate of a CH_3 radical species, we carried out an analysis of KED (ranging from 0 to 1) of the reactive mode of transition state in **step 1** (TS_1). The analysis reveals that most of the kinetic energy is concentrated in the motion of the nascent methyl radical (cf. KED_{CH_3} in all **A**–**E** cases exceeds 0.78; Figure S7), with the highest KED_{CH_3} = 0.88 calculated for the native **E**. This result suggests that the shallow Int_1 state can be bypassed by a ballistic trajectory of CH_3 toward the key H–S bond, leading to direct methane production after passing TS_1. Such dynamically controlled reactivity goes beyond the traditional transition state theory description, and the KED analysis was proposed by us as an applicable tool when nonequilibrium reactions are suspected,[43]

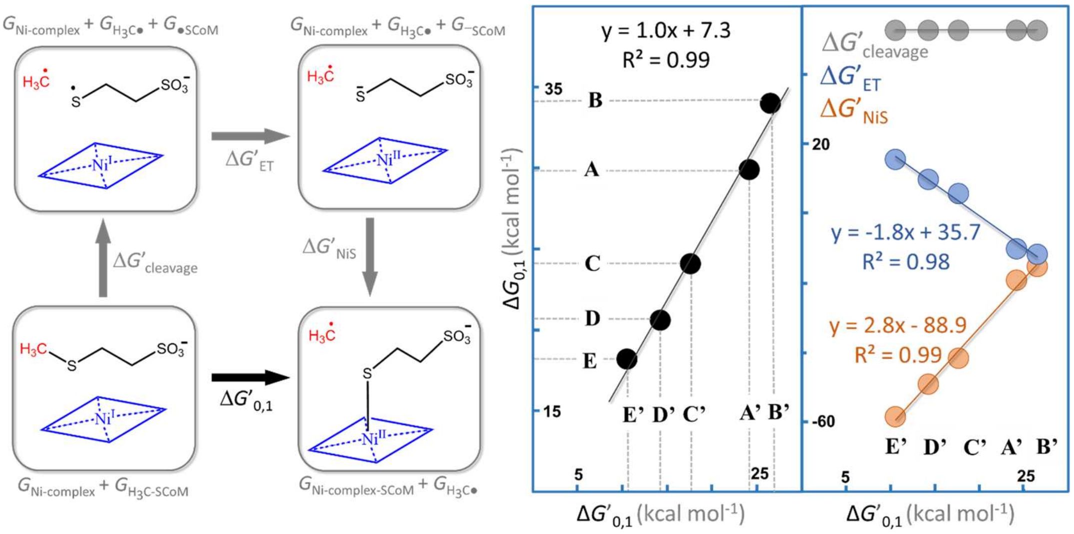


Figure 4. Thermodynamic cycle of the first catalytic step involving the formation of the transient methyl radical coupled with one-electron oxidation of the Ni^I center and Ni-thiolate bond formation. The whole process is dissected into three steps: (i) the homolytic cleavage of the S–CH$_3$ bond ($\Delta G'_{cleavage}$), (ii) electron transfer from the Ni^I center to the S-radical ($\Delta G'_{ET}$); and (iii) the Ni–S bond formation ($\Delta G'_{NiS}$). Note that $\Delta G'$ values were calculated from Gibbs free energies of individual CPCM(ε_r = 4.0)-solvated species, as indicated in the figure and using eq 1; the sum of three sequential ΔG's steps equals to $\Delta G'_{0,1}$, which is analogous to the free energy of reaction $\Delta G_{0,1}$ from Figure 3B. The labels **A′**, **B′**, **C′**, **D′**, and **E′** are used in parallel to labels for the full models **A**–**E**; the "prime" symbol refers to the fact that the model consists of individual (infinitely separated) moieties such as the Ni$^{I/II}$-complex, H$_3$C–SCoM, $^{-/\cdot}$SCoM, CH$_3^\cdot$, and the Ni^{II}–SCoM complex. Of note, the individual Ni complexes from the upper-right corner of the cycle are calculated in the singlet ground spin state (details in Tables S4A and S5), while the Ni^{II}–SCoM complexes from the lower-right corner of the cycle are obtained in the triplet ground state, which is in line with the local triplet state of Ni in the Int$_1$ structures of models **A**–**E** from Figure 3. Importantly, if the triplet state for the individual Ni^{II}-complexes from the upper-right corner is considered, the correlation patterns seen in the rightmost graph in this figure are preserved (cf. Figure S8 and Table S4B): the change in $\Delta G'_{NiS}$ dominates the change in $\Delta G'_{0,1}$ and outcompetes the unfavorable change in $\Delta G'_{ET}$ in going from **B′**/**A′** to **E′**.

as is potentially the case of the CH$_3$S-CoM cleavage/CoB-SH oxidation sequence. Very high pro-reactive motion of the CH$_3$ fragment toward the H–S bond would prevent accumulation of the radical intermediate, potentially dangerous for the protein structure and its function. In this context, it is also worth stressing that Siegbahn, using a similar computational protocol, reported a single barrier for methane production with TS$_1$ followed by a barrierless **step 2**.[21] In either mechanism, the methyl radical would be very transient.

In **step 3**, the CoBS$^\cdot$ radical combines with the Ni^{II}-bound thiolate of CoM to generate a disulfide anion radical (Figure 3B). Notably, we recognize two distinct stages of **step 3**: (i) substrate translocation to approach SCoM (**step 3a**) and (ii) the S–S bond formation (**step 3b**). Concerning **step 3a**, the translocation of the CoBS-radical closer to SCoM in going from Int$_2$ to Int$_3$ is exergonic in all models **A**–**E**. Formation of the disulfidic bond in **step 3b** requires overcoming a barrier, which is found to be the highest ($\Delta G_3^{\neq}$ = 8.5 kcal mol^{-1}) for model **E**, with the native cofactor F430, and the lowest for **B** (5.0 kcal mol^{-1}). While $\Delta G_3^{\neq}$ is quite variable in the series **A**–**E**, it stays lower in energy as compared to the rate-determining $\Delta G_1^{\neq}$.

Overall, relative to the uncatalyzed reaction (Figure 3A), the participation of any of the studied Ni macrocycles enormously reduces the highest barrier of the reaction highlighted in red in Scheme 1. However, the early-stage biosynthetic precursors anchored in models **A** and **B** still possess very high barriers (~38 and ~34 kcal mol^{-1} for **B** and **A**, respectively) to catalyze the reaction, while two late-stage precursors in models **C** and **D** become better but still inefficient catalysts, rendering their respective rate-determining barriers roughly 3 and 7 kcal mol^{-1} higher than the barrier displayed by model **E** with the native F430.

Reactivity Factor Analyses of the Key Reaction Step 1. To elucidate the factors contributing to energy differences of key catalytic **step 1** in the series **A**–**E**, we decomposed this step into three elementary events and constructed the thermodynamic cycle shown in Figure 4. The first elementary step is homolytic cleavage of the substrate S–CH$_3$ bond and is energetically independent of the chemical character of the Ni complex ($\Delta G'_{cleavage}$). In contrast, the second and third events are dependent on the properties of the Ni complex because they correspond to Ni^I-to-thiyl electron transfer (ET) and Ni^{II}-thiolate bond formation ($\Delta G'_{ET}$ and $\Delta G'_{NiS}$), respectively. Of note, the $\Delta G'_{cleavage}$, $\Delta G'_{ET}$, and $\Delta G'_{NiS}$ terms in Figure 4 are obtained using free energies of individual moieties (at infinite separation) such as the *isolated* Ni complex (in the absence of amino residues and coenzymes), *isolated* coenzyme M, etc., as indicated in the figure. The thermodynamic dissection of **step 1** into three terms is sensible, as evidenced by the correlation between the free energy of reaction $\Delta G_{0,1}$ from Figure 3B and the sum of the three thermodynamic terms $\Delta G'_{0,1}$ (= $\Delta G'_{cleavage}$ + $\Delta G'_{ET}$ + $\Delta G'_{NiS}$), where the correlation slope reaches nearly the ideal value of 1 (Figure 4, *left graph*). The detailed inspection of the two decisive individual contributions $\Delta G'_{ET}$ and $\Delta G'_{NiS}$ is provided in the following sections.

Ni$^{I/II}$ Redox Potential and Its Correlation with Reactivity in Step 1. The change of $\Delta G'_{ET}$ is given by the difference in reduction potentials of Ni complexes (cf. Figure 4 and Tables S4 and S5). Importantly, $E°$ increases as F430 biosynthetic precursors gradually become more similar to F430, suggesting an appealing possibility that the reduction potential of the Ni center is responsible for the variability of

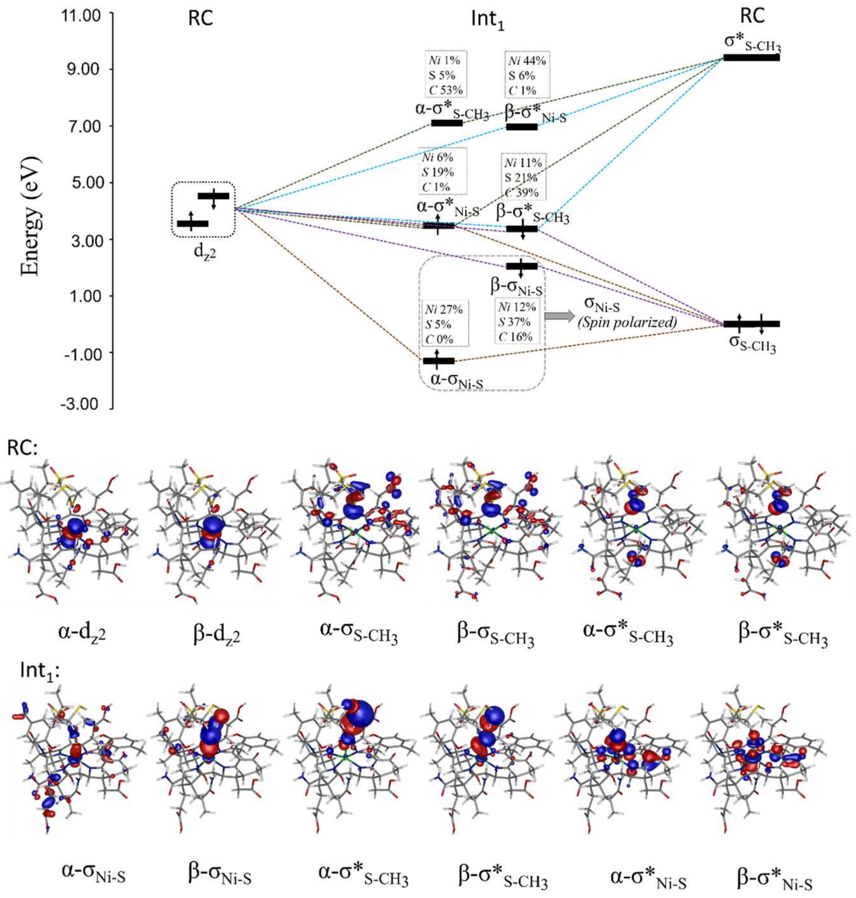


Figure 5. Representative MO diagram and molecular orbitals for **step 1** going from RC to Int$_1$ as shown for **E**. The analogous MO diagrams for models **A**–**D** are given in Figure S9. MO energies are shown relative to the σ_{S-CH_3} energy, which serves as a reference value of 0 eV.

the free energies of reaction and activation in **step 1**. From Figure 4 (*right graph*), $\Delta G'_{ET}$ correlates nicely with $\Delta G'_{0,1}$ and thus with the actual free energy of reaction $\Delta G_{0,1}$. However, the obtained correlation is counter-intuitive: an easier oxidation of Ni^{I} is linked to a more endergonic reaction (with a more positive $\Delta G_{0,1}$). In other words, the reductive cleavage of the S−CH$_3$ bond facilitated by the Ni cofactor is the most favorable in the native active-site **E**, despite the fact that F430 is the least capable of donating electrons among Ni complexes. This implies that the effect of $E°$ on reaction energetics must be outweighed by another contribution, which we reveal in the upcoming analysis.

Ni−S Bond Formation and Its Correlation with Reactivity in Step 1. The factor that dominates over the effect of $E°$ is the free energy of bond formation between Ni^{II} and thiolate ($\Delta G'_{NiS}$). From Figure 4 (*right*), $\Delta G'_{NiS}$ correlates with $\Delta G'_{0,1}$ so that a stronger Ni^{II}−S bond yields a more favorable reaction. The strongest bond is calculated for the native F430 in **E**, which pulls down $\Delta G_{0,1}$ and also the barrier $\Delta G_1^{\neq}$. Moreover, the correlation slope for $\Delta G'_{NiS}$ vs $\Delta G'_{0,1}$ is 2.8 as compared to the slope of −1.8 calculated for $E°$ vs $\Delta G'_{0,1}$ (Figure 4). Thus, the change in the Ni−S bond strength across systems **A**−**E** dominates the change in $\Delta G'_{0,1}$, outweighing the counter-acting effect of $E°$ by a factor of ~1.5. As a result, the overall change in $[\Delta G'_{ET} + \Delta G'_{NiS}]$ quantitatively yields differences seen in free-energy profiles of **step 1** within **A**−**E** in Figure 3B. Our finding also suggests that the driving force for the biosynthesis of F430 consists in the electronic-structure adjustment of Ni to lower the barrier for methane production by strengthening the Ni−S bond formed during the catalytic **step 1**.

The representative molecular orbital (MO) diagram is given in Figure 5. It describes the four-electron three-center interaction between bonding/antibonding orbitals σ_{S-CH_3}/$\sigma^{*}_{S-CH_3}$ of coenzyme M and the Ni-d_{z^2} orbital in going from RC to Int$_1$ in **step 1**. First, we notice a strong spin polarization of the Ni−S bond, with bonding *α*- and *β*-electron to be spatially and energetically distorted so that the *α*-electron is lower in energy and located more on Ni (due to exchange stabilization of *α*-electrons on Ni^{II}), while the *β*-electron is higher in energy and located more on thiolate (cf. α-σ_{Ni-S} vs β-σ_{Ni-S} MOs). This immediately implies a significant covalent character of the Ni−S bond, with a pronounced thiyl character on S. Second, one of the two electrons initially located on Ni^{I} is transferred to the $\sigma^{*}_{S-CH_3}$ orbital, while one of the two σ_{S-CH_3}electrons translocates to σ^{*}_{Ni-S}. Among the set **A**−**E**, the strongest Ni−S bond (calculated for **E** as shown in Figure 4) is the most covalent as reflected by the atomic composition of α-σ_{Ni-S} and β-σ_{Ni-S} with the largest Ni character in α-σ_{Ni-S} and thiyl character on S in β-σ_{Ni-S} (Figure S10). Conversely, the weakest Ni−S bond (calculated for **B**; Figure 4) is the least covalent as evidenced by the lowest Ni character in α-σ_{Ni-S} and thiyl character on S in β-σ_{Ni-S} (Figure S10). The most pronounced electron donation from thiolate to Ni in Int$_1$ found for the native model **E** indicates the native corphinoid ligand to be a weaker electron donor compared to its biosynthetic precursors.

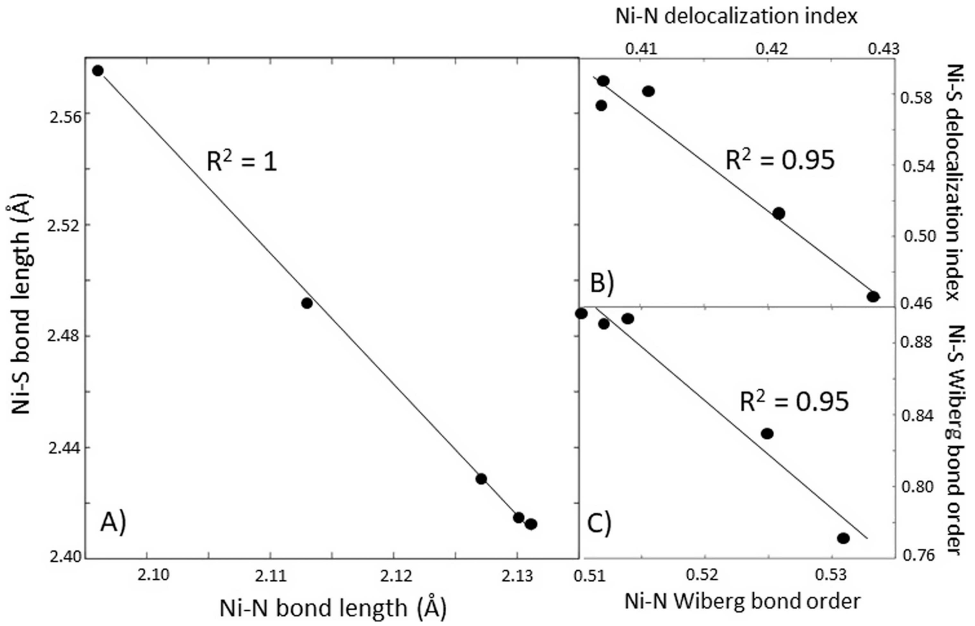


Figure 6. Correlation plots between the average length of the four bonds formed by Ni with the four ligating nitrogen atoms of porphyrin-like skeleton and Ni−S_{CoM} length (A); between the Ni−S and average Ni−N delocalization indices, obtained using the AIM formalism (B); and between the Ni−S and average Ni−N bond orders obtained through the Wiberg bond order analysis in Löwdin orthogonalized basis[49] (C). Correlations were performed for the Int_1 structures of the models **A**−**E**.

The electron-donation ability of the macrocyclic ligands in **A**−**E** was analyzed by means of AIM integration of the DFT-optimized electron density of their RC state (from Figure 3B). To this aim, the Ni ion was replaced by a point charge (q_P) whose value varies from 0 to 1*e* in 0.1*e* steps. It probes ligand polarization as q_P grows. Ligand polarization $P(q_P)$ is calculated as

$$P(q_p) = \sum_{i=1}^{n} |q_i(q_p) - q_i(0)| \tag{5}$$

where *n* is the number of atoms in the system, $q_i(q_P)$ is the charge of the *i*th atom in response to the point charge q_P, and $q_i(0)$ is the charge of the *i*th atom when $q_P = 0e$. As shown in Figure S11, $P(q_p)$ varies linearly in **A**−**E**, with the ligand in **A** being the most sensitive to increasing q_P (reflected in its slope $m_A = 1.34$, the highest in the set). Such a polarization gradually gets less responsive as the biosynthetic precursor approaches the native ligand F430 anchored in model **E**, for which the slope of $P(q_p)$ is 1.02. Thus, the corphinoid ligand in F430 is tuned to be the least responsive to the electric charge of the coordinated Ni ion, following the trend seen for the Ni−S bond strength and hence for the barrier height of the rate-determining step. The observed trend in ligand polarizability can be intuitively understood since the non-native cofactor **A** features the largest π-system which becomes increasingly saturated en route to **E**. Now, we will interrogate the connection of the distinct polarizabilities of ligands in **A**−**E** with the strengthening of the nascent Ni−S bond, which emerges as a key determinant of the efficacy of MCR.

The bonding between Ni^{II} and the macrocyclic ligands in **A**−**E** was calculated through AIM-derived atomic charges (q_{AIM}), which allowed a direct estimation of differential interaction strengths and delocalization indices as measures of electron sharing. From Figure S12, the average q_{AIM} of the four ligating N atoms is the least and the most negative for **E** (−1.15*e*) and **A** (−1.17*e*), respectively, while q_{AIM} for the Ni^{II} gradually gets more positive going back in the biosynthetic pathway (+1.17*e* for **E**, +1.21*e* for **A**). With these charge differences, we estimated the differential Ni^{II}-ligand interaction by means of Coulombic law to be ca. 12 kcal mol^{-1} weaker in **E** than in **A** (see Supporting Information for the details and Table S6 with the values for systems **A**−**E**). This quantity, derived from electrostatics, misses the covalent component of bonding. However, the fact that the ligating N atoms in the native F430 bear the lowest electron density translates into the smallest Ni−N bond order, which translates into the largest Ni−S bond order (and hence, the highest covalency) due to electron-donation competition. The electrostatic analysis agrees with the change in both Ni−N and Ni−S bond lengths, displaying the shortest Ni−S and the longest Ni−N distances in the native **E** (Figure 6). Such a negative proportionality between the Ni−N and Ni−S distances (and their bond orders) is also observed in Figure S13 for the simplified models used in the thermodynamic cycle analysis in Figure 4. As a final and qualitative note, the macrocyclic π-system is the least conjugated in model **E**, which implies the densest π-system within the **A**−**E** set, i.e., the most capable of displacing σ electron density from the π-region. This effect may contribute to a lower σ-electron density on N (due to σ/π-electron density polarization[48]), leading to a lower Ni−N covalency. We also largely attribute to this effect the trends observed in the reduction potential and splitting of the spin (singlet/triplet) states of the Ni^{II} center in **A**−**E**, discussed earlier in the text. As for the metal-ring π-interactions, we do not expect them to affect the trends in Ni−S/Ni−N covalency and thus the reactivity due to full electron occupancy of $d_\pi(xz/yz)$ orbitals. Although it is reasonable to expect electronic repulsion between d_π and π-ring electrons to be less pronounced for more conjugated models (e.g., **A**) compared to less conjugated **E**, the π-metal/ring repulsion is expected to be approximately the same in both redox states of Ni and,

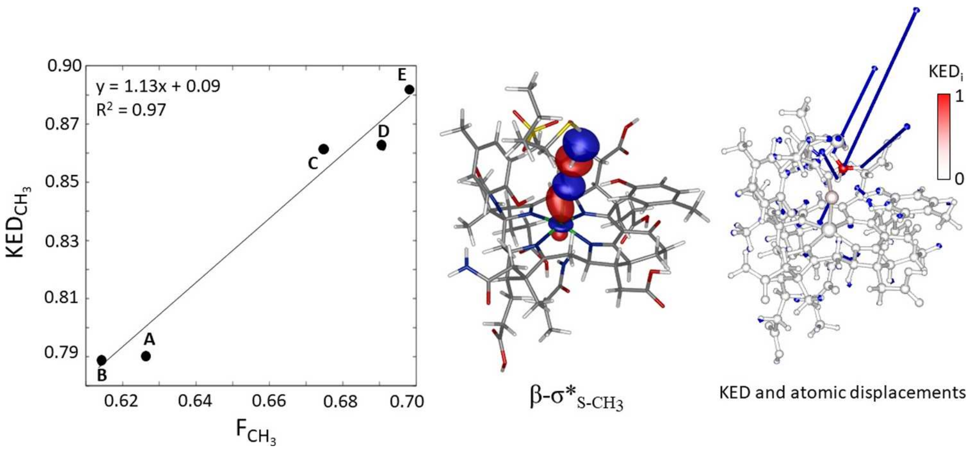


Figure 7. Correlation plot between fraction of kinetic energy of the reactive mode belonging to the transient CH_3 radical (KED_{CH_3}) and fraction of the atomic p_C orbital in $\sigma^*_{\mathrm{S-CH_3}}$ relative to the atomic p_S orbital (F_{CH_3}) for **A**−**E**, where F_{CH_3} calculated as ($p_C[\%]/(p_C[\%] + p_S[\%])$), (*left*). Note that an analogous correlation of $KED_{CH_3}$with the % of p_C in $\sigma^*_{\mathrm{S-CH_3}}$ is given in Figure S15. Relevant β-$\sigma^*_{\mathrm{S-CH_3}}$orbital for the key S−$CH_3$ cleavage step, associated with KED concentrated in the motion of the CH_3 radical (*center*). KED mapping and atomic displacements in the reactive mode of the S−CH_3 cleavage step (*right*).

therefore, to have a minor impact on the observed trends in **A**−**E**.

In addition, a consistent trend is obtained for Ni−S vs Ni−N Mulliken bond order derived from CASSCF calculations, which were performed on top of the DFT-optimized Int_1 structures of the models **A**−**E** (Figure S14). Here, the observed trend is rationalized through the configuration interaction (CI) vector analysis of the CASSCF wavefunctions: the admixture of the S p_z → Ni d_{z^2} excited states in the ground state gradually increases from **B**/**A** to **E**, while the contributions of N $p_{x/y}$ → Ni $d_{x^2-y^2}$ excited states generally decrease. It is also noticeable that the dominant determinant with the contribution of ∼60% to the total CASSCF wavefunction (in case of all models) corresponds to the antiferromagnetic coupling between the triplet configuration on the nickel ion and the methyl radical, while the two subsequent determinants (each with the weight of 15%) are characterized as methyl radicals combined with the open-shell singlet configuration on the nickel ion. Thus, ∼90% of the total CASSCF wavefunction comes from a single configuration with three singly occupied orbitals, which is fully consistent with the presented DFT calculations.

Connection between Kinetic Energy Distribution within the Transition-State Reactive Mode in Step 1 and Macrocycle Electron-Donation Ability. As already noticed, a β-electron from Ni-d_{z^2} transfers to the unoccupied antibonding $\sigma^*_{\mathrm{S-CH_3}}$ orbital along the reaction coordinate of **step 1** and thereby weakens the S−CH_3 bond. Electron transfer is accomplished at TS_1, yielding the essentially broken S−CH_3 bond with a kinetic energy distribution within the TS_1 reactive mode that is predominantly localized on the transient methyl radical (KED_{CH_3}= 0.88 for model **E**; see earlier text, Figure S7). The singly occupied β-$\sigma^*_{\mathrm{S-CH_3}}$ orbital at TS_1 (β-$\sigma^*_{\mathrm{S-CH_3}}$ in Int_1) keeps its antibonding character between the p_S and p_C atomic orbitals, the latter of which is dominant (e.g., ∼19 vs ∼43% for model **E**). Considering only these two components, the fraction of p_C in $\sigma^*_{\mathrm{S-CH_3}}$ relative to p_S, which is given as $F_{CH_3} = (p_C[\%]/(p_C[\%] + p_S[\%]))$, lies within the range of ∼0.6−0.7 across **A**−**E**. This is close to the ∼0.8−0.9 range for the corresponding KED_{CH_3} values. Indeed, as shown in Figure 7, F_{CH_3} correlates quantitatively with KED_{CH_3}. A larger asymmetry in β-$\sigma^*_{\mathrm{S-CH_3}}$ in favor of the CH_3 group, which yields a more concentrated KED on CH_3, is directly linked to a larger asymmetry in β-$\sigma_{\mathrm{Ni-S}}$ in favor of the sulfur component, which, in turn, and due to the above-mentioned spin polarization, is connected with a larger α-$\sigma_{\mathrm{Ni-S}}$ asymmetry in favor of the Ni component. The magnitude of KED_{CH_3} and its evolution across the models **A**−**E** is therefore a direct consequence of Ni−S bond covalency and corphinoid electron-donation ability.

Redox and Ni−S Bond Strength Factors and Their Effects on Reaction Step 3. As already shown above, **step 3** is not found to be rate-determining and therefore not as critical as **step 1**. However, it is still interesting to understand whether and how the interplay of redox vs Ni−S bond strength, which controls the energetics of **step 1**, also affects **step 3**. Following Figure 3B, **step 3** leads to formation of the disulfidic bond between coenzymes M and B, with the concomitant one-electron reduction of the Ni^{II} center and Ni−S bond cleavage. Thus, from the perspective of the cofactor, **step 3** can be considered as a reverse process of **step 1**.

Indeed, **step 3** is calculated to be the least exergonic for the model **E** with the native F430, while it becomes gradually more exergonic in going backward from late- to early-stage biosynthetic precursors (**step 3b** in Figure 3B). In parallel, the cleavage of the Ni−S bond gets increasingly feasible as reduction of the Ni^{II} center becomes less favorable in going from **E** to **A**. Formation of the S−S bond is independent of the Ni cofactor. In line with the trend in reaction free energy of the third step seen across the set, the same is seen for the barrier $\Delta G_3^{\neq}$. To gain a better insight in **step 3**, we deconstructed it through a thermodynamic cycle (Figure S16) to three distinct events: (i) the homolytic cleavage of the Ni−S bond ($\Delta G'_{\text{NiS-cleavage}}$), (ii) one-electron transfer from the anionic thiolate of SCoM to the Ni^{II} center ($\Delta G'_{\mathrm{ET}}$), and (iii) the S−S bond formation ($\Delta G'_{\mathrm{S-S}}$) between CoBS and SCoM. Overall, the sum of all three contributions ($\Delta G'_{0,3} = \Delta G'_{\text{NiS-cleavage}} + \Delta G'_{\mathrm{ET}} + \Delta G'_{\mathrm{S-S}}$) is calculated to be the least negative for **E**.

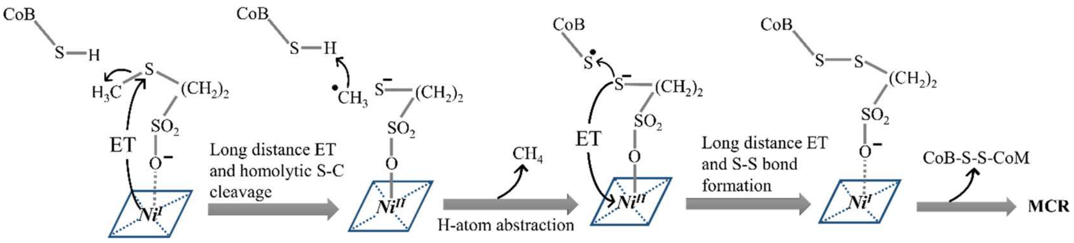
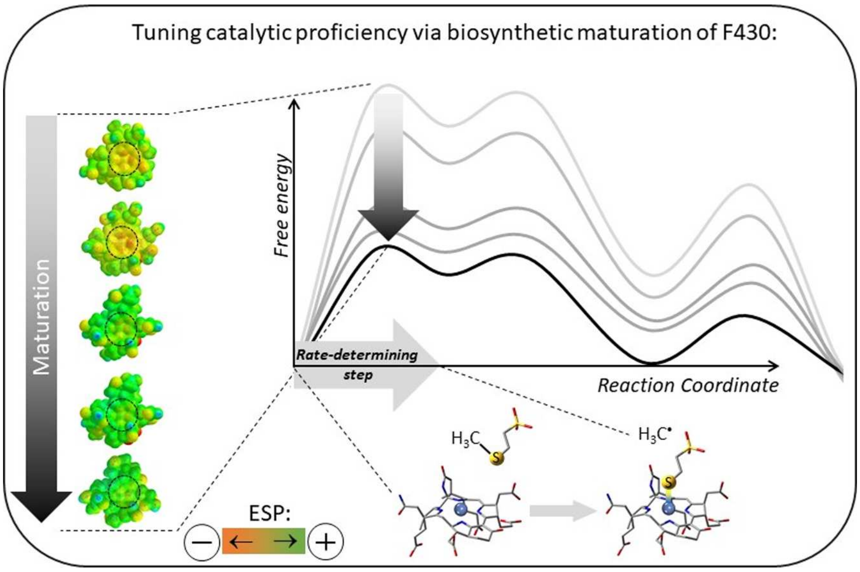


Figure 8. Alternative mechanism for production of methane by MCR as reported in ref 22.

Scheme 2. Biosynthesis of the F430 Cofactor Includes Four Sequentially Modified Precursors, Which Are Found to Be Less Efficient Catalysts in Production of CH_4 Than the Native F430[a]



[a]The driving force for the highest catalytic competence of F430 is the formation of the strongest Ni−S bond in the rate-determining step that is allowed by the lowest electron-donation ability of the corphinoid ligand, as reflected by the least negative ESP in the zone bounded by dashed circles in the ESP-contoured structures.

Since $\Delta G'_{0,3}$ correlates satisfactorily with the reaction free energy of **step 3** ($\Delta G_{0,3}$; Figure 3B), we conclude that the native **E** has the least favorable **step 3** due to the least favorable Ni−S bond cleavage. While both the reductive cleavage of S−CH_3 and the Ni−S bond formation are essentially accomplished at TS_1 in **step 1**, the oxidative S−S bond formation (i.e., the reduction of Ni) is more advanced than disruption of the Ni−S bond at TS_3 in **step 3** (Figure S17). Thus, asynchronicity in favor of the Ni reduction, which is most pronounced in **E**, outcompetes the unfavorable Ni−S bond cleavage. As a consequence, the opposite change in $\Delta G_3^{\neq}$ is less pronounced as compared to the change in rate-determining $\Delta G_1^{\neq}$ when passing from **A** to **E** (Figure S18).

Key Reaction Factors in the Context of a Recently Proposed Mechanism. Another mechanism proposed in literature for MCR-catalyzed production of CH_4 is sketched in Figure 8.[22] Here, we briefly make a remark on the viability of such a mechanism in light of our findings on the redox properties of the cofactor. As we elaborate in this work, the biosynthetic maturation of F430 through four distinct steps systematically fine-tunes the rate-determining **step 1** from Figure 1. In contrast, the alternative pathway reported in ref 22 seems triggered by long-distance electron transfer from Ni^I to the remote S−CH_3 bond of coenzyme M. In such a mechanism, the Ni center remains coordinated by the sulfonate group of the coenzyme M along the whole catalytic process. If the Ni−OSO_2R bond were not significantly strengthened during the oxidation of the Ni center, the electron transfer from Ni to the distant S−CH_3 bond (Ni oxidation) would be central to this rate-determining barrier, in contrast to the canonical mechanism in Figure 1. Since Ni oxidation is less favorable in going from early-stage through late-stage precursors to the native F430, the alternative mechanism recently proposed by Ragsdale and co-workers[22] could therefore be incompatible with the assumption that maturation of the cofactor F430 evolved to make production of CH_4 effective. This deserves further investigation.

■ CONCLUSIONS

Methanogenesis is a form of anaerobic respiration. As such, it has been evolutionary tuned to be as effective as possible. A final step in such a unique metabolic pathway produces methane from the reaction between a thiol and a thioether, which is catalyzed by MCR. The power horse of this enzyme is the active site hosting the Ni-containing F430 cofactor, whose catalytic properties were studied under the prism of evolutionary driving force for its biosynthesis. To this aim, we took advantage of four recently discovered biosynthetic precursors of F430 and investigated their reactivity relative to the native F430. Indeed, we found that F430 is best-suited for catalysis, displaying the lowest barrier for the rate-determining step involving the reductive cleavage of the thioether S−CH_3 bond by the Ni^{I} center (Scheme 2). Surprisingly, the native F430 cofactor has the highest reduction potential and, therefore, from this perspective, it would be the least effective reductant for the cleavage of the $S-CH_3$ bond. In fact, there is another factor that makes F430 the most effective in facilitating this critical step and that outweighs the unfavorable reduction potential: the strength of the Ni−S bond, which is formed upon the reductive $S-CH_3$ cleavage and which is the strongest across the series of the studied active-site models anchoring F430 and its four biosynthetic precursors (Scheme 2). The strongest Ni−S bond is attributed to the highest covalent character, result of the weakest electron-donation ability of the native porphyrin-like F430 skeleton, which arises from complex (energy-demanding) chemical modifications in the biosynthetic pathway of F430. We also found that the transient methyl radical formed upon the reductive cleavage of $S-CH_3$ concentrates most of the kinetic energy of the reactive mode at the corresponding transition state, which may facilitate a subsequent ballistic hydrogen atom abstraction from coenzyme B. Such a *dynamic* feature, which is again best-suited for the native F430 TS (~90% of the kinetic energy of the reactive mode), is a direct consequence of the composition of the $\sigma^{*}_{S-CH_3}$ orbital at the TS, which in turn depends on the electronic-structure properties of the Ni cofactor, namely, the electron-donation ability of the corphinoid skeleton.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.3c00469.

> Depiction and cartesian coordinates of all models employed for the calculation of reduction potentials and for the modeling of reactivity; representative MO diagrams for models **A**−**D**; energetics for all studied thermodynamic cycles; and spin densities, kinetic energy distributions, and electron density analyses (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Martin Srnec − *J. Heyrovský Institute of Physical Chemistry, Czech Academy of Sciences, Prague 182 23, Czech Republic;* orcid.org/0000-0001-5118-141X; Email: martin.srnec@jh-inst.cas.cz

Authors

Priyam Bharadwaz − *J. Heyrovský Institute of Physical Chemistry, Czech Academy of Sciences, Prague 182 23, Czech Republic*

Mauricio Maldonado-Domínguez − *J. Heyrovský Institute of Physical Chemistry, Czech Academy of Sciences, Prague 182 23, Czech Republic;* orcid.org/0000-0002-4260-3032

Jakub Chalupský − *J. Heyrovský Institute of Physical Chemistry, Czech Academy of Sciences, Prague 182 23, Czech Republic;* orcid.org/0000-0002-1584-9272

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.3c00469

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This project was supported by the Grant Agency of the Czech Republic (21-10383S). Computer time at the IT4I supercomputer center (project LM2018140 of the MSMT CR) is gratefully acknowledged.

■ REFERENCES

(1) Prakash, D.; Wu, Y.; Suh, S. J.; Duin, E. C. Elucidating the process of activation of methyl-coenzyme M reductase. *J. Bacteriol.* **2014**, *196*, 2491−2498.

(2) Lide, D. R. (Ed.) *CRC handbook of chemistry and physics*; CRC press, 2004; Vol. *85*.

(3) Allen, K. D.; Wegener, G.; White, R. H. Discovery of multiple modified F-430 coenzymes in methanogens and anaerobic methanotrophic archaea suggests possible new roles for F-430 in nature. *Appl. Environ. Microbiol.* **2014**, *80*, 6403−6412.

(4) Richard, R. M.; Ball, D. W. Ab initio calculations on the thermodynamic properties of azaborospiropentanes. *J. Mol. Model.* **2008**, *14*, 871−878.

(5) Thauer, R. K. Biochemistry of methanogenesis: a tribute to Marjory Stephenson: 1998 Marjory Stephenson prize lecture. *Microbiology* **1998**, *144*, 2377−2406.

(6) Krüger, M.; Meyerdierks, A.; Glöckner, F. O.; Amann, R.; Widdel, F.; Kube, M.; Shima, S. A conspicuous nickel protein in microbial mats that oxidize methane anaerobically. *Nature* **2003**, *426*, 878−881.

(7) Shima, S.; Krueger, M.; Weinert, T.; Demmer, U.; Kahnt, J.; Thauer, R. K.; Ermler, U. Structure of a methyl-coenzyme M reductase from Black Sea mats that oxidize methane anaerobically. *Nature* **2012**, *481*, 98−101.

(8) Scheller, S.; Goenrich, M.; Boecher, R.; Thauer, R. K.; Jaun, B. The key nickel enzyme of methanogenesis catalyses the anaerobic oxidation of methane. *Nature* **2010**, *465*, 606−608.

(9) Shima, S.; Thauer, R. K. Methyl-coenzyme M reductase and the anaerobic oxidation of methane in methanotrophic Archaea. *Curr. Opin. Microbiol.* **2005**, *8*, 643−648.

(10) Timmers, P. H. A.; Welte, C. U.; Koehorst, J. J.; Plugge, C. M.; Jetten, M. S. M.; Stams, A. J. M. Reverse methanogenesis and respiration in methanotrophic archaea. *Archaea* **2017**, *2017*, No. 1654237.

(11) Lemaire, O. N.; Wagner, T. A Structural View of Alkyl-Coenzyme M Reductases, the First Step of Alkane Anaerobic Oxidation Catalyzed by Archaea. *Biochemistry* **2022**, *61*, 805−821.

(12) Thauer, R. K. Methyl (alkyl)-coenzyme M reductases: nickel F-430-containing enzymes involved in anaerobic methane formation and in anaerobic oxidation of methane or of short chain alkanes. *Biochemistry* **2019**, *58*, 5198−5220.

(13) Lane, N. *The Vital Question: Why is Life the Way it is?*; Profile Books, 2016.

(14) Ermler, U.; Grabarse, W.; Shima, S.; Goubeaud, M.; Thauer, R. K. Crystal structure of methyl-coenzyme M reductase: the key enzyme of biological methane formation. *Science* **1997**, *278*, 1457−1462.

(15) Dey, M.; Li, X.; Kunz, R. C.; Ragsdale, S. W. Detection of organometallic and radical intermediates in the catalytic mechanism of methyl-coenzyme M reductase using the natural substrate methyl-coenzyme M and a coenzyme B substrate analogue. *Biochemistry* **2010**, *49*, 10902−10911.

(16) Signor, L.; Knuppe, C.; Hug, R.; Schweizer, B.; Pfaltz, A.; Jaun, B. Methane formation by reaction of a methyl Thioether with a photo-excited nickel Thiolate—a process mimicking methanogenesis in Archaea. *Chem. − Eur. J.* **2000**, *6*, 3508−3516.

(17) Grabarse, W.; Mahlert, F.; Duin, E. C.; Goubeaud, M.; Shima, S.; Thauer, R. K.; Victor, L.; Ermler, U. On the mechanism of biological methane formation: structural evidence for conformational changes in methyl-coenzyme M reductase upon substrate binding. *J. Mol. Biol.* **2001**, *309*, 315−330.

(18) Duin, E. C.; McKee, M. L. A new mechanism for methane production from methyl-coenzyme M reductase as derived from

density functional calculations. *J. Phys. Chem. B* **2008**, *112*, 2466−2482.

(19) Pelmenschikov, V.; Blomberg, M. R.; Siegbahn, P. E.; Crabtree, R. H. A mechanism from quantum chemical studies for methane formation in methanogenesis. *J. Am. Chem. Soc.* **2002**, *124*, 4039−4049.

(20) Pelmenschikov, V.; Siegbahn, P. E. Catalysis by methyl-coenzyme M reductase: a theoretical study for heterodisulfide product formation. *J. Biol. Inorg. Chem.* **2003**, *8*, 653−662.

(21) Chen, S. L.; Blomberg, M. R.; Siegbahn, P. E. How Is Methane Formed and Oxidized Reversibly When Catalyzed by Ni-Containing Methyl-Coenzyme M Reductase? *Chem. − Eur. J.* **2012**, *18*, 6309−6315.

(22) Patwardhan, A.; Sarangi, R.; Ginovska, B.; Raugei, S.; Ragsdale, S. W. Nickel−sulfonate mode of substrate binding for forward and reverse reactions of methyl-SCoM reductase suggest a radical mechanism involving long-range electron transfer. *J. Am. Chem. Soc.* **2021**, *143*, 5481−5496.

(23) Chen, S. L.; Blomberg, M. R.; Siegbahn, P. E. An investigation of possible competing mechanisms for Ni-containing methyl−coenzyme M reductase. *Phys. Chem. Chem. Phys.* **2014**, *16*, 14029−14035.

(24) Wongnate, T.; Sliwa, D.; Ginovska, B.; Smith, D.; Wolf, M. W.; Lehnert, N.; Simone, R.; Ragsdale, S. W. The radical mechanism of biological methane synthesis by methyl-coenzyme M reductase. *Science* **2016**, *352*, 953−958.

(25) Moore, S. J.; Sowa, S. T.; Schuchardt, C.; Deery, E.; Lawrence, A. D.; Ramos, J. V.; Billig, S.; Birkemeyer, C.; Chivers, P. T.; Howard, M. J.; Rigby, S. E. J.; Layer, G.; Warren, M. J. Elucidation of the biosynthesis of the methane catalyst coenzyme F430. *Nature* **2017**, *543*, 78−82.

(26) Becke, A. D. A new mixing of Hartree−Fock and local density-functional theories. *J. Chem. Phys.* **1993**, *98*, 1372−1377.

(27) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.

(28) (a) Hay, P. J.; Wadt, W. R. Ab initio effective core potentials for molecular calculations. Potentials for the transition metal atoms Sc to Hg. *J. Chem. Phys.* **1985**, *82*, 270−283. (b) Wadt, W. R.; Hay, P. J. Ab initio effective core potentials for molecular calculations − potentials for main group elements Na to Bi. *J. Chem. Phys.* **1985**, *82*, 284−298. (c) Hay, P. J.; Wadt, W. R. Ab initio effective core potentials for molecular calculations. Potentials for K to Au including the outermost core orbitals. *J. Chem. Phys.* **1985**, *82*, 299−310.

(29) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A. Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. *J. Chem. Phys.* **1982**, *77*, 3654−3665.

(30) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. *J. Comput. Chem.* **2003**, *24*, 669−681.

(31) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16 Revision C.01*; Gaussian, Inc.: Wallingford CT, 2016.

(32) Biegler-König, F. W.; Bader, R. F. W.; Tang, T.-H. Calculation of the Average Properties of Atoms in Molecules. II. *J. Comput. Chem.* **1982**, *3*, 317−328.

(33) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. Aqueous Solvation Free Energies of Ions and Ion-Water Clusters Based on an Accurate Value for the Absolute Aqueous Solvation Free Energy of the Proton. *J. Phys. Chem. B* **2006**, *110*, 16066−16081.

(34) Holliger, C.; Pierik, A. J.; Reijerse, E. J.; Hagen, W. R. A spectroelectrochemical study of factor F430 nickel (II/I) from methanogenic bacteria in aqueous solution. *J. Am. Chem. Soc.* **1993**, *115*, 5651−5656.

(35) Jaun, B.; Pfaltz, A. Coenzyme F430 from methanogenic bacteria: reversible one-electron reduction of F430 pentamethyl ester to the nickel (I) form. *J. Chem. Soc., Chem. Commun.* **1986**, *513*, 1327−1329.

(36) Zilbermann, I.; Maimon, E.; Cohen, H.; Meyerstein, D. Redox chemistry of nickel complexes in aqueous solutions. *Chem. Rev.* **2005**, *105*, 2609−2626.

(37) Kratky, C.; Fässler, A.; Pfaltz, A.; Kräutler, B.; Jaun, B.; Eschenmoser, A. Chemistry of corphinoids: structural properties of corphinoid nickel (II) complexes related to coenzyme F430. *J. Chem. Soc., Chem. Commun.* **1984**, 1368−1371.

(38) Chalupský, J. *Charmol: Program for molecular graphics*, 2018. https://sourceforge.net/projects/charmol (accessed Jun 19, 2022).

(39) (a) Widmark, P.-O.; Malmqvist, P.-Å.; Roos, B. O. Density matrix averaged atomic natural orbital (ANO) basis sets for correlated molecular wave functions. *Theor. Chim. Acta* **1990**, *77*, 291−306. (b) Roos, B. O.; Lindh, R.; Malmqvist, P.-Å.; Veryazov, V.; Widmark, P.-O. Main Group Atoms and Dimers Studied with a New Relativistic ANO Basis Set. *J. Phys. Chem. A* **2004**, *108*, 2851−2858. (c) Roos, B. O.; Lindh, R.; Malmqvist, P.-Å.; Veryazov, V.; Widmark, P.-O. New Relativistic ANO Basis Sets for Transition Metal Atoms. *J. Phys. Chem. A* **2005**, *109*, 6575−6579.

(40) Stoychev, G. L.; Auer, A. A.; Neese, F. Automatic Generation of Auxiliary Basis Sets. *J. Chem. Theory Comput.* **2017**, *13*, 554−562.

(41) (a) Douglas, M.; Kroll, N. M. Quantum electrodynamical corrections to the fine structure of helium. *Ann. Phys.* **1974**, *82*, 89−155. (b) Hess, B. A. Relativistic electronic-structure calculations employing a two-component no-pair formalism with external-field projection operators. *Phys. Rev. A* **1986**, *33*, 3742−3748. (c) Jansen, G.; Hess, B. A. Revision of the Douglas-Kroll transformation. *Phys. Rev. A* **1989**, *39*, 6016−6017.

(42) (a) Ruedenberg, K.; Schmidt, M.; Gilbert, M.; Elbert, S. Are atoms intrinsic to molecular electronic wavefunctions? I. The FORS model. *Chem. Phys.* **1982**, *71*, 41−49. (b) Roos, B. O. The complete active space self-consistent field method and its applications in electronic structure calculations. *Adv. Chem. Phys.* **1987**, *69*, 399−445.

(43) Yanai, T.; Kurashige, Y.; Mizukami, W.; Chalupský, J.; Lan, T. N.; Saitow, M. Density matrix renormalization group for ab initio Calculations and associated dynamic correlation methods: A review of theory and applications. *Int. J. Quantum Chem.* **2015**, *115*, 283−299.

(44) Maldonado-Domínguez, M.; Bím, D.; Fučík, R.; Čurík, R.; Srnec, M. Reactive mode composition factor analysis of transition states: The case of coupled electron−proton transfers. *Phys. Chem. Chem. Phys.* **2019**, *21*, 24912−24918.

(45) Bharadwaz, P.; Maldonado-Domínguez, M.; Srnec, M. Bifurcating reactions: distribution of products from energy distribution in a shared reactive mode. *Chem. Sci.* **2021**, *12*, 12682−12694.

(46) Maldonado-Dominguez, M.; Srnec, M. Understanding and predicting post H-atom abstraction selectivity through reactive mode composition factor analysis. *J. Am. Chem. Soc.* **2020**, *142*, 3947−3958.

(47) Sun, Y.; Benabbas, A.; Zeng, W.; Kleingardner, J. G.; Bren, K. L.; Champion, P. M. Investigations of heme distortion, low-frequency vibrational excitations, and electron transfer in cytochrome c. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 6570−6575.

(48) Wiberg, K. B.; Frisch, M. J. Effect of Conjugation of Electron Distributions. Separation of σ and π Terms. *J. Chem. Theory Comput.* **2016**, *12*, 1220−1227.

(49) Lu, T.; Chen, F. Multiwfn: a multifunctional wavefunction analyzer. *J. Comput. Chem.* **2012**, *33*, 580−592.

https://doi.org/10.1021/jacs.3c00469