

Data-Driven Identification of Metal-free Orthogonal Bioorthogonal Cycloaddition Pairs

Joao M. J. M. Ravasco^{¶,†,‡}, Jaime A. S. Coelho^{§,‡,*}

[¶] Research Institute for Medicines, Faculty of Pharmacy, University of Lisbon, 1649-003 Lisboa, Portugal

[§] Centro de Química Estrutural, Institute of Molecular Sciences, Faculty of Sciences, University of Lisbon, 1749-016 Lisboa, Portugal

ABSTRACT: Mutually orthogonal bioorthogonal reactions are a thriving area of research in chemical biology. Here we present a predictive-driven approach to identify mutually orthogonal pairs among two bioorthogonal reactions: the metal-free 1,3-dipolar cycloaddition (1,3-DC) reaction and the inverse electron-demand Diels–Alder (IEDDA) reaction. Parametrization of both 1,3-dipoles and dipolarophiles structures allowed the development of statistically robust models for predicting the second order rate constants of 1,3-DC reactions. Combination of predictive models were used to identify potential mutually orthogonal reactions among sets of structurally different pairs.

INTRODUCTION

Bioorthogonal reactions are widely used for monitoring and controlling biological functions through labelling or decaging of biomolecules and payloads of interest.^{1–3} Advances in the bioorthogonal toolbox have rendered these reactions more effective, selective and widespread, enabling the construction of innovative theranostic and delivery systems for in vivo applications.^{4–6} Recently, the technology reached the milestone of first-in-human usage with a trans-cyclooctene/tetrazine (TCO/Tz) prodrug activation system recently entering Phase I clinical trials (ClinicalTrials.gov, Identifier: NCT04106492).^{5–7} A current theme in bioorthogonal chemistry is the development of mutually orthogonal reactions towards deepening the understanding of biological systems.^{8,9} This has been achieved by empirically identifying suitable pairs among reported mechanistically and kinetically distinct reactions with reaction rates varying by several orders of magnitude.¹⁰ Examples include the orthogonality between most polar reactions with cycloaddition reactions^{11–13}; but also amongst cycloadditions, such as the inverse electron-demand Diels–Alder reaction (IEDDA) using a TCO/Tz pair with the copper-, or strain-promoted azide-cyclooctyne cycloaddition.¹⁴ More demanding triple strategies combining such reactivities have also been gaining momentum on expanding the current paradigm on multiplexed biological approaches.^{15–19}

Prediction of the reaction kinetics has been a cornerstone of the advances made in bioorthogonal chemistry. The use of DFT calculations together with the distortion/interaction analysis has been used to determine the activation energies and consequently the rate constants of bioorthogonal reactions.^{20–23} In multiplexed applications, kinetic divergence among bioorthogonal pairs is a widely employed strategy. By using such computations, Houk and co-workers have shown the advantage of using heatmaps (i.e., a matrix) of calculated rate constants for identifying new mutually

orthogonal reaction pairs.²⁰ However, the need to locate transition state geometries renders this approach time consuming. For instance, a 10×10 matrix, which displays 100 reactions (e.g., IEDDA involving 10 dienes and 10 dienophiles), would require optimizing 100 transition state structures. To cover a broader chemical space of bioorthogonal compounds in a more convenient way, we envisioned the use of statistical data analysis as alternative to predict rate constants. We have recently shown that multivariate regression analysis can be used to model and predict the kinetics of IEDDA reactions using low computation cost by parametrizing both diene and dienophile reagents.²⁴ This approach would require the parametrization of only $10 + 10$ reagents for effectively generate a 10×10 matrix of 100 reactions. Metal-free 1,3-dipolar cycloaddition reactions (1,3-DC) have contributed together with IEDDA to the widespread success of bioorthogonal chemistry for therapeutic and bioimaging purposes, and have been explored in multiplexed approaches as kinetically disparate pairs.

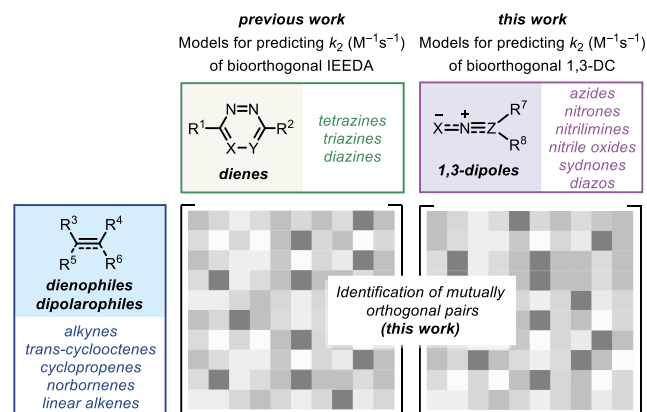


Figure 1. Matrices of predicted rate constants for bioorthogonal IEDDA and 1,3-DC reactions.

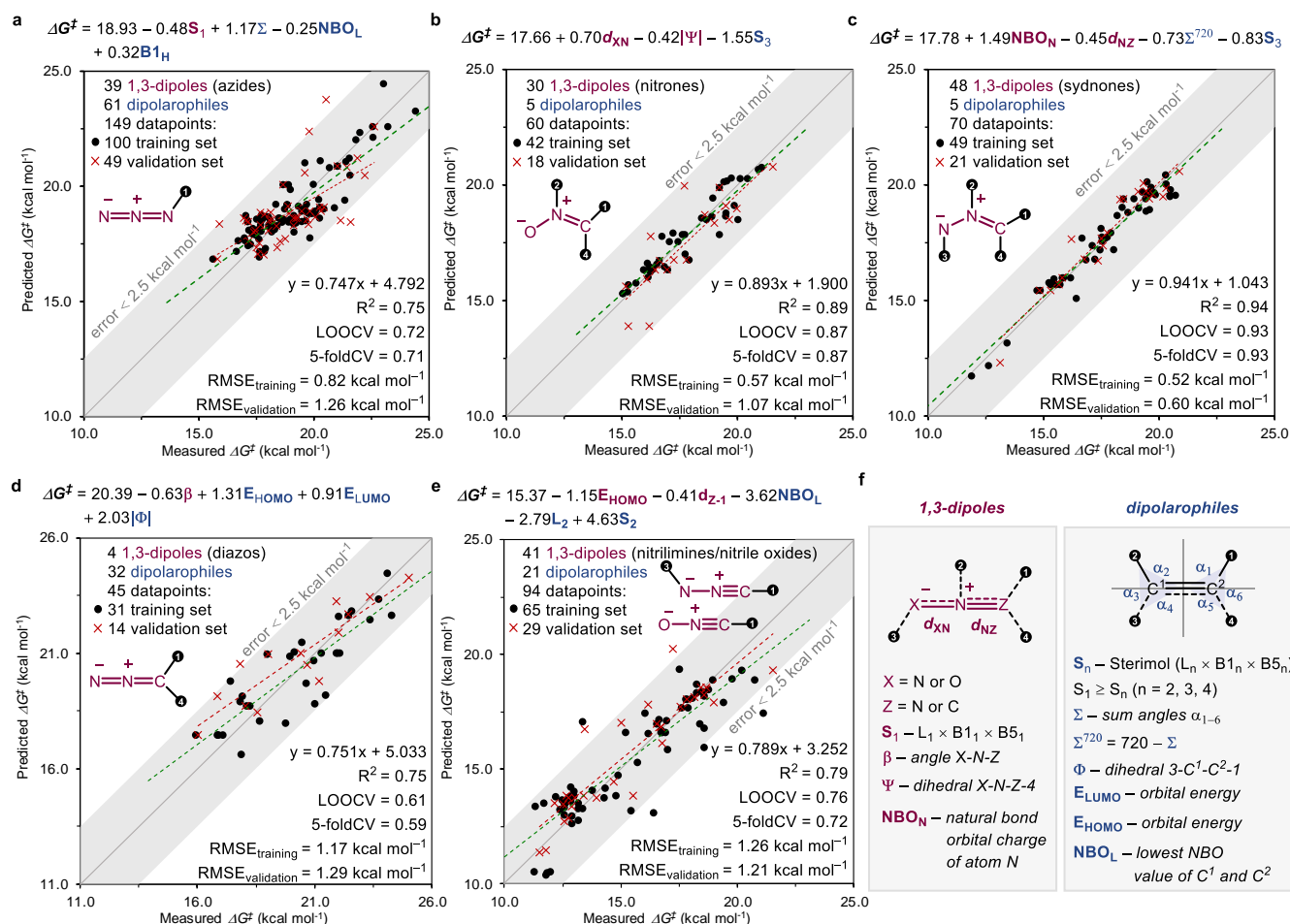


Figure 2. Multivariate models for 1,3-DC reactions involving (a) azides, (b) nitrones, (c) sydnones, (d) diazos, and (e) nitrilimines and nitrile oxides. (f) Parameters related to 1,3-dipoles and dipolarophiles.

Herein we modeled the 1,3-DC as the bioorthogonal reaction counterpart of the previously modeled IEDDA reaction and evaluated the benefit of using statistical data analysis and integrated models in the identification of new mutually orthogonal pairs (Figure 1).

RESULTS AND DISCUSSION

We started our study by modeling the bioorthogonal 1,3-DC reaction. Literature curation resulted in the collection of experimental rate constants of 400 reactions involving 168 structurally different 1,3-dipoles and 88 dipolarophiles (free energy range of nearly 15 kcal mol⁻¹, see SI). These reagents were parametrized using molecular descriptors determined after structure optimization using DFT calculations (M06-2X/6-31G(d) level of theory). Calculated descriptors included geometric, steric and electronic features such as Sterimol parameters (L , B_1 and B_5), distances (d), bond angles and dihedral angles, vibration frequencies and intensities, and NBO charges and HOMO/LUMO energies. For the parametrization of the 1,3-dipoles we oriented the structures from left-to-right according to the order XNZ and with the highest Sterimol coefficient of the Z substituents placed on the top (Figure 2f), while the dipolarophiles structures where oriented according to our previous report (i.e., highest Sterimol coefficient placed to the top right quadrant)²⁴.

We created five data sets according to the 1,3-dipole structure (Figure 2): azides (149 reactions, 39 1,3-dipoles, 61 dipolarophiles), nitrones (60 reactions, 30 1,3-dipoles, 5 dipolarophiles), sydnones (70 reactions, 48 1,3-dipoles, 5 dipolarophiles), diazos (45 reactions, 4 1,3-dipoles, 32 dipolarophiles), nitrilimines and nitrile oxides (94 reactions, 41 1,3-dipoles, 21 dipolarophiles).

Highly predictive models were identified for all five individual data sets with a goodness-of-fit R^2 values in the range of 0.74 to 0.94. Furthermore, cross validation analysis (LOO, $Q^2 = 0.61$ –0.93 and k-fold, $Q^2 = 0.59$ –0.93) and external validation analysis (data set partitioned into 7:3 training/validation sets, $RMSE_{\text{training}} = 0.52$ –1.30 kcal mol⁻¹, $RMSE_{\text{validation}} = 0.60$ –1.29 kcal mol⁻¹) suggest robust models. These models include terms relative to 1,3-dipoles and dipolarophiles. The descriptors used to describe 1,3-dipoles across the five models are the geometric parameter d_{XN} , d_{NZ} , d_{Z-1} , β , Ψ , Sterimol S_1 , NBO charge NBO_N , and HOMO energy (Figure 1f). For dipolarophiles, the descriptors included in these models are the geometric parameters Σ , Σ^{720} and Φ , Sterimol parameters B_1 , L_2 , S_3 and S_2 , NBO charge NBO_L , and HOMO and LUMO energies (Figure 1f). None of the parameters describing the structure of 1,3-dipoles are repeated across the different datasets, which is not surprising taking into consideration that each dataset corresponds to a unique class of 1,3-dipoles.

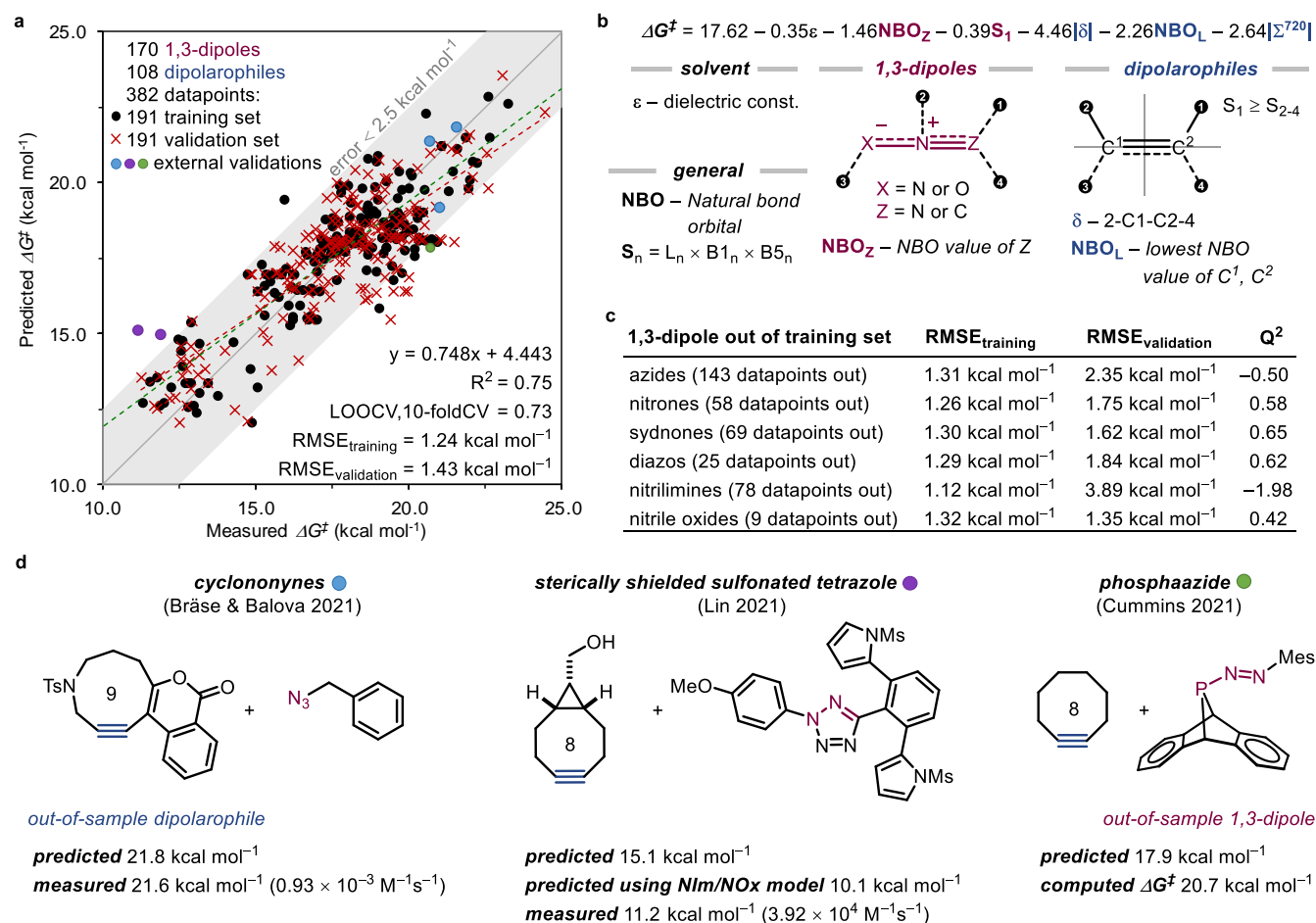


Figure 3. (a) Comprehensive model for the 1,3-DC reaction. (b) model and parameters (c) LODO (leave one-class-of-dipole out) analysis (d) Selected external validations using out-of-sample dipolarophiles or 1,3-dipoles. Ts – toluenesulfonyl or tosyl, Ms – mesyl or methanesulfonyl, Mes – mesityl.

Conversely, as several dipolarophiles structures are repeated across the different datasets, NBO_L , Σ/Σ^{720} and S_3 are repeated twice across the five models. Despite the fact that, in general, reaction rates are affected by solvent, we observed that dielectric constant of the solvent was not statistically relevant for these individual models, which could be attributed to the low diversity of solvents in each individual subset. All the parameters used in these models are chemically comprehensible, nevertheless the most complex parameter is Σ^{720} , which describes the structure of dipolarophiles. This descriptor is the difference between the sum of the bend angles of a perfectly planar structure (i.e., 720, resembling ethylene) and the sum of the bend angles of the dipolarophile core (Σ), thus describing how far from planar the dipolarophile structures are. For instance, in the sydnones model, this parameter is inversely correlated to ΔG^\ddagger , suggesting faster reaction rates for distorted dipolarophiles. Furthermore, Σ^{720} in combination with NBO_L can be used to represent the chemical space of dipolarophiles including alkynes and alkenes both cyclic and acyclic (see SI).

Comprehensive model for 1,3-DC reactions

The robustness of these individual models suggested that a comprehensive model for 1,3-DC reactions could be developed. To this end, we built a data set consisting of 382 1,3-DC reactions employed in biological systems, which

included all six classes of 1,3-dipoles. By applying the same methodology as before, a highly predictive model was obtained with a goodness-of-fit R^2 value of 0.75 (Figure 3). Cross validation analysis (LOO, $Q^2 = 0.73$ and k-fold, $Q^2 = 0.73$) and external validation analysis (data set portioned into 1:1 training/validation sets, $\text{RMSE}_{\text{training}} = 1.25 \text{ kcal mol}^{-1}$, $\text{RMSE}_{\text{validation}} = 1.43 \text{ kcal mol}^{-1}$) suggest a robust model. Moreover, we did not identify any outliers. This comprehensive model includes two parameters related to 1,3-dipoles: NBO_Z and S_1 , three parameters related to dipolarophiles: δ , NBO_L and $|\Sigma^{720}|$, and the dielectric constant of the solvent (ε). In other words, reaction rate of 1,3-DC increases for polar solvents (high ε), for 1,3-dipoles with high NBO charge at atom Z and sterically shielded (high Sterimol parameter of the Z-substituent), and for dipolarophiles with high NBO charges and far from planar (high $|\delta|$ and $|\Sigma^{720}|$). Not surprisingly, the previously highlighted descriptors NBO_L and $|\Sigma^{720}|$ are statistically relevant for the comprehensive model. The presence of the dielectric constant of the solvent contrasts with the individual models, which could be explained by the greater variability of the solvent in the full dataset. Interestingly, neither HOMO nor LUMO energies are used to parametrize 1,3-dipole and dipolarophiles structures in this comprehensive model.

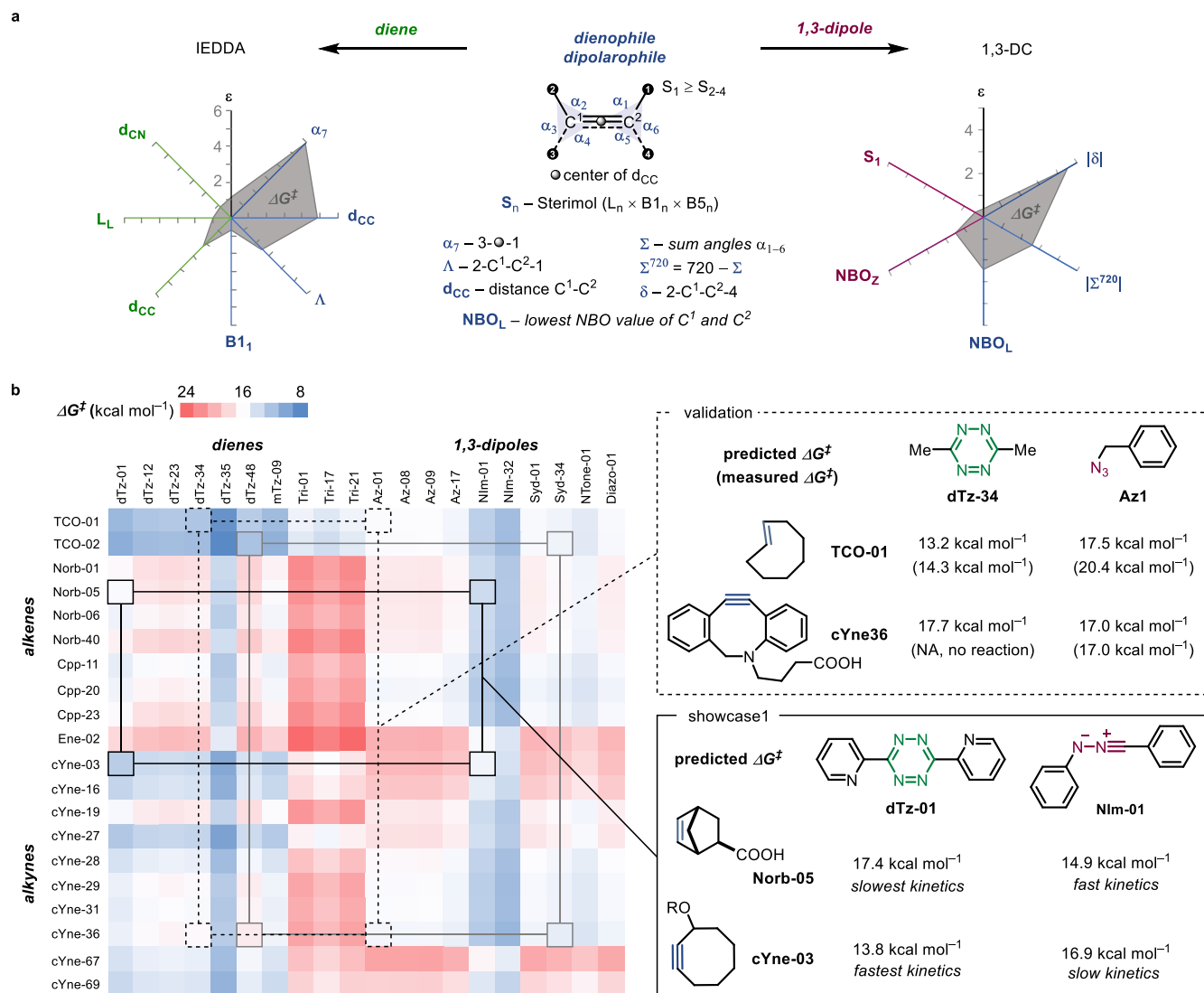


Figure 4. (a) Representation of the terms for the comprehensive models for the IEDDA and 1,3-DC reaction. (b) Predictive matrix of 10 dienes, 10 1,3-dipoles, 10 alkenes and 10 alkynes structures. For complete list of structures and values see Supporting Information.

To demonstrate the ability of the model to predict out-of-sample structures, we performed 1) a leave-one-class-of-dipoles out cross validation analysis (LODO) and 2) external validations by predicting several reaction rates recently reported in the literature. In LODO analysis, the full dataset is split into different folds according to the class of 1,3-dipole structure and for each class the model is trained with all folds except one that is used for validation. This analysis shows that in general the model is able to predict out-of-sample structures (Figure 3c, $RMSE_{\text{validation}} \leq 2.35$ kcal mol⁻¹, for azides, nitrones, sydrones, diazos and nitrile oxides), with a higher error found for the prediction of the k_2 of reactions involving nitrilimines ($RMSE_{\text{validation}} = 3.89$ kcal mol⁻¹). It is worth mentioning that the low LODO_{azides} arises from the fact that a large data of dipolarophiles is removed when azides-containing reactions are not included in the training set (see SI for analysis of the chemical space of each dataset). Finally, we tested the comprehensive model using three recently reported case studies. In the first case, the model was able to predict with high precision the reaction rates of the cycloadditions of different

out-of-sample cyclononynes with benzyl azide (e.g., 21.8 vs 21.6 kcal mol⁻¹, Figure 3d).²⁵ In the second case, the prediction of the reaction rate involving sterically shielded sulfonated nitrilimines²⁶ resulted in a greater error (e.g., 15.1 vs 11.2 kcal mol⁻¹, Figure 3d), which was not surprising due to the low LODO_{nitrilimines}. In fact, we could increase the precision of this prediction by acknowledging that this is not an out-of-sample structure and using the individual model developed for the subset of nitrilimines (10.1 vs 11.2 kcal mol⁻¹, Figure 3d). Nevertheless, it is worth mentioning that this is an overextrapolation considering the datapoints comprising nitrilimine pairs (i.e., this is the fastest reaction rate involving nitrilimines), and thus validating the high predictive skill of the model. In the third case, the prediction of a more challenging out-of-sample 1,3-dipole phosphazide²⁷ resulted in a 2.8 kcal mol⁻¹ error compared to the reported DFT calculated transition state energy (17.9 vs 20.7 kcal mol⁻¹, experimental reaction rate not reported, Figure 3d).

Identification of mutually orthogonal bioorthogonal reactions

With access to two statistically robust models composed of independent parameters to describe the common pair dienophile/dipolarophile of the two cycloadditions (Figure 4a), we questioned whether mutually orthogonal bioorthogonal reactions could be identified. As proof-of-concept, we selected structurally diverse reactive pairs comprised of 10 dienes (6-disubstituted tetrazines, dTz; 1 monosubstituted tetrazine, mTz; and 3 triazines, Tri), 10 1,3-dipoles (4 azides, Az; 2 nitrilimines, NIm; 2 sydnone, Syd; 1 nitron, NTone; and 1 diazo, Diazo), and 20 dienophiles/dipolarophiles (10 alkenes, TCO, Norb, Cpp, Ene; and 10 cyclic alkynes, cYne), and created a matrix of the respective predicted ΔG^\ddagger values (Figure 4b). Identification of mutually orthogonal pairs among the selected reactants could be achieved by detecting ΔG^\ddagger values that differ by several units. To validate this approach, we selected a mutually orthogonal cycloaddition described in the literature and evaluated the skill of the matrix to predict such reactivity. Remarkably, taking in consideration the corresponding RMSE, the predictive matrix was able to describe the mutual orthogonality involving the TCO/cyclooctyne/Tz/azide combination reported by Hilderbrand and co-workers²⁸ (Figure 4b, dashed lines). Finally, several showcases of unreported potential mutually orthogonal bioorthogonal cycloaddition were identified involving: 1) norbornene/cyclooctyne/Tz/nitrilimine (Figure 4b, solid black line), 2) TCO/cyclooctyne/Tz/sydnone (Figure 4b, solid grey line), and 3) TCO/cyclooctyne/Tz/Diazo (not shown, See SI).

CONCLUSION

In this study, we parametrized the structure of 1,3-dipoles and dipolarophiles and developed predictive multivariate models for the determination of second order rate constants of 1,3-dipolar cycloadditions. Analysis of the developed comprehensive model revealed that i) reaction rates are greatly affected by manipulation of dipolarophiles and less by 1,3-dipoles (higher normalized coefficients for dipolarophiles terms); ii) ring strain poses as the most important descriptor for dipolarophiles as depicted by high coefficients for $|\Phi|$ and $|\Sigma^{720}|$ terms; iii) 1,3-dipoles structures could be effectively tuned by varying the NBO charge and size of substituent of **Z** (**NBO_Z** and **S_I**).

The combination of two comprehensive models for the bioorthogonal IEDDA reaction and 1,3-DC was used to generate matrixes of predicted reaction rates for several bioorthogonal cycloadditions. This workflow demonstrates the ability of a predictive data-driven approach for identifying mutually orthogonal bioorthogonal cycloadditions. We believe that statistical modeling can be combined with decision analysis to improve the effectiveness of the approach for further identifying novel mutually orthogonal reactions—either to account for information of reactions that do not occur (reaction rate not available) or to use a sub model depending on the reagent structure in oppose to the comprehensive model. Furthermore, we envision the use of this approach for fine-tuning a pair in a specific reaction, being crucial the quality of the training set and model choice for the performance.

AUTHOR INFORMATION

Corresponding Author

*jaimeacoelho@campus.ul.pt

Present Addresses

†NOVA Medical School, Faculdade de Ciências Médicas, Universidade Nova de Lisboa, 1169-056 Lisboa, Portugal

Author Contributions

*J.M.J.M.R and J.A.S.C. contributed equally. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

Financial support from Fundação para a Ciência e a Tecnologia (SFRH/BD/120829/2016, UIDB/00100/2020, UIDP/00100/2020, LA/P/0056/2020, UIDB/04138/2020; UIDP/04138/2020) is gratefully acknowledged. J.A.S.C. thanks the Fundação para a Ciência e a Tecnologia (FCT) for Scientific Employment Stimulus 2020/02383/CEECIND. We acknowledge Dr. João N. Rosa for helpful discussions.

REFERENCES

1. Patterson, D. M.; Nazarova, L. A.; Prescher, J. A., Finding the right (bioorthogonal) chemistry. *ACS Chem. Biol.* **2014**, *9* (3), 592-605.
2. Sletten, E. M.; Bertozzi, C. R., Bioorthogonal Chemistry: Fishing for Selectivity in a Sea of Functionality. *Angew. Chem. Int. Ed.* **2009**, *48* (38), 6974-6998.
3. McKay, C. S.; Finn, M. G., Click chemistry in complex mixtures: bioorthogonal bioconjugation. *Chem. Biol.* **2014**, *21* (9), 1075-101.
4. Devaraj, N. K., The Future of Bioorthogonal Chemistry. *ACS Cent. Sci.* **2018**, *4* (8), 952-959.
5. Kumar, G. S.; Lin, Q., Light-Triggered Click Chemistry. *Chem. Rev.* **2021**, *121* (12), 6991-7031.
6. Bilodeau, D. A.; Margison, K. D.; Serhan, M.; Pezacki, J. P., Bioorthogonal Reactions Utilizing Nitrones as Versatile Dipoles in Cycloaddition Reactions. *Chem. Rev.* **2021**, *121* (12), 6699-6717.
7. Srinivasan, S.; Yee, N. A.; Wu, K.; Zakharian, M.; Mahmoodi, A.; Royzen, M.; Oneto, J. M. M., SQ3370 Activates Cytotoxic Drug via Click Chemistry at Tumor and Elicits Sustained Responses in Injected and Non-Injected Lesions. *Adv. Therap.* **2021**, *4* (3), 2000243.
8. Smeenk, M. L. W. J.; Agramunt, J.; Bongers, K. M., Recent developments in bioorthogonal chemistry and the orthogonality within. *Curr. Opin. Chem. Biol.* **2021**, *60*, 79-88.
9. Hu, Y.; Schomaker, J. M., Recent Developments and Strategies for Mutually Orthogonal Bioorthogonal Reactions. *ChemBioChem* **2021**, *22* (23), 3254-3262.
10. Narayanam, M. K.; Liang, Y.; Houk, K. N.; Murphy, J. M., Discovery of new mutually orthogonal bioorthogonal cycloaddition pairs through computational screening. *Chem. Sci.* **2016**, *7* (2), 1257-1261.
11. Lu, H.; Zhou, Q.; Deshmukh, V.; Phull, H.; Ma, J.; Tardif, V.; Naik, R. R.; Bouvard, C.; Zhang, Y.; Choi, S.; Lawson, B. R.; Zhu, S.; Kim, C. H.; Schultz, P. G., Targeting Human C-Type Lectin-like Molecule-1 (CLL1) with a Bispecific Antibody for Immunotherapy of Acute Myeloid Leukemia. *Angew. Chem. Int. Ed.* **2014**, *53* (37), 9841-9845.
12. Kim, C. H.; Axup, J. Y.; Dubrovskaya, A.; Kazane, S. A.; Hutchins, B. A.; Wold, E. D.; Smider, V. V.; Schultz, P. G., Synthesis of

Bispecific Antibodies using Genetically Encoded Unnatural Amino Acids. *J. Am. Chem. Soc.* **2012**, *134* (24), 9918-9921.

13. Hudak, J. E.; Barfield, R. M.; de Hart, G. W.; Grob, P.; Nogales, E.; Bertozzi, C. R.; Rabuka, D., Synthesis of Heterobifunctional Protein Fusions Using Copper-Free Click Chemistry and the Aldehyde Tag. *Angew. Chem. Int. Ed.* **2012**, *51* (17), 4161-4165.

14. Karver, M. R.; Weissleder, R.; Hilderbrand, S. A., Bioorthogonal Reaction Pairs Enable Simultaneous, Selective, Multi-Target Imaging. *Angew. Chem. Int. Ed.* **2012**, *51* (4), 920-922.

15. Winz, M. L.; Linder, E. C.; Becker, J.; Jaschke, A., Site-specific one-pot triple click labeling for DNA and RNA. *Chem. Commun.* **2018**, *54* (83), 11781-11784.

16. Simon, C.; Lion, C.; Spriet, C.; Baldacci-Cresp, F.; Hawkins, S.; Biot, C., One, Two, Three: A Bioorthogonal Triple Labelling Strategy for Studying the Dynamics of Plant Cell Wall Formation In Vivo. *Angew. Chem. Int. Ed.* **2018**, *57* (51), 16665-16671.

17. Simon, C.; Lion, C.; Huss, B.; Blervacq, A. S.; Spriet, C.; Guerardel, Y.; Biot, C.; Hawkins, S., BLISS: Shining a light on lignification in plants. *Plant Signal Behav* **2017**, *12* (8).

18. Hu, Y.; Roberts, J. M.; Kilgore, H. R.; Lani, A. S. M.; Raines, R. T.; Schomaker, J. M., Triple, Mutually Orthogonal Bioorthogonal Pairs through the Design of Electronically Activated Sulfamate-Containing Cycloalkynes. *J. Am. Chem. Soc.* **2020**, *142* (44), 18826-18835.

19. Nikic, I.; Plass, T.; Schraidt, O.; Szymanski, J.; Briggs, J. A.; Schultz, C.; Lemke, E. A., Minimal tags for rapid dual-color live-cell labeling and super-resolution microscopy. *Angew Chem Int Ed Engl* **2014**, *53* (8), 2245-9.

20. Liu, F.; Liang, Y.; Houk, K. N., Bioorthogonal Cycloadditions: Computational Analysis with the

Distortion/Interaction Model and Predictions of Reactivities. *Acc. Chem. Res.* **2017**, *50* (9), 2297-2308.

21. Fernandez, I.; Bickelhaupt, F. M., The activation strain model and molecular orbital theory: understanding and designing chemical reactions. *Chem. Soc. Rev.* **2014**, *43* (14), 4953-4967.

22. Ess, D. H.; Houk, K. N., Theory of 1,3-dipolar cycloadditions: distortion/interaction and frontier molecular orbital models. *J. Am. Chem. Soc.* **2008**, *130* (31), 10187-10198.

23. Ess, D. H.; Houk, K. N., Distortion/interaction energy control of 1,3-dipolar cycloaddition reactivity. *J. Am. Chem. Soc.* **2007**, *129* (35), 10646-10647.

24. Ravasco, J. M. J. M.; Coelho, J. A. S., Predictive Multivariate Models for Bioorthogonal Inverse-Electron Demand Diels-Alder Reactions. *J. Am. Chem. Soc.* **2020**, *142* (9), 4235-4241.

25. Danilkina, N. A.; Govdi, A. I.; Khlebnikov, A. F.; Tikhomirov, A. O.; Sharoyko, V. V.; Shtyrov, A. A.; Ryazantsev, M. N.; Bräse, S.; Balova, I. A., Heterocycloalkynes Fused to a Heterocyclic Core: Searching for an Island with Optimal Stability-Reactivity Balance. *J. Am. Chem. Soc.* **2021**, *143* (40), 16519-16537.

26. Kumar, G. S.; Racioppi, S.; Zurek, E.; Lin, Q., Superfast Tetrazole-BCN Cycloaddition Reaction for Bioorthogonal Protein Labeling on Live Cells. *J. Am. Chem. Soc.* **2021**.

27. Riu, M.-L. Y.; Transue, W. J.; Rall, J. M.; Cummins, C. C., An Azophosphine Synthetic Equivalent of Mesitylphosphazide and Its 1,3-Dipolar Cycloaddition Reactions. *J. Am. Chem. Soc.* **2021**, *143* (20), 7635-7640.

28. Karver, M. R.; Weissleder, R.; Hilderbrand, S. A., Bioorthogonal Reaction Pairs Enable Simultaneous, Selective, Multi-Target Imaging. *Angew. Chem. Int. Ed.* **2012**, *51* (4), 920-922.