Investigation of the mechanism of the SpnF-catalyzed [4+2]-cycloaddition reaction in the biosynthesis of spinosyn A

Byung-sun Jeon[a], Mark W. Ruszczycky[b,1], William K. Russell[c], Geng-Min Lin[a], Namho Kim[b], Sei-hyun Choi[a], Shao-An Wang[a], Yung-nan Liu[b], John W. Patrick[c], David H. Russell[c], and Hung-wen Liu[a,b,1]

[a]Department of Chemistry, University of Texas at Austin, Austin, TX 78712; [b]Division of Chemical Biology and Medicinal Chemistry, College of Pharmacy, University of Texas at Austin, Austin, TX 78712; and [c]Department of Chemistry, Texas A&M University, College Station, TX 77843

Edited by Jerrold Meinwald, Cornell University, Ithaca, NY, and approved August 7, 2017 (received for review June 12, 2017)

The Diels–Alder reaction is one of the most common methods to chemically synthesize a six-membered carbocycle. While it has long been speculated that the cyclohexene moiety found in many secondary metabolites is also introduced via similar chemistry, the enzyme SpnF involved in the biosynthesis of the insecticide spinosyn A in *Saccharopolyspora spinosa* is the first enzyme for which catalysis of an intramolecular [4+2]-cycloaddition has been experimentally verified as its only known function. Since its discovery, a number of additional standalone [4 + 2]-cyclases have been reported as potential Diels–Alderases; however, whether their catalytic cycles involve a concerted or stepwise cyclization mechanism has not been addressed experimentally. Here, we report direct experimental interrogation of the reaction coordinate for the [4 + 2]-carbocyclase SpnF via the measurement of α-secondary deuterium kinetic isotope effects (KIEs) at all sites of $sp^2 \rightarrow sp^3$ rehybridization for both the nonenzymatic and enzyme-catalyzed cyclization of the SpnF substrate. The measured KIEs for the nonenzymatic reaction are consistent with previous computational results implicating an intermediary state between formation of the first and second carbon–carbon bonds. The KIEs measured for the enzymatic reaction suggest a similar mechanism of cyclization within the enzyme active site; however, there is evidence that conformational restriction of the substrate may play a role in catalysis.

Diels–Alderase | isotope effect | enzyme | spinosyn | Bayesian analysis

The enzyme SpnF from *Saccharopolyspora spinosa* catalyzes an intramolecular [4 + 2]-cycloaddition (**4** → **5**) during biosynthesis of the insecticide spinosyn A (**1**) (Fig. 1) (1, 2). At the time of its discovery, the search for a naturally selected Diels–Alderase had already led to the characterization of several enzymes proposed to catalyze [4 + 2]-cycloadditions (3). In all cases, however, these potential Diels–Alderases were found to be multifunctional, often catalyzing a priming reaction that activates the substrate toward cyclization. Thus, the detailed mechanisms of cyclization and the catalytic roles played by the enzymes were unclear and difficult to address. In contrast, the priming reaction (**3** → **4**) during spinosyn A biosynthesis is catalyzed by the separate enzyme SpnM (1). The resulting 1, 4-dehydration product **4** can then undergo stereospecific cyclization in aqueous solution with a half-life of ∼25 min at 300 K; however, in the presence of SpnF, a rate enhancement (i.e., k_{cat}/k_{non}) of ∼500-fold is observed, making SpnF the first reported example of a standalone [4 + 2]-carbocyclase (1). This separation of activities has made SpnF an ideal system for mechanistic study.

Since the discovery of SpnF, a number of additional enzymes, such as VtsJ (4), PyrI4 (5), PyrE3 (5), TclM (6), TbtD (7), and AbyU (8), have been characterized as uniquely catalyzing [4+2]-cycloadditions; however, investigation of these enzymes has been limited primarily to crystallographic and computational studies in large part because of the structural complexity of their substrates and the limited options available to study the mechanisms of biological cycloaddition reactions (3). For the purposes of the following discussion, we regard an intermediate as a local minimum in the free energy along the reaction coordinate and a concerted cyclization as one lacking any such intermediate. Consequently, there has yet to be a direct experimental interrogation of the reaction coordinates that defines whether the catalytic cycles of these [4 + 2]-carbocyclases proceed via concerted or stepwise mechanisms.

Development of a convergent synthetic strategy (2) has made it possible to prepare isotopologs of the SpnF substrate, allowing study of the SpnF reaction via the measurement of secondary deuterium kinetic isotope effects (KIEs). Previous chemical model studies have indicated that concerted formation of both C–C bonds in a [4 + 2]-cycloaddition reaction (e.g., the C4–C12 and C7–C11 bonds in **1**) should result in inverse deuterium KIEs of about 5% at all four carbons undergoing sp^2 to sp^3 rehybridization (9, 10). In contrast, if cyclization is stepwise, with the first step most rate-limiting, then unit to normal isotope effects would be expected for carbons involved in the second bond formation (9, 11). Since a stepwise intramolecular cyclization for the SpnF-catalyzed cycloaddition (Fig. 2) is expected to proceed with C7–C11 bond formation first, the C4D and C12D secondary deuterium KIEs are of particular interest in determining the sequence of bond formation. With this goal in mind, we prepared four site-specifically deuterated isotopologs of **4** (C4D, C7D, C11D, and C12D) and measured the corresponding secondary deuterium KIEs on both the nonenzymatic and enzyme-catalyzed reactions. The C2D and C14D isotopologs were also

Significance

Computational and structural studies of putative Diels–Alderases and the reactions that they catalyze have provided testable models of the reaction coordinates and mechanisms of catalysis; however, there has yet to be a direct experimental interrogation of these hypotheses. Herein, the α-secondary deuterium kinetic isotope effects at all sites of rehybridization on the nonenzymatic and SpnF-catalyzed [4+2]-cycloaddition during biosynthesis of spinosyn A are reported. Results implicate an intermediary state between formation of the first and second carbon–carbon bonds and provide evidence that conformational restriction of the substrate may play a role in catalysis.

Author contributions: M.W.R. and H.-w.L. designed research; B.-s.J., W.K.R., G.-M.L., N.K., S.-h.C., S.-A.W., Y.-n.L., and J.W.P. performed research; W.K.R., J.W.P., and D.H.R. contributed new reagents/analytic tools; M.W.R. and H.-w.L. analyzed data; and B.-s.J., M.W.R., and H.-w.L. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: mwr8@utexas.edu or h.w.liu@mail.utexas.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1710496114/-/DCSupplemental.

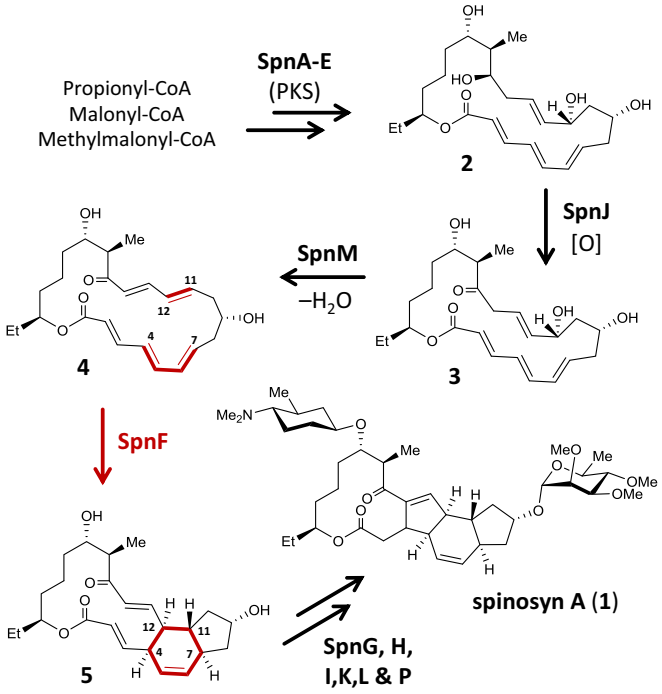

Fig. 1. Biosynthetic pathway of spinosyn A in *S. spinosa* (1, 2). SpnA–E together constitute a polyketide synthase (PKS) system.

prepared to assess the possible involvement of these two carbon centers in the cyclization reaction as predicted by a recent computational investigation (12).

Results and Discussion

The site-specific KIEs were measured using the method of internal competition (14) and whole-molecule MS to monitor changes in the isotopolog enrichment of residual substrate (15–17). As discussed in *SI Appendix*, *SI Text*, this approach requires using isotopically enriched substrates to enhance observable changes in isotopolog populations and minimize any measurement bias caused by natural abundance labeling. Natural abundance and site-specifically deuterated isotopologs of the macrolactone **2** were prepared by connecting the three fragments **A**, **B**, and **C** via two C–C bond coupling reactions as shown in Fig. 3. A Julia–Kocienski olefination between fragments **A** and **B** yielded the Δ^{12}-(*E*) olefin, and a palladium-catalyzed Stille coupling with fragment **C** led to the C-5/C-6 linkage. Subsequent cyclization of the resulting linear polyketide by Yamaguchi macrolactonization completed the assembly of **2**. The corresponding SpnM substrate (**3**) was then prepared and isolated from reactions with the biosynthetic enzyme SpnJ (18). This modular synthetic approach facilitated preparation of the six isotopologs of **3** by incorporating the appropriately deuterated fragments (**A**–**C**) as detailed in *SI Appendix*, *SI Text*.

Addition of SpnM to a mixture of natural abundance **3** and a site-specifically deuterated isotopolog in H_2O/DMSO at 30 °C generated the isotopically enriched **4** in situ. After complete consumption of **3**, residual **4** was isolated by HPLC from six reaction aliquots spanning ∼10–90% conversion of **4** to **5** in the presence of sufficient SpnF to ensure that reaction flux is predominantly through the enzyme-catalyzed reaction. Parallel experiments were also performed in the absence of SpnF to study the nonenzyme-catalyzed cyclization. The observed isotopolog ratio of each isolated sample of residual SpnF substrate (**4**) was then determined by electrospray ionization time-of-flight MS by taking the ratio of the $M+1$ and M peak intensities for the sodiated ion of **4** and correcting for natural abundance labeling (19). Experimental and analytical details are provided in *Materials and Methods*. At least three experimental trials, each consisting of approximately six isolated samples of residual **4**, were performed for all site-specific KIEs measured.

Since the expected secondary KIEs are all near unity, the ratio of isotopolog concentrations in the residual starting material (heavy over light) can be expressed as a function R of the fraction of reaction f parameterized in terms of the isotope effect ${}^{D}k$ and initial enrichment ratio ρ according to (20)

$$R(f;{}^{D}k,\rho)=\rho(1-f)^{1/{}^{D}k-1},\quad 0\le f<1. \qquad [1]$$

We used a hierarchical Bayesian regression analysis using Markov chain Monte Carlo (MCMC) (21, 22) rather than least squares fitting of Eq. **1** to model uncertainty in the measured KIEs caused by both within- and between-experiment variation while at the same time, estimating posterior probability distributions for the KIEs. Details regarding this analysis, including construction of priors, are described in *Materials and Methods*. Fig. 4 lists posterior point estimates of the site-specific KIEs as well as examples of enrichment plots for the C4D and C12D labeling patterns. The full dataset and simulated posterior marginal probability distributions of the KIEs are provided in *SI Appendix*, Figs. S3 and S4 and Tables S2 and S3.

Given a quasisteady state, the KIE on the cyclization of **4** to **5** in n elementary reactions (steps) can be expressed as the weighted average

$${}^{D}k=\sum_{i=1}^{n}S_i\,{}^{D}\kappa_i, \qquad [2]$$

where ${}^{D}\kappa_i$ is the effective KIE associated with progression from the reference state (i.e., **4** free in solution) to the ith transition state, and S_i is Ray's sensitivity index for the ith step (23–25). Each sensitivity index can be written as a reciprocal sum of forward and reverse commitments and reflects the degree to which a step may be considered "rate-limiting" (24, 25). Applied to the nonenzymatic reaction, the roughly 5% inverse values for

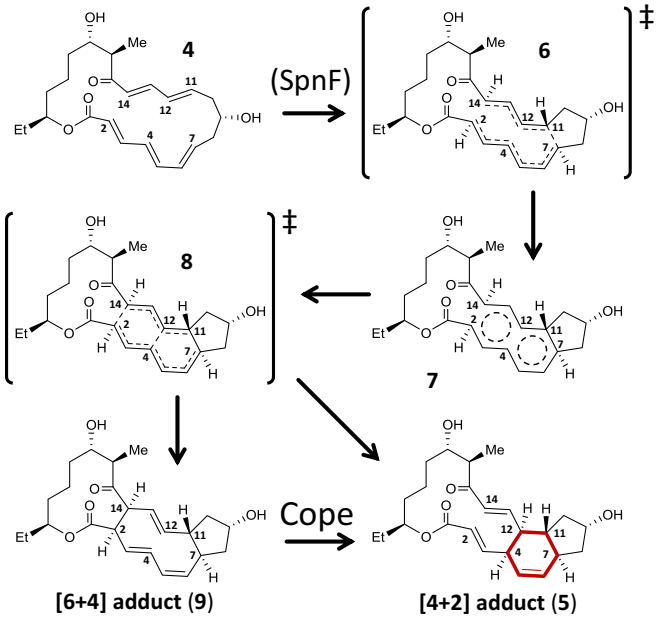

Fig. 2. Mechanistic proposal for the [4 + 2]-cycloaddition of the SpnF substrate **4**. Lewis structures represent putative local minima (intermediates) and maxima (transition states) along the generalized free energy curves of the bifurcating reaction coordinate.

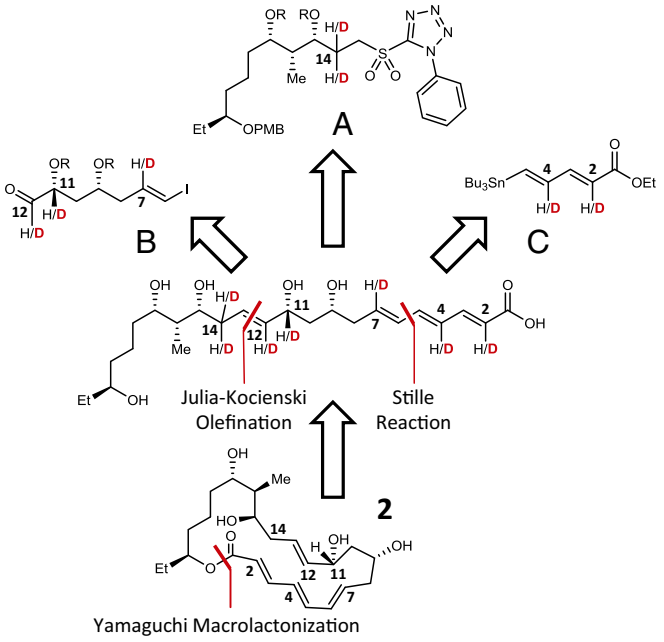

Fig. 3. Modular, retrosynthetic design for the preparation of natural abundance and site-specifically deuterated isotopologs of **2**. Carbons are numbered according to their final position in the macrolactone. Full synthetic schemata and protocols are provided in *SI Appendix*, *SI Text* and Figs. S5–S7.

$^{C7D}k_{non}$ and $^{C11D}k_{non}$ are indicative of partial C7–C11 bond formation in the transition state of the most rate-limiting step. Likewise, the values centered near unity for $^{C2D}k_{non}$, $^{C4D}k_{non}$, and $^{C14D}k_{non}$ suggest little to no $sp^2 \rightarrow sp^3$ rehybridization at these positions in this transition state compared with the reference state (**4**). The ∼4% normal value for $^{C12D}k_{non}$ may be explained by redistribution of electron density along the C11–C15 π-system, leading to a loosening of the C12–H bonding environment (9, 11, 26).

These results are consistent with recent computational models of the reaction (12, 13, 27, 28). In particular, a simplified gas-phase model of the nonenzymatic reaction investigated by Hess and Smentek (13) identified a reaction coordinate with a single potential energy saddle point (similar to **6** in Fig. 2) followed by two apparent inflection points forming a caldera-like region. Subsequent computations by the Houk and Singleton laboratories using variational transition-state theory identified a local free energy minimum described as a 10π-aromatic species (depicted as structure **7** in Fig. 2) residing about the caldera between two variational transition states along the reaction coordinate (12). The first of these transition states (**6**) is similar to that identified by Hess and Smentek (13), and the trajectory of **7** can bifurcate on progression through the second transition state (**8**) to form either a $[4+2]$- or $[6+4]$-cycloadduct (i.e., **5** vs. **9**), where the latter then converts to **5** via a facile Cope rearrangement (12). The second of these transition states (**8**) is ∼2.3 kcal/mol lower in free energy compared with the first (12). This implies that the sensitivity index at 300 K for the first transition state is nearly 50-fold greater than that for the second (24, 25). In other words, conversion of **4** to **7** via **6** is predicted to be significantly rate-limiting in the nonenzymatic reaction.

The KIEs calculated from Hess and Smentek's gas-phase models of **6** vs. **4** (13) are also shown in Fig. 4 and are directionally consistent with the observed values. Because of uncertainty in the measurements, there are ∼65 and 25% posterior probabilities that the $^{C7D}k_{non}$ and $^{C11D}k_{non}$ isotope effects, respectively, are between 0.90 and 0.95; however, the probability that they are less than 0.90 is no greater than 10% (*SI Appendix*, Fig. S3). Therefore, the reduced magnitude of $^{C7D}k_{non}$ and $^{C11D}k_{non}$ vs. the computed values suggests that the actual transition state (represented by **6**) is likely earlier than that predicted by the computational gas-phase model. The KIEs are not consistent, however, with a pericyclic $[4+2]$-transition state for a synchronous concerted cyclization, which would predict inverse effects at all four sites of $sp^2 \rightarrow sp^3$ rehybridization (9, 10, 29). In so far as the 10π-aromatic state can be considered an intermediate following **6**, these observations suggest that nonenzymatic cyclization of **4** may indeed be stepwise, albeit not in the usual sense of a biradical or dipolar intermediate. It should be emphasized, however, that the KIE measurements do not address existence of the $[6+4]$-cycloadduct (**9**) along the reaction coordinate, because they only provide information regarding the first, more rate-limiting elementary reaction.

The enzymatic KIEs can also be interpreted in terms of Eq. **2**, where $^{D}k_{enz}$ corresponds to the isotope effect on V/K for SpnF (14, 25). If substrate binding were the most rate-limiting (i.e., $S_1 \approx 1$), then roughly unit isotope effects would have been expected at all six positions. If product dissociation were the most rate-limiting (i.e., $S_n \approx 1$), then the observed KIE would reflect the equilibrium isotope effect on $sp^2 \rightarrow sp^3$ rehybridization (23, 25), and four of six KIEs would be at least 5% inverse regardless of whether the product released is a $[4+2]$- or $[6+4]$-cycloadduct (9, 29). That these outcomes were not observed suggests that the intermediary steps of the catalytic cycle dominate the observed enzymatic isotope effects.

Unlike the nonenzymatic cyclization, computational study of the SpnF-catalyzed reaction has been hampered, because a structure of the SpnF–substrate complex is not yet available. Nevertheless, the enzymatic results largely mirror those for the nonenzymatic reaction with the exception of $^{C4D}k_{enz}$, which is notably inverse, whereas the value of $^{C12D}k_{enz}$ is comparable with that of the nonenzymatic case as shown in Fig. 4. While the posterior marginal probability that $^{C4D}k_{enz} < 0.98$ is greater than 90%, the ∼5% normal value for $^{C12D}k_{enz}$ argues against significant C4–C12 bond formation in the transition state of the most rate-limiting step. Structural investigations of the $[4+2]$-carbocyclases SpnF, AbyU, and PyrE3 all implicate the formation of a closed complex on substrate binding into a largely hydrophobic active site (8, 30, 31). Therefore, the inverse value for $^{C4D}k_{enz}$ may reflect conformational restriction of **4** within a low-volume active site of the Michaelis complex, resulting in a stiffer bonding environment for the C4–H center and an inverse binding isotope effect (32).

Conclusions

As the number of recognized $[4+2]$-carbocyclases continues to grow, experimental interrogation of reaction coordinates proposed from structural and computational studies will be necessary for a complete mechanistic understanding of these enzymes. The experimental results and computational predictions are collectively most consistent with the hypothesis that the nonenzymatic cyclization of **4** involves an intermediary state before complete C4–C12 bond formation. A possible candidate for such an intermediate (i.e., local free energy minimum) is the 10π-aromatic state (e.g., **7**) identified in recent computational studies (12). The results also suggest that the SpnF-catalyzed reaction is similar to that of the nonenzymatic cyclization but likely involves a reduction in the entropy of activation after the substrate is bound in the active site, thereby facilitating catalysis and the observed rate enhancement (k_{cat}/k_{non}). These observations combined with prior structural investigations (8, 30, 31) suggest that $[4+2]$-carbocyclases in general may operate, in part, by reducing the conformation space available to the cyclizing species in a closed, induced fit complex, a criteria worthy

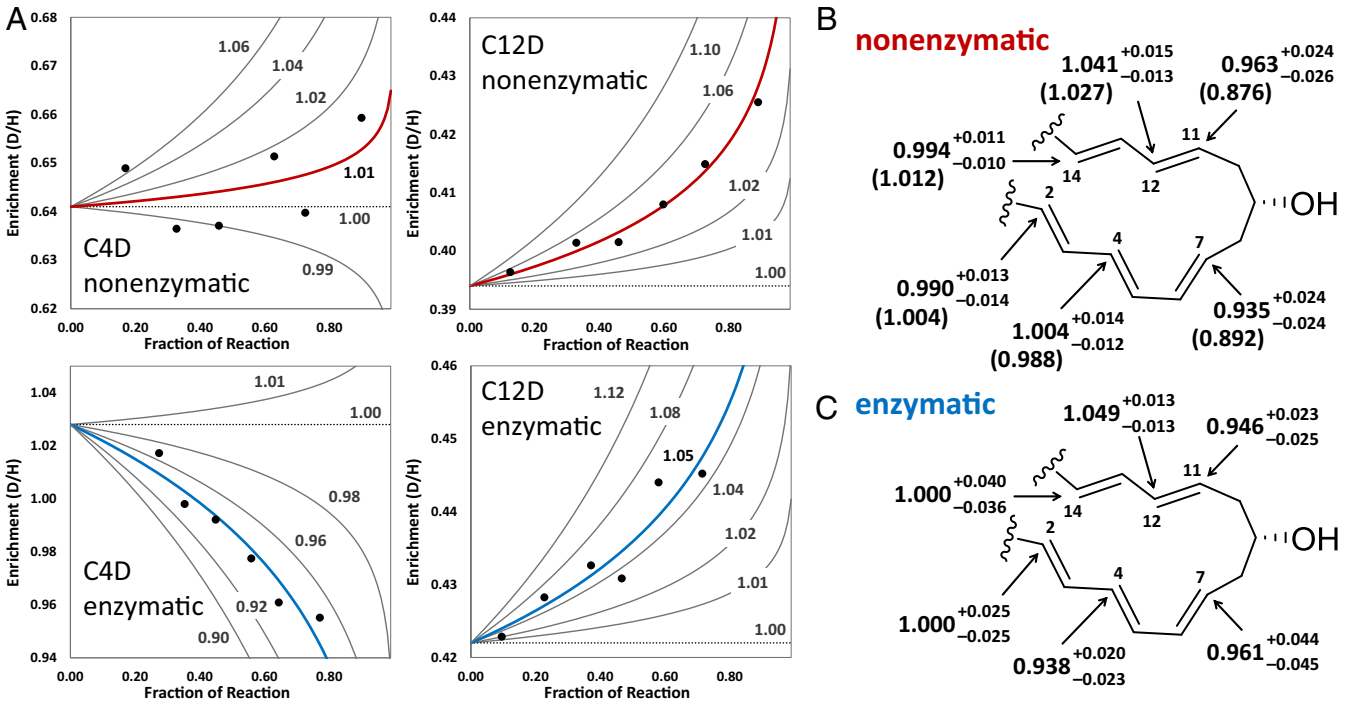

Fig. 4. Summary of the secondary deuterium KIEs on the enzymatic and nonenzymatic cyclization of **4**. (*A*) Examples of enrichment plots from experimental trials for the C4D and C12D KIEs for both the nonenzymatic and enzymatic reactions. Black circles denote the observed enrichments (D/H, heavy over light) in the isolated residual substrate. Fitted lines (red/blue) are drawn using the median values of the posterior marginal probability distributions for the initial enrichment and KIE corresponding to the experimental trials shown (*Materials and Methods*). (*B*) Overall site-specific KIEs for the nonenzymatic reaction. Values are reported as the median (50th percentile) of the simulated posterior marginal probability distribution for each KIE along with the 68% highest posterior density interval (equivalent to ±1 SD about the median of a Gaussian distribution). Isotope effects computed from the gas-phase model system (13) at 300 K are shown in parentheses. (*C*) Site-specific KIEs for the SpnF-catalyzed cyclization with values reported as in *B*. The full simulated posterior marginal probability distributions of each KIE as well as all enrichment plots are shown in *SI Appendix*, Figs. S3 and S4.

of consideration for de novo protein design/reengineering to develop Diels–Alderases in the future.

Materials and Methods

Synthesis of Deuterated Compounds. Complete synthetic protocols and schemata (*SI Appendix*, Figs. S5–S7) for all deuterated compounds are provided in *SI Appendix*, *SI Text*.

Sample Preparation. The α-secondary deuterium KIE at each site of interest was measured using the method of internal competition (14) in at least three separate experimental trials. In each trial, site-specifically labeled and unlabeled SpnM substrates (**3**), which are stable, were mixed together (~1.5 mM total concentration) in 50 mM Tris buffer (pH 8.0) containing 10% DMSO at 30 °C before adding SpnM. Initial ratios of the isotopologs were generally close to 1 : 1, although they ranged from no less than 1 : 4 to no greater than 4 : 1. Reaction mixtures also included 0.47 mM *p*-methoxyacetophenone as an internal standard to standardize the observed HPLC peak integrations. The species **4** was generated in situ by the addition of SpnM to 16 μM, which converts all of the observable SpnM substrate (**3**) to SpnF substrate (**4**) within ~4 min. During this time, a by-product of the SpnM reaction with the same mass as **4** also forms; however, controls showed that, after its formation, this by-product is both stable and baseline-separated from **4** by HPLC.

No SpnF was present during measurement of nonenzymatic KIEs, and reaction aliquots were taken after all of the SpnM substrate had been consumed at six different time points ranging from ~10 to 90% total reaction (~90-min time course). Proportionately larger aliquots were taken at later time points to compensate for the reduced concentration of residual substrate. During measurement of enzymatic KIEs, SpnF was added after all of the SpnM substrate had been consumed and immediately before taking the initial quench. SpnF was added in sufficient quantity, so that the final quench at ~80% total reaction was taken at ~6 min after adding SpnF (rationale is in *SI Appendix*, *SI Text*). In all other respects, the nonenzymatic and enzymatic samples were treated identically.

On removal, aliquots were quenched by mixing 1 : 1 with MeCN containing 5% DMSO at 0 °C, which arrests both enzymatic and nonenzymatic cyclization. Protein was removed by centrifugation, and the residual substrate (**4**) was isolated using C18 reversed-phase HPLC (Varian Microsorb-MV; 100-5 C18 250 × 4.6 mm) with UV detection at 254 nm. The column was eluted with water (A) vs. MeCN (B) using the following gradient: start at 25% B, linear to 50% B in 35 min, linear to 70% B in 3 min, isocratic at 70% B for 2 min, and linear return to 25% B in 2 min. A solvent blank was run interposed between each HPLC separation to prevent memory effects. Isolated **4** was concentrated in a speed vacuum concentrator and redissolved in a 1 : 1 mixture of MeCN and water for MS analysis. Aliquots of the isolated samples were also reanalyzed by HPLC to ensure that they were free of all other reaction products and by-products. Fractions of reaction were determined from HPLC peak integrations as described in *SI Appendix* along with an assessment of error in variables (*SI Appendix*, *SI Text*). Total sample sizes and outlier criteria are provided in *SI Appendix*, *SI Text*.

Isotopolog Enrichments. Isolated residual substrate was analyzed by MS at the Laboratory for Biological Mass Spectrometry at Texas A&M University using a Waters SYNAPT G2 HDMS System in the positive ion high-sensitivity mode. Samples were first diluted using a 1:1 solution of MeOH:H_2O and injected at a flow rate of 0.5 μL/min with signal averaging over 1-s intervals. Three 3-min continuous direct infusions were run in series for each sample. The first two 3-min infusions were used to flush the instrument to prevent memory effects. Data from these flushes were discarded. Signal intensities obtained from the third 3-min acquisition were averaged at each mass over the ~175 spectra acquired to obtain a final spectrum for analysis. Control experiments did not show any evidence of ion fragmentation in the mass spectrometer, even at higher source ionization energies. All samples isolated in a given experimental trial were analyzed together without retuning of the instrument. Details regarding processing of MS spectra, correcting for natural abundance labeling, and an analysis of bias because of isotope effects from natural abundance labeling are provided in *SI Appendix*, *SI Text*. An example MS spectrum is shown in *SI Appendix*, Fig. S1.

Statistical Analysis. Each of the 12 measured KIEs was analyzed independently of the others using a hierarchical Bayesian regression model (33, 34). If n is the number of independent experimental trials for a given KIE, then the model involves $2n+3$ parameters. At the highest level is the KIE of interest denoted ${}^{D}k$, for which a log-normal prior was constructed with median 0 and variance 0.0325, which reparameterizes to

$$p\left({}^{D}k\right) = \frac{1}{{}^{D}k}\mathcal{N}\left(\log{}^{D}k|0, 0.0325\right), \quad {}^{D}k > 0. \qquad [3]$$

The distribution in Eq. **3** has a median of unity and ensures that the prior probability of ${}^{D}k$ lying on a positive interval $[a, b]$ is equal to the prior probability that it is on $[1/b, 1/a]$. This can be interpreted physically as treating positive and negative differences in the free energy of activation between two isotopogs equally. The variance of the log-normal distribution (0.0325) places $\sim$68% of the prior probability mass between KIEs of 0.83 and 1.20, consistent with previous reports of α-secondary deuterium KIEs on rehybridization reactions at 300 K (9, 10, 29, 35–38). The prior also allows for the possibility of large KIEs, consistent with previous estimates of theoretical maxima (11, 29, 39) on the intervals [0.70, 0.83] and [1.20, 1.43] with a combined probability of 25%. Extreme KIEs that exceed the predicted maxima have a prior probability less than 5% with this prior. A sensitivity analysis described in *SI Appendix, SI Text* shows that the posterior inferences for ${}^{D}k$ were insensitive to changes in the scale of this prior, unless it was increased so as to include extreme values with significant probability.

Between-experiment variation is modeled by characterizing the ith trial with its own observed isotope effect ${}^{D}k_i$ and initial enrichment ρ_i. The prior probability distribution of ρ_i is treated as uniform on (0, 10), consistent with the experimental design, and ${}^{D}k_i$ is modeled as dependent on ${}^{D}k$ according to

$${}^{D}k_i = {}^{D}k + \eta_i, \qquad [4]$$

where the error term η_i is a truncated Gaussian random variable with scale σ_b and mode 0. The prior distributions for ${}^{D}k_i$ (conditional on ${}^{D}k$ and σ_b) and ρ_i are then given by

$$p\left({}^{D}k_i|{}^{D}k, \sigma_b\right) = \mathcal{N}\left({}^{D}k_i|{}^{D}k, \sigma_b^2\right), \quad {}^{D}k_i > 0, \qquad [5a]$$

$$p(\rho_i) = \mathcal{U}(0, 10). \qquad [5b]$$

Based on previous recommendations for hierarchical analyses with small numbers of groups (40), a half-Cauchy distribution on $(0, \infty)$ with scale 0.2 was used for $p(\sigma_b)$, which admits the possibility for large variation between experiments with significant probability. Sensitivity analysis indicated that doubling the scale from 0.2 to 0.4 had no meaningful impact on the posterior marginal probability intervals for ${}^{D}k$. Finally, within-experiment variation is modeled as the SD about the regression, denoted σ_w, and the prior distribution for σ_w was uniform on support (0, 1). The joint prior distribution is then given by

$$p\left({}^{D}k, \sigma_b, \sigma_w, {}^{D}\mathbf{k}, \boldsymbol{\rho}\right) = p\left({}^{D}k\right)p\left(\sigma_b\right)p\left(\sigma_w\right)\prod_{i=1}^{n}p\left({}^{D}k_i|{}^{D}k, \sigma_b\right)p(\rho_i), \qquad [6]$$

where ${}^{D}\mathbf{k}$ denotes the n vector of ${}^{D}k_i$ parameters, and $\boldsymbol{\rho}$ denotes the n vector of ρ_i parameters.

The likelihood function is equal to the probability of the observed data conditional on the parameters. The likelihood function was thus constructed by modeling the jth enrichment from the ith experimental trial as

$$r_{ij} = R\left(f_{ij}; {}^{D}k_i, \rho_i\right) + \varepsilon_{ij}, \qquad [7]$$

where the function R is given by Eq. **1**, f_{ij} is the jth fraction of reaction in the ith trial, and ε_{ij} is the jth ith realization of the random Gaussian variable ε with mean zero and variance σ_w^2. Therefore, if $\mathbf{f}_i$ and $\mathbf{r}_i$ are vectors of the n_i fractions of reaction and enrichment measurements in the ith experimental trial, respectively, then the likelihood function based on the ith experimental trial is given by

$$p\left(\mathbf{r}_i|{}^{D}k, \sigma_b, \sigma_w, {}^{D}k_i, \rho_i, \mathbf{f}_i\right) = \prod_{j=1}^{n_i}\mathcal{N}\left(r_{ij}|R\left(f_{ij}; {}^{D}k_i, \rho_i\right), \sigma_w^2\right). \qquad [8]$$

The likelihood function over all independent experimental trials is then just the product

$$p\left(\mathbf{r}|{}^{D}k, \sigma_b, \sigma_w, {}^{D}\mathbf{k}, \boldsymbol{\rho}, \mathbf{f}\right) = \prod_{i=1}^{n}p\left(\mathbf{r}_i|{}^{D}k, \sigma_b, \sigma_w, {}^{D}k_i, \rho_i, \mathbf{f}_i\right), \qquad [9]$$

where $\mathbf{f}$ and $\mathbf{r}$ are the fractions of reaction and enrichment vectors over all experimental trials. The posterior joint distribution of the parameters conditional on the observations is given by Bayes' Theorem and is proportional to the product of the prior joint distribution in Eq. **6** and likelihood function in Eq. **9**: that is,

$$p\left({}^{D}k, \sigma_b, \sigma_w, {}^{D}\mathbf{k}, \boldsymbol{\rho}|\mathbf{f}, \mathbf{r}\right) \propto p\left(\mathbf{r}|{}^{D}k, \sigma_b, \sigma_w, {}^{D}\mathbf{k}, \boldsymbol{\rho}, \mathbf{f}\right)p\left({}^{D}k, \sigma_b, \sigma_w, {}^{D}\mathbf{k}, \boldsymbol{\rho}\right), \qquad [10]$$

because the distribution of $\mathbf{f}$ is assumed to provide no information regarding the distribution of $\mathbf{r}$ given $\mathbf{f}$ (33).

MCMC (33, 34, 41, 42) was used to sample from the posterior joint distribution in Eq. **10** using the affine-invariant ensemble sampling algorithm proposed by Goodman and Weare (21) and implemented in Python as the emcee package (22). The emcee sampler uses an ensemble of walkers to generate separate MCMC chains. A total of 200 walkers were used for all experiments except the C7D enzymatic KIE, where the number was 246. Starting positions for the walkers were based on a Gaussian ball about the maximum likelihood estimate of the parameters under a linear approximation. Each walker was iterated 5,000 times after an initial equilibration period (burn-in) of 1,000 iterations, which were discarded. This produced a total sample size of at least 1×10^6 points, which were used without thinning to construct the simulated posterior joint distribution. Effective sample sizes were calculated for each parameter from the autocorrelation times (22, 42) and found to be at least 10,000 for each parameter (*SI Appendix*, Tables S2 and S3), which is consistent with recommendations for interval estimation (34). Potential scale reduction factors (43) were calculated for each parameter and found to be less than 1.06 in all cases (*SI Appendix*, Tables S2 and S3), consistent with equilibration. In all cases, the mean acceptance fraction was between 0.3 and 0.5, indicating that the walkers were neither becoming trapped nor representative of a random walk (22). Visual inspection of randomly selected walker chains (*SI Appendix*, Fig. S2) indicated the absence of any apparent drift or trapping, which is in line with the quantitative evaluations. *SI Appendix*, Table S2 provides summary statistics for all parameters based on the simulated posterior joint distributions obtained in the analysis of each nonenzymatic KIE. *SI Appendix*, Table S3 does the same for each enzymatic KIE. Simulated posterior marginal distributions for each KIE (i.e., the ${}^{D}k$ parameter) as well as sampled curves from the posterior joint distribution are shown in *SI Appendix*, Figs. S3 and S4.

Computed KIEs. Computational methods used to calculate KIEs are described in *SI Appendix, SI Text*.

ACKNOWLEDGMENTS. We thank Ken Houk and Dan Singleton for helpful discussions during preparation of this manuscript. This work was supported by NIH Grant R01 GM040541 and Welch Foundation Grant F-1511.

1. Kim HJ, Ruszczycky MW, Choi Sh, Liu Yn, Liu Hw (2011) Enzyme-catalysed [4+2] cycloaddition is a key step in the biosynthesis of spinosyn A. *Nature* 473:109–112.
2. Kim HJ, et al. (2014) Chemoenzymatic synthesis of spinosyn A. *Angew Chem Int Ed* 53:13553–13557.
3. Jeon BS, Wang SA, Ruszczycky MW, Liu Hw (2017) Natural [4 + 2]-cyclases. *Chem Rev* 117:5367–5388.
4. Hashimoto T, et al. (2015) Biosynthesis of versipelostatin: Identification of an enzyme-catalyzed [4 + 2]-cycloaddition required for macrocyclization of spirotetronate-containing polyketides. *J Am Chem Soc* 137:572–575.
5. Tian Z, et al. (2015) An enzymatic [4 + 2] cyclization cascade creates the pentacyclic core of pyrroindomycins. *Nat Chem Biol* 11:259–265.
6. Wever WJ, et al. (2015) Chemoenzymatic synthesis of thiazolyl peptide natural products featuring an enzyme-catalyzed formal [4 + 2] cycloaddition. *J Am Chem Soc* 137:3494–3497.
7. Hudson GA, Zhang Z, Tietz JI, Mitchell DA, van der Donk WA (2015) In Vitro biosynthesis of the core scaffold of the thiopeptide thiomuracin. *J Am Chem Soc* 137:16012–16015.
8. Byrne MJ, et al. (2016) The catalytic mechanism of a natural Diels-Alderase revealed in molecular detail. *J Am Chem Soc* 138:6095–6098.
9. Storer JW, Raimondi L, Houk KN (1994) Theoretical secondary kinetic isotope effects and the interpretation of transition state geometries. 2. The Diels-Alder reaction transition state geometry. *J Am Chem Soc* 116:9675–9683.
10. Beno BR, Houk KN, Singleton DA (1996) Synchronous or asynchronous? An "experimental" transition state from a direct comparison of experimental and theoretical kinetic isotope effects for a Diels-Alder reaction. *J Am Chem Soc* 118:9984–9985.
11. Henderson RW, Pryor WA (1972) Secondary hydrogen isotope effect for the transition from olefin to free radical. *Int J Chem Kinet* 4:325–330.
12. Patel A, et al. (2016) Dynamically complex [6 + 4] and [4 + 2] cycloadditions in the biosynthesis of spinosyn A. *J Am Chem Soc* 138:3631–3634.

13. Hess BA Jr, Smentek L (2012) Concerted, highly asynchronous, enzyme-catalyzed [4 + 2] cycloaddition in the biosynthesis of spinosyn A; computational evidence. *Org Biomol Chem* 10:7503–7509.
14. Cleland WW (1982) Use of isotope effects to elucidate enzyme mechanisms. *Crit Rev Biochem* 13:385–429.
15. Cassano AG, et al. (2007) Inaccuracies in selected ion monitoring detemination of isotope ratios obviated by profile acquisition: Nucleotide $^{18}O/^{16}O$ measurements. *Anal Biochem* 367:28–39.
16. Harris ME, et al. (2010) Kinetic isotope effects for RNA cleavage by 2′-O-transphosphorylation: Nucleophilic activation by specific base. *J Am Chem Soc* 132:11613–11621.
17. Berti PJ, Blanke SR, Schramm VL (1997) Transition state structure for the hydrolysis of NAD^+ catalyzed by diphtheria toxin. *J Am Chem Soc* 119:12079–12088.
18. Kim HJ, Pongdee R, Wu Q, Hong L, Liu Hw (2007) The biosynthesis of spinosyn in Saccharopolyspora spinosa: Synthesis of the cross-bridging precursor and identification of the function of SpnJ. *J Am Chem Soc* 129:14582–14584.
19. Wang B, et al. (2007) Isotopologue distributions of peptide product ions by tandem mass spectrometry: Quantitation of low levels of deuterium incorporation. *Anal Biochem* 367:40–48.
20. Bigeleisen J, Wolfsberg M (1958) Theoretical and experimental aspects of isotope effects in chemical kinetics. *Adv Phys Chem* 1:15–76.
21. Goodman J, Weare J (2010) Ensemble samplers with affine invariance. *Commun Appl Math Comput Sci* 5:65–80.
22. Foreman-Mackey D, Hogg DW, Lang D, Goodman J (2013) emcee: The MCMC hammer. *Publ Astron Soc Pac* 125:306–312.
23. Stein RL (1981) Analysis of kinetic iosotpe effects on complex reactions utilizing the concept of the virtual transition state. *J Org Chem* 46:3328–3330.
24. Ray WJ Jr (1983) Rate-limiting steps: A quantitative definition. Application to steady-state enzymic reactions. *Biochemistry* 22:4625–4637.
25. Ruszczycky MW, Anderson VE (2006) Interpretation of V/K isotope effects for enzymatic reactions exhibiting multiple isotopically sensitive steps. *J Theor Biol* 243: 328–342.
26. Hout RF, Levi BA, Hehre WJ (1983) A method for the calculation of normal-mode vibrational frequencies using symmetry coordinates. Application to the calculation of secondary deuterium isotope effects on carbocations. *J Comp Chem* 4:499–505.
27. Medvedev MG, et al. (2017) Quantifying possible routes for SpnF-catalyzed formal Diels-Alder cycloaddition. *J Am Chem Soc* 139:3942–3945.
28. Gordeev EG, Ananikov VP (2015) Computational study of a model system of enzyme-mediated [4 + 2] cycloaddition reaction. *PLoS One* 10:e0119984.
29. Gajewski JJ, Peterson KB, Kagel JR, Huang YCJ (1989) Transition-state structure variation in the Diels-Alder reaction from secondary deuterium kinetic isotope effects. The reaction of nearly symmetrical dienes and dienophiles is nearly synchronous. *J Am Chem Soc* 111:9078–9081.
30. Fage CD, et al. (2015) The structure of SpnF, a standalone enzyme that catalyzes [4+2] cycloaddition. *Nat Chem Biol* 11:256–258.
31. Zheng Q, et al. (2016) Enzyme-dependent [4 + 2] cycloaddition depends on lid-like interaction of the *N*-terminal sequence with the catalytic core in PyrI4. *Cell Chem Biol* 23:352–360.
32. Schramm VL (2007) Binding isotope effects: Boon and bane. *Curr Opin Chem Biol* 11:529–536.
33. Gelman A, et al. (2014) *Bayesian Data Analysis* (CRC, Boca Raton, FL), 3rd Ed.
34. Kruschke JK (2015) *Doing Bayesian Data Analysis* (Elsevier, London), 2nd Ed.
35. Singleton DA, Thomas AA (1995) High-precision simultaneous determination of multiple small kinetic isotope effects at natural abundance. *J Am Chem Soc* 117:9357–9358.
36. Gajewski JJ, Peterson KB, Kagel JR (1987) Transition-state structure variation in the Diels-Alder reaction from secondary deuterium kinetic isotope effects: The reaction of a nearly symmetrical diene and dienophile is nearly synchronous. *J Am Chem Soc* 109:5545–5546.
37. Seltzer S (1965) The mechanism of the Diels-Alder reaction of 2-methylfuran with maleic anhydride. *J Am Chem Soc* 87:1534–1540.
38. Van Sickle DE, Rodin JO (1964) The secondary deuterium isotope effect on the Diels-Alder reaction. *J Am Chem Soc* 86:3091–3094.
39. Streitwieser A Jr, Jagow RH, Fahey RC, Suzuki S (1958) Kinetic isotope effects in the acetolyses of deuterated cyclopentyl tosylates. *J Am Chem Soc* 80:2326–2332.
40. Gelman A (2006) Prior distributions for variance parameters in hierarchical models. *Bayesian Anal* 1:515–533.
41. MacKay DJC (2003) *Information Theory, Inference, and Learning Algorithms* (Cambridge Univ Press, Cambridge, UK).
42. Kass RE, Carlin BP, Gelman A, Neal RM (1998) Markov chain Monte Carlo: A roundtable discussion. *Am Stat* 52:93–100.
43. Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. *Stat Sci* 7:457–472.