Article

Development of Parallel On-the-Fly *Crystal* Algorithm for Reaction Discovery in Large and Complex Molecular Systems

Ankit Pandey, Gustavo J. Costa, Mushfiq Alam, Bill Poirier,* and Ruibin Liang*

Read Online

| ACCESS | | Metrics & More | | Article Recommendations | | SI Supporting Information |

ABSTRACT: The parallel on-the-fly *Crystal* algorithm is a new, efficient global search algorithm for exploring single-state potential energy surfaces and conical intersection seam spaces of a wide range of molecules. Despite major developments, its application to complex molecular systems, especially in the condensed phase, remains challenging due to the high dimensionality of the configurational space. In this work, we address this challenge and extend its applicability to the reaction discovery of large and complex molecular photoswitches in various molecular environments, including in the condensed phase with explicit solvent molecules. This is achieved by performing an explicit exploration of a comparatively large *Crystal* configurational subspace, while gradually relaxing the remaining degrees of freedom. The new *Crystal* algorithm is applied to the reaction discovery of bilirubin and donor–acceptor Stenhouse adducts, a next-generation class of molecular photoswitches, in vacuum and in the aqueous solution. To this end, we designed an automated and systematic workflow for *Crystal* to discover and characterize new minima and low-energy reaction pathways in these challenging and complex systems. Our findings demonstrate the algorithm's effectiveness in quickly exploring the configuration space and uncovering kinetically accessible products, offering new insights into the intricate chemical reactivities of these molecules and the roles of molecular environments on the reaction pathways. The results underscore the promising potential of parallelized global exploration methods for reaction discovery in biomolecular systems.
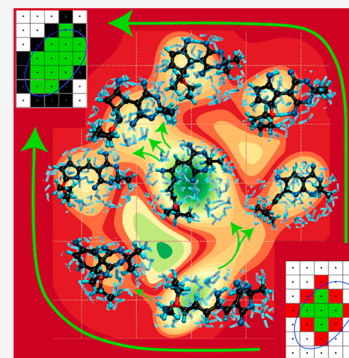
INTRODUCTION

Reaction discovery through molecular modeling is an emerging field that has seen significant progress in recent years.[1–14] Computational predictions of reaction products and pathways in complex molecular systems have become useful for interpreting and guiding new experiments. Central to the computational reaction discovery is the efficient exploration of the low-energy regions of potential energy surfaces (PESs). To achieve this goal, one class of approaches utilizes trajectories to dynamically sample the configuration space of the molecular systems. Among this type of approach, usually, the sampling is enhanced by imposing external biasing forces on a certain set of reaction coordinates, such as metadynamics,[15–17] or changing the global temperature and pressure of the system, such as the nanoreactors.[7] Another class of approaches does not involve running trajectories, but instead overcomes reaction barriers and locates new product minima by evaluating the energies, gradients and Hessians at existing static geometries (e.g., the growing string method[18–20]). Among these algorithms, some methods were designed to attempt a global search of the PES, such as the Tabu search method,[21–23] the anharmonic downward distortion approach,[24] machine-learning-based approaches,[25] and the *Crystal* algorithm.[13,26–29] In principle, the advantage of global algorithms is their ability to exhaustively explore configurational space to identify all minima and transition states. The global nature of these algorithms ensures the identification of

new products and low-energy pathways that might not be sampled by trajectory-based approaches. This is because the search results of global algorithms are less sensitive to choices of the initial structure and the set of selected collective variables, as compared to trajectory-based enhanced sampling simulations.

The *Crystal* algorithm[13,26–29] is based on the simple idea of performing an exhaustive search among all combinations of discretized moves of all selected reaction coordinates, subject to a maximal energy constraint. The moves are performed on a high-dimensional lattice spanned by these reaction coordinates, with predefined lattice point spacings in each dimension. Notably, the number of lattice points with energies below a certain preset value only scales polynomially (see Supporting Information for discussion) with respect to the number of degrees of freedom (d.o.f.s) of the system, rather than exponentially as expected for a brute-force global search algorithm. The reduced scaling behavior lays the foundation for the efficiency of the *Crystal* algorithm, making it suitable for
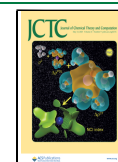
discovering reactions with low energy barriers that are kinetically and thermodynamically relevant. *Crystal* prevents the re-evaluation of previously visited lattice points by maintaining lists of "accepted" and "rejected" lattice points, which bears a resemblance to the "Tabu list" of the Tabu search method.[21−23] However, unlike the Tabu method, *Crystal* performs a comprehensive search of the entire energetically accessible configurational space and is thus theoretically guaranteed to find all reaction pathways below the specified maximum energy threshold.

Recently, the *Crystal* algorithm has undergone several major developments. One such development is the generalization to on-the-fly electronic structure calculations of potential energies,[26] eliminating the need to construct a predetermined global PES prior to the search. *Crystal*'s efficiency in this context has been greatly increased by an improved search algorithm and the implementation of a massively parallelizable version of the code.[13] *Crystal* has also been generalized to search for low-energy conformations in the conical intersection seam space[13] in addition to single-state PESs, facilitating photochemical reaction discovery. These improvements have broadened *Crystal*'s applicability to larger and more realistic systems. Although other global search algorithms have been recently developed, such as the anharmonic-downward-distortion based approaches developed by Morokuma et al.,[2,30−33] *Crystal* offers unique advantages: it does not require Hessian evaluations, is massively parallelizable, and does not assume a harmonic-like PES structure.

Despite its recent success, the application of the *Crystal* algorithm has been restricted to modestly sized molecular systems, often smaller than a few tens of atoms. The main reason is that the number of low-energy lattice points increases rapidly (albeit not exponentially) as the number of d.o.f.s spanning the lattice space grows. Currently, the *Crystal* algorithm cannot efficiently search in a lattice space of more than 30 dimensions with on-the-fly electronic structure calculations. For systems larger than a few tens of atoms, it is computationally impractical to include all d.o.f.s explicitly in the *Crystal* lattice space, even with relatively inexpensive quantum chemistry methods such as semiempirical methods. This is largely due to the large number of low-lying lattice points in the high-dimensional space, as well as the much slower speed at which the PES of larger systems is evaluated on-the-fly.

In this paper, we substantially develop the *Crystal* algorithm further, to greatly enhance its capability for reaction discovery in large and complex molecular systems, including condensed phase systems. The central idea behind the latest developments is to allow the d.o.f.s outside of *Crystal*'s explicit lattice configuration subspace to relax, in response to the changes of the d.o.f.s inside *Crystal*'s subspace. In other words, this approach enables exhaustive sampling of all of the d.o.f.s that are essential to the chemical events, while letting the rest of the system adjust to the motions of these essential d.o.f.s. There are several key technical advantages of this new approach over the previous version, which fixed the values of d.o.f.s outside the explicit *Crystal* subspace. The first is that the new algorithm enables the use of larger steps of *Crystal* moves, i.e., larger lattice spacings, making the conformational exploration more efficient while effectively avoiding steric clashes caused by d.o.f.s outside of the *Crystal* subspace. This is especially beneficial for reaction discovery in condensed-phase systems. Second, the new algorithm minimizes the chance of

inadvertently crossing hidden high energy barriers as a result of using large steps. This is critical for retaining *Crystal*'s ability to faithfully discover kinetically relevant reaction pathways below a defined energy barrier limit. Third, the new *Crystal* algorithm is readily integrated with the quantum mechanics/molecular mechanics (QM/MM) evaluation of the PES, making it feasible for PES exploration in condensed phase molecular systems.

All the above-mentioned advances greatly expand the applicability of *Crystal* toward reaction discovery in larger molecular systems of importance in biomedical and material sciences. To test the effectiveness of this new approach, we applied it to the electronic-ground-state reaction discovery of complex molecular photoswitches in the gas phase and aqueous solution. Molecular photoswitches can be reversibly switched between different isomeric forms by different wavelengths of light.[34] They can be covalently or non-covalently embedded in large molecular systems and make the latter photoresponsive. Their scope of applications has been growing in fields ranging from photopharmacology to photoresponsive materials,[35−39] where light is used for controlling the activity of drugs in target biomolecules and changing the mechanical and photophysical properties of materials.

One class of visible-light responsive photoswitches, known as donor−acceptor Stenhouse adducts (DASAs), has gained significant interest recently.[40−42] DASAs are "multi-step" photoswitches with multiple isomers that can transform into one another reversibly under the control of light, thus opening new doors to manipulating complex molecular properties.[43] The photoswitching of DASAs is mediated by conical intersections[8,44] that facilitate non-radiative decay to different ground-state minima through various pathways. Understanding these pathways and identifying "unproductive" pathways that cause fatigue-induced damage, i.e., reduction of quantum yield, requires knowledge of the relevant stable and metastable states that can be accessed by the molecule.[8,45,46] In ref 8, Sanchez et al. applied an in silico approach to identify such geometries for Meldrum's acid DASA derivative, the simplest of the DASAs. Subsequently, Raucci et al.[9] employed metadynamics to discover many previously unknown conformers of DASA on the ground state.

DASA is unstable in aqueous environments, so it is typically studied in nonpolar solvents. Computational studies of DASAs have been performed in various environments, including the gas phase,[47] toluene[9] and trifluoromethanesulfonic acid.[48] However, there is growing interest in water-soluble DASA molecules. Water-assisted isomerization has been explored in ref 49. Studying DASA in water can reveal key solute−solvent interactions and provide insights into designing stable water-insoluble DASAs, and predicting water's effect on photo-isomerization. These insights could guide the development and tailoring of DASA derivatives with desirable properties.

Another class of biologically significant molecules in which photochemistry plays a crucial role are tetrapyrroles, including chlorophyll, porphyrins, phycobilins, and bilirubin.[50−53] Among these, bilirubin plays an important role in human physiology, exhibiting neuroprotective effects at physiological concentrations.[54] It acts as a free radical scavenger, protecting brain tissue from oxidative stress and aiding resistance against age-related illnesses. However, excessive bilirubin becomes neurotoxic. In its natural state, bilirubin is water-insoluble and binds to albumin, but conjugation with glucuronic acid in the
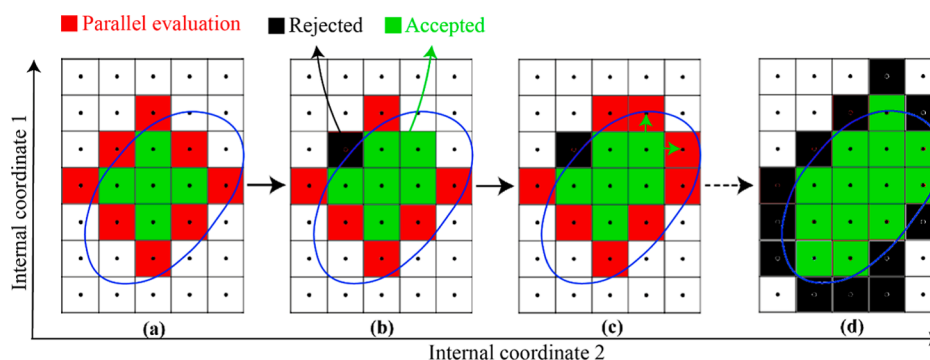
Figure 1. Overview of the *Crystal* algorithm. (a) The accepted lattice points in the previous search iteration are represented by the green lattice points. The blue contour denotes the updated energy cutoff in the current step. The red lattice points are being evaluated by parallelized secondary jobs. (b) Two of these points have completed energy evaluation. One has been rejected due to high energy (black), while the other has been accepted (new green point). (c) The unexplored neighbors of the newly accepted points are added to the evaluation queue. (d) The end result after all evaluations are completed for the current energy cutoff, which can be increased in the next iteration.

liver makes it water-soluble, facilitating excretion.[55,56] However, immature liver function in newborns can hamper this process, leading to kernicterus.[57] Light therapy is used to treat this condition by photoisomerizing bilirubin, making it water-soluble and easier to eliminate.[58] In its native Z,Z isomer form, bilirubin is stabilized by six intramolecular hydrogen bonds. Photoisomerization disrupts some of these bonds, allowing water to form hydrogen bonds with bilirubin. The thermal isomerization of bilirubin is known to proceed via multiple intermediates[59] and slow in the dark. A smaller portion of bilirubin is irreversibly converted to water-soluble lumirubin, aiding its removal from the body.[60−62] Bilirubin is also structurally closely related to several key chromophores in a variety of natural photoreceptors, such as phycocyanobilin in phytochromes.

Understanding the PES of bilirubin could aid in identifying and stabilizing favorable isomeric forms. This knowledge could support drug development aimed at enhancing bilirubin clearance by stabilizing specific conjugated forms. It may also shed light on the energetics of crossing the blood−brain barrier and other factors influencing bilirubin's interaction with the brain, offering potential strategies to inhibit its toxic configurations.

In this paper, we employ the newly developed *Crystal* algorithm to explore the ground state PESs of Meldrum's acid DASA in the gas phase, implicit solvent, and condensed phases, as well as bilirubin in the gas phase.

One of our primary goals is to comprehensively discover all key types of DASA reactions, including intramolecular proton transfer, photoisomerization around double bonds, and electrocyclization. These three key reaction types encompass nearly all reactivities observed in a wide array of modern molecular photoswitches applied in biomedical and material sciences, including but not limited to derivatives of azobenzenes, stilbenes, spiropyrans, diarylethenes and hydrazone photoswitches. Consequently, automatic computational reaction discovery for this single and complex DASA molecule is highly challenging, serves as a solid demonstration of our method's applicability to a wide array of photoactive molecules. The computational efficiency of the new *Crystal* algorithm is on par with and sometimes even better than dynamics-based methods utilized for similar systems[8,9] (see SI). Furthermore, our approach has the advantage that it provides *guaranteed* discovery (at least in principle) of all energetically accessible reaction pathways. Overall, our results

demonstrate the excellent performance of the new *Crystal* algorithm for reaction discovery in large and complex biomolecular systems of practical interest in biomedical and material science.

■ METHODS

Below, we first give an overview of the original *Crystal* algorithm, followed by a discussion of the limitations of the previous version of this algorithm, which serve as the motivations for the new developments. Next, the rationales of new developments are discussed, followed by a full summary of the new algorithm. The computational details of the three systems are then described.

For a clear presentation of our most up-to-date approach while avoiding too much diversion here in the main text, in the Supporting Information we include important data to document the significant improvements over the previous version of this method thoroughly. The new and previous *Crystal* algorithms are compared in terms of the quality of the reaction discovery and computational efficiency, along with comparisons with other related approaches. These additional data sets highlight the major challenges we encountered in developing the new method, supporting our motivations and rationales for the improvements.

Overview of the Original Crystal Algorithm. The *Crystal* algorithm was originally designed as a global PES exploration algorithm that exhaustively searches for all geometries below a predefined maximal energy threshold. The search begins with a definition of conformational subspace spanned by an ensemble of key d.o.f.s. *Crystal* then discretizes each dimension spanning this subspace and creates a list of lattice points representing it. Starting from an initial lattice point representing a minimum in the PES, the search algorithm explores the configuration subspace by iteratively accepting lattice points in the neighborhood of existing points with energies lower than a specified energy cutoff value. The energies of lattice points may be evaluated with on-the-fly electronic structure calculations, such as ab initio or semi-empirical methods.[13,26] In the original implementation, d.o.f.s outside the lattice space are fixed to their initial values during the entire search. This process is summarized in Figure 1. For all lattice points in the list, neighbors that have not yet been evaluated are added to an evaluation queue. Secondary jobs process the evaluation queue in parallel, recursively selecting

lattice points from the queue, evaluating them, and proceeding to the next point. The energy cutoff value is then incrementally raised to explore the configuration space progressively until it reaches the predefined maximal energy threshold. At the end of the exploration, in principle, the algorithm will identify all lattice points that are energetically accessible from the initial minima by crossing transition states below the predefined maximal energy threshold. Following the exploration phase, a subsequent refinement phase will be carried out starting from these lattice points, and geometry optimizations will pinpoint the exact locations of all new minima and transition states in the explored region of the configurational subspace. The use of energy threshold and cutoff values as selection criteria for accepting lattice points greatly reduces the computational cost and makes it scale polynomially rather than exponentially (see Supporting Information for details) with respect to the subspace dimension.

Limitations of the Original Crystal Algorithm in Conformational Sampling of Large Systems. Despite the good scaling behavior, due to the global nature of the algorithm, the upper limit of subspace dimensionality manageable by *Crystal* is about 30 d.o.f.s. For smaller molecules, such as those discussed in our previous studies,[13,26−29] *Crystal* can quickly scan a large number of geometries by including most of the essential low-frequency d.o.f.s in the lattice space, while leaving the high-frequency modes outside the subspace and fixing them during the search. However, large molecular systems often have many low-frequency modes that are coupled to one another. In this case, due to the fixing of the d.o.f.s outside the lattice space, a much larger number of d.o.f.s must be included in the lattice space, greatly increasing the dimensionality and the computational cost. Thus, practical considerations of computational speed necessitate significantly increasing the lattice point spacing explored by *Crystal* in most of the dimensions spanning the subspace. However, this can lead to several serious problems. First, many important and distinct minima between the large spacing may fail to be detected. Second, high energy barriers beyond the maximal energy threshold may be overcome unintendedly because of the large stepping in the lattice space. This can lead to the discovery of reaction pathways that are kinetically inaccessible and many physically irrelevant minima. For instance, in our experience, major steric repulsions were overcome accidentally, resulting in the discovery of drastic bonding topology rearrangement of the original molecule, sometimes leading to the disfiguration or disintegration of the entire molecule. These observed new reactions are mostly spurious in that it would have required a much higher energy threshold to detect them had the lattice spacing been chosen to be sufficiently small. As a remedy, in these cases, extensive subsequent evaluations of transition states are necessary to determine whether the observed reaction path is kinetically feasible, typically requiring many expensive path optimizations using methods such as the nudged elastic band (NEB). Third, the use of large lattice spacing can easily lead to many high-energy structures due to steric clashes, decreasing the acceptance ratio of new lattice points and reducing the computational efficiency.

The Supporting Information provides a detailed discussion of the previous *Crystal* method's limitations, which were discovered by our simulation data reported therein (Figures S1 and S2).

Developments of Crystal Algorithm. Therefore, for large molecular systems, we reason that it is essential to divide the d.o.f.s into two categories: (1) the d.o.f.s key to the chemical reactivity and explicitly explored in the *Crystal* lattice space (subsystem, explicit d.o.f.s), and (2) those that relax in response to the displacements of the d.o.f.s in the first category (environment, implicit d.o.f.s). In this scheme, the energy of each lattice point corresponds to the structure optimized with the constraints that all explicit d.o.f.s take their respective values specified by that lattice point. This scheme of dividing the total system is analogous to the time scale separation that has been extensively exploited in enhanced sampling simulations with collective variables (CV) and trajectory dynamics (such as umbrella sampling).[63] In these techniques, the slow CVs that are rate-limiting for the process of interest are explicitly biased by external forces. At the same time, the fast d.o.f.s are supposed to quickly and automatically respond to the changes in the slow CVs during the dynamics.

To compensate for the increased cost of constrained optimizations, we introduced an additional step into the workflow to allow for the use of large lattice spacing, while reducing the chances of missing minima, inadvertently stepping over high energy barriers or encountering high energy structures. The following approach proved effective through our test: the implicit d.o.f.s can be optimized gradually using a relaxed scan during one large *Crystal* move (Figure 2).
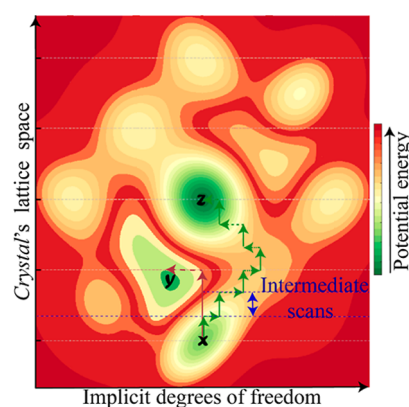
Figure 2. Starting from the minimum "x", a single big-step *Crystal* move using large lattice spacing (the solid and vertical red arrow) can erroneously overcome a high barrier, followed by relaxation of implicit d.o.f.s (the dashed and horizontal red arrow) into a minimum "y". This minimum is not kinetically accessible from "x" given the current energy cutoff. It should not have been discovered with the current energy cutoff if a smaller lattice spacing had been used. On the other hand, the relaxed scan approach (small green arrows) gradually optimizes the implicit d.o.f.s in intermediate points of a big-step *Crystal* move, effectively directing the system close to the minimum energy pathway, eventually finding a kinetically accessible minimum "z". This approach thus enables the use of larger lattice spacings without discovering kinetically unfeasible reaction pathways dictated by the energy cutoff.

For instance, if a single *Crystal* move dictates that a torsion in the explicit d.o.f.s be changed from 0 to 60°, it is realized by a relaxed scan of the system with that torsion gradually increased at 20° intervals toward its target value, while the remaining explicit d.o.f.s are constrained to their respective values. The intermediate energies of this relaxed scan are closely monitored so as not to exceed the maximal energy threshold at any given intermediate point during the single *Crystal* move. Thus, the
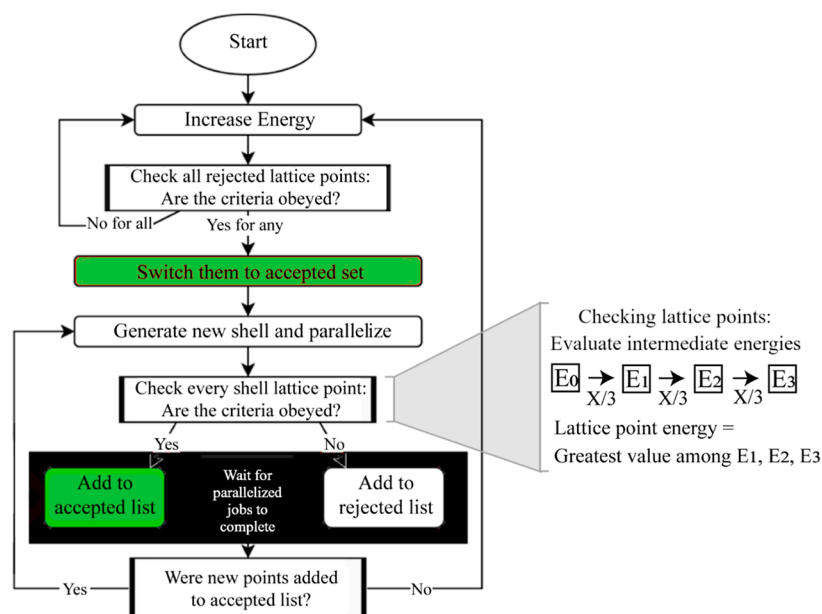
Figure 3. Flowchart of the parallel on-the-fly *Crystal* algorithm. Left: Overview of the general *Crystal* algorithm (see refs 13 and 26 for further details). Right: The method for lattice point energy evaluation has been adapted to large lattice spacing. In this example, a relaxed scan through three intermediate small steps is performed during a big step X (*Crystal* move). At each of the small steps (X/3), the implicit d.o.f.s are optimized, and the energy is evaluated. *Crystal* records the highest energy from these evaluations as the barrier for taking the big step X. This barrier is subject to the control by the maximal energy threshold to filter out kinetically inaccessible pathways during the discovery.

likelihood of stepping over kinetically inaccessible barriers is decreased, as relaxing the implicit d.o.f.s. helps direct the system to follow the minimum-energy pathway (Figure 3). Importantly, the computational cost of these additional relaxed scan steps only scales linearly with respect to the number of intermediate points in the relaxed scan for each explicit d.o.f. However, this approach allows for the use of large lattice point spacings, which substantially decreases the cost. As a result, the improvement in computational efficiency was found to be massive (see Supporting Information for a detailed discussion of the trade-off). A flowchart illustrating the overall scheme of the new *Crystal* algorithm is provided in Figure 3. This idea, in principle, is analogous to the multistep integration scheme in nonadiabatic dynamics simulations,[64,65] where the slower nuclear d.o.f.s are integrated using a much larger step than the faster electronic d.o.f.s., which relax in response to the former in the intermediate steps.

The above-mentioned developments of *Crystal* bring several major advantages over the original version. First, it enables the use of much fewer explicit d.o.f.s. to capture the essential features of possible chemical reactions while allowing the rest of the d.o.f.s (implicit d.o.f.s) to reach their energetically favorable configurations. This will significantly save the computational cost due to dimensionality reduction. Second, the constrained geometry optimization of the implicit d.o.f.s, although more expensive than single-point energy calculations, will automatically avoid a large number of energetically unfavorable configurations that would have been generated if they were kept frozen. Thus, it will increase the acceptance ratio of explored lattice points and further save computational costs. Third, the increased likelihood of accepting new lattice points as well as the gradual relaxation of the implicit d.o.f.s makes the algorithm more tolerant to the use of large lattice spacing, which further reduces the computational cost.

The following two technical details are worth noting. First, the *Crystal* increases its energy cutoffs gradually after starting at

the value of the first lattice point (which is chosen at a minimum). Second, the constraint-optimized geometry of the previous lattice point is used to generate the starting geometry of the next lattice point. Taken together, the exploration will largely retain the continuity of the minimum in the subspace of the implicit d.o.f.s in the initial structure. When multiple minima exist in the subspace of implicit d.o.f.s, the method will most likely keep the system at the minimum most similar to the initial geometry. Arguably, we view it as an advantage rather than a disadvantage in most scenarios. This is because the explicit d.o.f.s are supposed to adequately sample the most important chemical reactivity, while the implicit d.o.f.s are supposed to have relatively less important conformational transitions. During the exploration, keeping the system at a consistent minimum in the subspace of implicit d.o.f.s will reduce the fluctuations in energy caused by significant conformational changes in these d.o.f.s, thus making the result more interpretable. If the conformational transitions in the implicit d.o.f.s are truly important, which was observed for the proton hopping reactions for DASA (see "Results"), they can be further included in *Crystal*'s explicit d.o.f.s. subspace in a restart simulation.

The developments described above were applied to investigate Meldrum's acid DASA in both the gas phase and condensed phases, as well as bilirubin in the gas phase. The results show that parallel on-the-fly *Crystal* is an effective tool not only for exploring the explicit d.o.f.s. but also for facilitating a more comprehensive examination of the implicit d.o.f.s. This new version of the *Crystal* algorithm, even with relatively low energy thresholds, enables the discovery of reaction pathways that are overlooked by existing methods.[9]

In the Supporting Information (Figure S3), we compare the simple constrained-optimization scheme and the relaxed scan scheme, the latter being the most up-to-date version of the method presented in this work. In particular, we show that the efficiency gained from the relaxed scans approach is more than

incremental—it represents an entire order of magnitude improvement over our previous method but also achieves a significant leading edge over other top-of-the-line competitive approaches. We also benchmarked this new method's computational efficiency against existing approaches applied to similar systems.

Meldrum's Acid DASA in Implicit Solvent. Our workflow starts with a *Crystal*-based exploration phase, during which the PES is scanned using large lattice spacings in combination with small intervals in relaxed scans. This is followed by a refinement phase, during which the identified geometries at the lattice points are optimized and classified. During the exploration phase, we employ the semiempirical electronic structure method GFN2-xTB in the gas phase.[66] This method can effectively optimize geometries of molecular systems with hundreds of atoms, providing reasonable accuracy at a significantly faster speed than DFT.[66] In the exploration phase, all explicit d.o.f.s used lattice spacings of 45°, with the implicit d.o.f.s. optimized after every 15° interval in the relaxed scan. In ref 13, this choice of 45° steps was determined to strike an acceptable balance between computational speed and accuracy for a preliminary evaluation. For the angular coordinates, the *Crystal* lattice spacing was set to 20°, with three equally spaced intermediate points for relaxed scans. The reference starting geometry and the explicit d.o.f.s. explored by *Crystal* are shown in Figure 4.
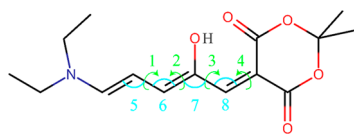
Figure 4. 8-Dimensional configuration space of explicit d.o.f.s explored for Meldrum's acid DASA by *Crystal*. Torsional d.o.f.s are labeled as 1−4 (green), while angular d.o.f.s are labeled as 5−8 (cyan). All remaining implicit d.o.f.s were optimized. All torsional d.o.f.s for DASA were defined using carbon atoms.

In principle, *Crystal* works with any coordinate system. The reasons for choosing these explicit d.o.f.s. for DASA are explained below. The objective of the search was to thoroughly explore the reactions commonly observed reactions in a wide variety of molecular photoswitches, such as isomerization, cyclization, and intramolecular proton transfer. They can be mostly induced by torsions and angles around double bonds and hydrogen-bonded electronegative atoms such as oxygen atoms.

Moreover, for reasons explained in our discussion of coordinate driving method in the Supporting Information, including bond lengths in the explicit d.o.f.s typically introduces difficulties toward the automated discovery of complex systems. Particularly, for bond lengths to be reasonable scanning variables in large molecules, one typically needs to have extensive prior knowledge of all important reactive bonds; but including all potentially important bonds can easily increase computational cost due to the high dimensionality of the subspace. On the other hand, so long as there is an overlap between the explicit d.o.f.s and the bonds of interest, *Crystal* has the ability to induce bond-breaking/ formation reactions without including these bonds among the explicitly treated d.o.f.s. We frequently observed that the bond length automatically adjusted to the motions of the explicit d.o.f.s that drive the system near the transition state. This

phenomenon is viewed as an advantage of this approach, as explained in the Results and Conclusion sections. In other words, changing angles and dihedrals in these cases is sufficient to bring the atoms close to each other and facilitate bond-forming reactions. Thus, we prefer this choice for easy automation of the process.

The refinement stage involves optimizing the low-energy lattice points identified by *Crystal* using density functional theory (DFT), a more accurate electronic structure method. The ωB97x-D3 functional[67] was used in combination with the def2-TZVP(-f) basis set and the COSMO implicit solvent model[68] ($\varepsilon = 2.38$), i.e., ωB97x-D3/def2-TZVP(-f)/COSMO. The geometries are classified based on the *cis/trans* isomerization of the linear triene chain and the bond connectivity. The hydrogen atom marked in the Figure 4 was observed to hop to other atoms in the first round of *Crystal* simulation, despite not being included among the explicit d.o.f.s. This hopping behavior, commonly reported in the literature, is known to occur during the cyclization of the triene chain. Hence, we performed a restart of the exploration phase, during which three additional d.o.f.s were assigned to this hydrogen atom.

Initial minimum energy pathway pathways (MEPs) between unique geometries were generated using internal-coordinate geodesic interpolation,[69] followed by optimization using the NEB[70] approach with frozen end points method. The NEB calculations were performed at ωB97x-D3/def2-TZVP(-f)/ COSMO level of theory.

All energy and gradient evaluations were conducted using the TeraChem software package,[71−74] which utilizes GPU nodes for accelerated electronic structure calculations. *Crystal*'s energy threshold was set to 40 kcal/mol. The exploration phase employed up to 30 secondary jobs over a span of 2 days. The finer runs were conducted with shorter runtimes, typically lasting less than a day. All constrained optimizations and relaxed scans were performed with the geomeTRIC optimizer[75] integrated into the TeraChem package. The NEB calculations were performed using the DL-FIND optimizer[76] interfaced with TeraChem package.

Meldrum's Acid DASA in the Aqueous Solution. Meldrum's acid DASA was solvated and equilibrated within a periodic boundary condition (PBC) simulation box of 55.9 × 58.4 × 60.9 Å³ dimensions. The generalized AMBER force field (GAFF) procedure[77,78] was used to parametrize the force fields of the DASA molecule. The SPC/Fw water model[79] was employed. Molecular dynamics (MD) equilibration simulations were initiated in the constant *NVT* ensemble for 1.2 ns at 300 K temperature, followed by equilibration in the constant *NPT* ensemble for 1 ns at 300 K temperature and 1 atm pressure, all with a 1 fs time step. A 10 Å cutoff was applied for nonbonded interactions, and the coordinates of DASA were allowed to relax during the simulation. Following classical MD simulation, the system was truncated to include the DASA and the water molecules within 10 Å of the DASA molecule, forming a subsystem with open boundary conditions (Figure 5). This subsystem was then subject to a short ground-state QM/MM MD simulation of 2.5 ps with a time step of 0.5 fs. The QM region included the DASA molecule only and was treated at BH&HLP/6-31G* level of theory. The MM region included the remainder of the system. The QM and MM regions were coupled through the electrostatic embedding scheme. The last snapshot of the QM/MM MD trajectory served as the initial structure for the *Crystal* simulation.
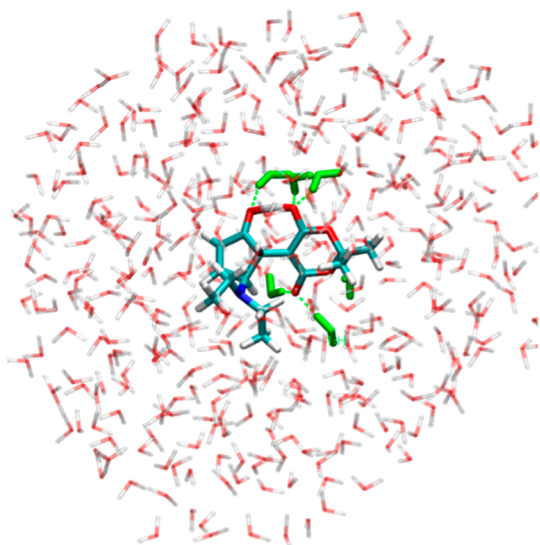
Figure 5. Meldrum's acid DASA in aqueous solution (explicit solvent). The green color denotes water molecules showing hydrogen bonds with the solute. The explicit internal coordinate space utilized by *Crystal* was kept unchanged from the gas phase calculations (see Figure 4).

In the *Crystal* simulations, QM/MM calculations provided on-the-fly evaluations of the PES and its gradients. The QM region included only the DASA molecule and was treated using the GFN2-xTB level of theory. The MM region contained all the water molecules and was treated with the same force field used in the MD simulation. The DASA molecule and all water molecules having any atoms initially located within 4 Å of any of its atoms were defined as the active region, where their coordinates were allowed to move, while the remaining water molecules were fixed. This resulted in 53 active water molecules out of a total of 390.

The condensed-phase QM/MM *Crystal* simulation was performed using CPU nodes exclusively. GFN2-xTB calculations were performed with the ORCA software package,[80] which was interfaced with the Chemshell[81] software package. All constrained optimizations and relaxed scans were performed with the DL-FIND optimizer[76] interfaced with the ChemShell package. The *Crystal* simulations were performed with an energy threshold of 40 kcal/mol, employing 45-degree torsions without including any hydrogen atoms in the explicit space. As in the gas-phase study, three intermediate points were used to optimize implicit d.o.f.s through the relaxed scan. The geometries were classified based on the *cis/trans* isomerization of the linear triene chain and any changes in bond connectivity. Finer lattice-spacing restarts were not performed.

For the NEB calculations, internal coordinate geodesic interpolations were first calculated between the end point geometries to generate the initial guesses of MEPs for NEB optimization. The NEB calculations were performed using ωB97x-D3/6-31G*/MM method in the aqueous solution. The relative energies and transition states of the condensed-phase geometries in aqueous solution were directly compared with gas-phase values. This was achieved by extracting the DASA molecule from the solution and reoptimizing the minima and pathways in the gas phase, where the ωB97x-D3/6-31G* was used.

Bilirubin. For bilirubin, double-bond isomerization and their relationship with hydrogen bonds are considered critical due to their biological significance. However, studying double-bond isomerization typically requires multireference electronic structure methods due to the near-degenerate frontier molecular orbitals near the transition state.[82] In our test, single-reference methods such as GFN2-xTB encountered convergence failure issues during the double bond rotations. For this reason, we used the semiempirical PM6 method[83] in combination with the FOMO-CASCI method[84] for *Crystal*'s exploration phase as well as for the subsequent geometry refinement. Finer lattice-spacing restarts were not performed.

Energy calculations were performed using the MOPAC software package, interfaced with the standalone GeomeTRIC software package[75] for optimizations. The explicitly d.o.f.s are labeled in Figure 6. A lattice spacing of 60° was used for all
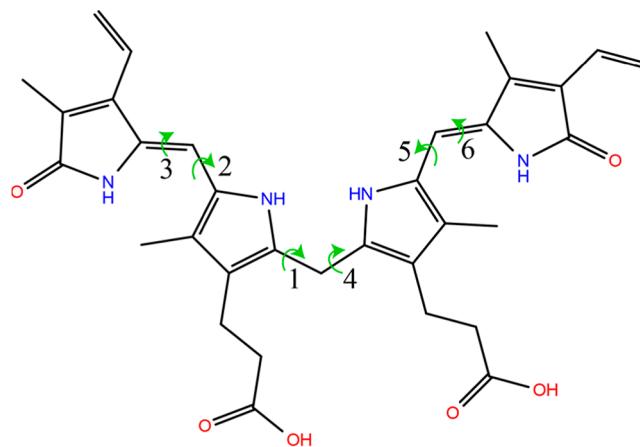

Figure 6. Chemical structure of the bilirubin molecule. The six explicit torsional d.o.f.s in the *Crystal*'s subspace are labeled as indices. Specifically, d.o.f.s 1,3,4 and 6 involve rotation about the CCCN dihedral angle, whereas dihedrals 2 and 5 involve rotation about the NCCC dihedral angle.

torsions, with implicit d.o.f.s minimizations carried out at every 20° interval in the relaxed scan. The energy cutoff for *Crystal* was set at 0.06 hartree (37.7 kcal/mol).

The resulting geometries were categorized into three isomeric forms: **Z−Z**, **E−Z/Z−E**, and **E−E**. Due to the differences in the energy scales governing H-bond changes and isomerization, these isomers do not readily interconvert to one another without light or catalysts. Therefore, unique geometries are examined separately within each isomeric form. For this purpose, two geometries are considered distinct if they differ in H-bond connections or if at least one of *Crystal*'s explicit dihedrals—specifically, dihedrals 1, 2, 4, and 5 in Figure 6—differs by 45°. H-bonds were identified based on two simultaneous criteria: (1) the H-to-acceptor distance is less than 2.5 Å, and (2) the donor-H-acceptor angle is greater than 120°.

The Supporting Information provides the thresholds of constrained optimizations and detailed data regarding computational costs. It also provides a benchmark of computational efficiency against the previous version of the *Crystal* algorithm and other approaches.
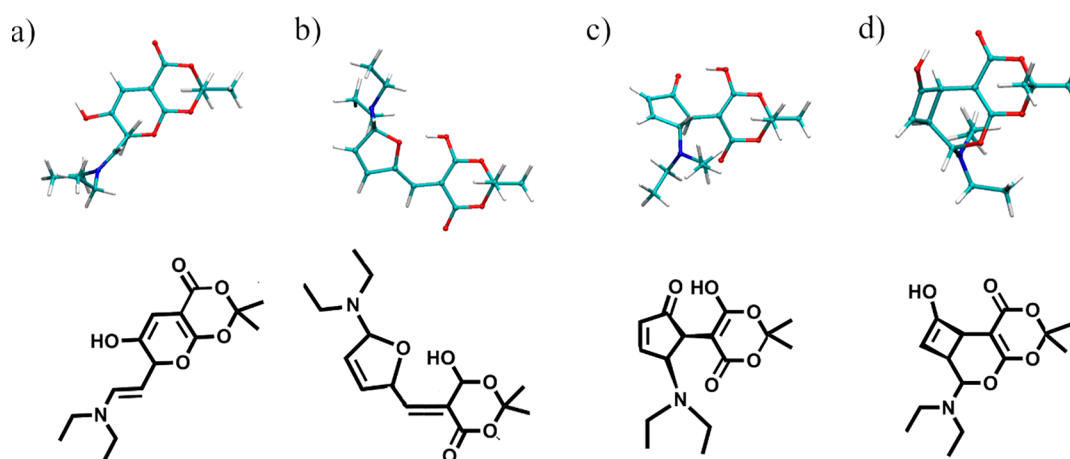
Figure 7. Starting from Meldrum's acid DASA in its linear form (Figure 4), four isomers identified by *Crystal* are shown as representative examples. The hydrogen atom that is observed to be prone to hopping between heavy atoms is labeled in the chemical structures at the bottom.
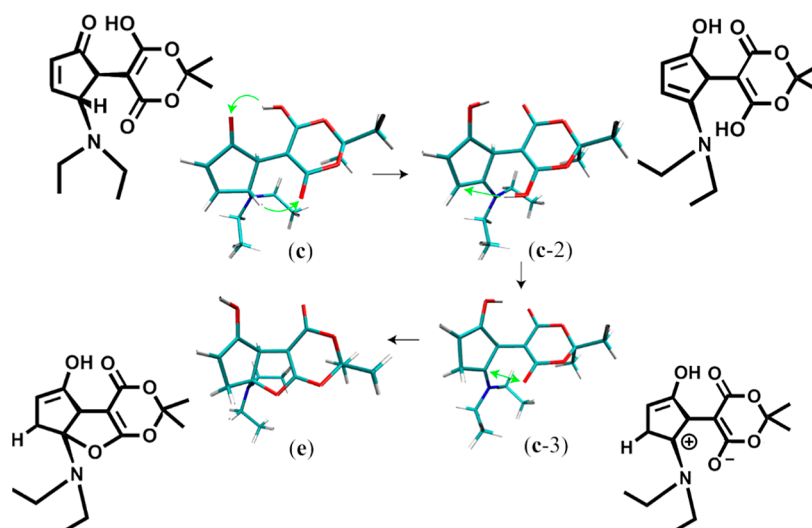
Figure 8. Restarting *Crystal* simulation from species **c** produced species **c-2**, **c-3**, and **e**. The two hydrogen atoms involved in the transfers are explicitly labeled. Species **c-2** forms **c-3** and **e** via double-hydrogen hopping (green arrows), with additional hydrogen transfers. All transitions have barriers below 25 kcal/mol relative to species **c** (see Table 1).

■ RESULTS

Meldrum's Acid DASA in Implicit Solvent. Figure 7 illustrates several representative low-energy isomers discovered by our newest version of *Crystal* algorithm. Species **a**, **b**, **c**, and **d** were all reported by Raucci et al.[9] As observed from these structures, the labeled hydrogen atom in Figure 7 is prone to hopping between heavy atoms. To better account for this motion, three additional d.o.f.s were assigned to this hydrogen during a second-round *Crystal* simulation restarting from species **c**. This led to the discovery of a double-hydrogen hopping pathway with relatively low energy barriers (Figure 8), which has not been reported in the literature (Table 1).

The MEP of the double-hydrogen-hopping reaction is shown in Figure 9. It is observed that the transfer of H between the two neighboring oxygen atoms has already happened before the TS is reached. The double-hydrogen hopping reaction could be an important consideration when working with DASA molecules that have a similar acceptor ring to DASA.

Restarting *Crystal* from species **a**, **b**, and **d** with 45-degree lattice spacings reproduced all the geometries reported in ref 9,

Table 1. Transition Barriers along the Double-Hydrogen Hopping Reaction Calculated via the NEB Approach, Utilizing the ωB97x-D3/def2-TZVP(-f)/COSMO Method

transitions	barrier height relative to **c** (kcal/mol)
c to **c-2**	23.89
c-2 to **c-3**	15.59
c-3 to **e**	5.88

demonstrating the ability of this method to perform comprehensive reaction discovery in large systems with complex reactivity.

DASA in the Aqueous Solution. Our results indicate that the aqueous solution environment can influence the geometries of the DASA molecule at the same minimum. For example, the linear geometry in the gas-phase minimum, when solvated in aqueous solution, exhibits significant out-of-plane bending, as shown in Figure 10.

Unique to the aqueous solution environment, *Crystal* identified a stable cyclic structure where the hydrogen atom remains attached to the hydroxyl group immediately after ring cyclization (Figure 11 and Table 2). The structure at this
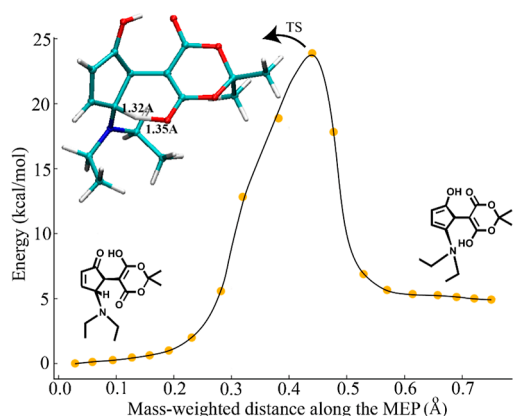
Figure 9. PES along the MEP for the double-hydrogen hopping reaction, showing the transition state (TS) between species **c** (left) and **c-2** (right). At the TS, the first hydrogen transfer (O-to-O near the top) is complete, while the second transfer (C-to-O, marked by the distance values) is midway between the two carbons. Notably, no local minimum exists to support the first hydrogen transfer without the simultaneous second hydrogen transfer.

minimum, **f-2**, was determined to be only 0.5 kcal/mol deep in the MEP. Additionally, multiple low-energy minima were observed following this pathway, suggesting that this geometry can be an intermediate to other, more stable species. Two such species, i.e., the **f-3** and **f-4**, with relatively low energy barriers are shown in Figure 11. Notably, the **f-3** is a previously unreported zwitterionic species, where the nitrogen atom receives hydrogen from a carbon atom rather than from the hydroxyl group (see Figure 11).

In aqueous environments, cyclic isomers are known to be thermodynamically favored.[49,85] According to ref 49, coordinated water molecules may play a catalytic role in facilitating the transfer of a proton from the hydroxyl group to the carbonyl group. It is widely believed that cyclic forms are strongly correlated with the hydrogen atom residing in the carbonyl group, a phenomenon that is consistently observed in the gas phase. However, our findings suggest that the stabilization of the **f-2** intermediate through hydrogen bonding in the condensed phase (Figure 12) enables the retention of the hydrogen within the hydroxyl group during cyclization. This stabilization could potentially lower the cyclization barrier.

For the **f-2** to **f-3** transition, the MEP is illustrated in Figure 13, with a ~1.3 Å distance for hydrogen transfer reaction. It has a 36.5 kcal/mol barrier. In DASAs with modified donor

and acceptor components, these factors—proton transfer facilitation and hydrogen bond stabilization—are likely to play varying roles in influencing the **f-2** to **f-3** pathway.

Figure 14 summarizes the geometries and energies identified by *Crystal* in the aqueous solution, and the corresponding gas phase energies. A total of 61 unique isomeric conformers were discovered, whose coordinates are provided in the Supporting Information. The gas-phase energies were evaluated by removing the water molecules, followed by reoptimization of structures and pathways. All condensed-phase and gas-phase calculations were performed using the ωB97x-D3 functional with the 6-31G* basis set.

Notably, the branch **c** to **e** of the implicit solvent case, shown earlier in Figure 8, is not among the geometries discovered during the condensed phase results (Figure 14). These differences might have arisen from changes in the transition barrier heights in the condensed phase. Our calculations confirmed that the **c** to **c-2** transition barrier is elevated in the condensed phase compared to the gas phase.

Bilirubin in the Gas Phase. The *Crystal* algorithm captures both double-bond isomerization and the associated changes in the H-bond patterns concurrently, generating various geometries for each of the Z-Z, Z-E/E-Z, and E-Z isomeric forms. A total of 423 unique geometries were obtained, whose coordinate files are provided in the Supporting Information. After removing duplicates considering symmetry, the 10 lowest-energy unique geometries identified for each isomeric form are summarized in Tables S1–S3. The unique hydrogen bonding patterns observed among these geometries are illustrated in the Figures 15, S4 and S5. When comparing these results with other studies in the literature, it is important to consider variations in H-bond identification criteria, such as distance and angle cutoffs, as well as differences in basis sets and levels of theory, which can lead to discrepancies.

For example, in ref 86, one geometry was evaluated for each of the three isomers, using a larger cutoff distance of 3 Å without considering donor-H-acceptor angular constraints. This led to the identification of 10 H-bonds. Additionally, carbon atoms were treated as hydrogen donors in this study,[86] which was not the case in ours. A closer comparison of the hydrogen-bond distance values reported in ref 86 shows that our lowest-energy structures exhibit significantly more H-bonds within the 2.5 Å cutoff. Some of our higher-energy isomers closely match their descriptions.

In general, we observe the expected trend of frequencies of different H-Bond patterns. The most stable configurations for each isomer tend to involve a higher number of hydrogen


(a)   (b)
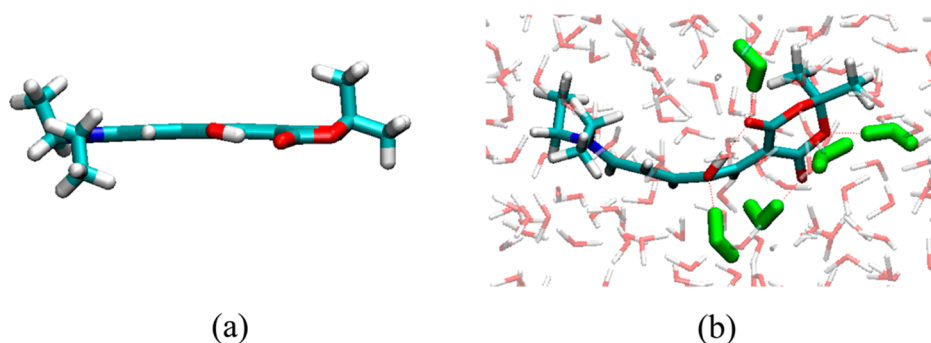
Figure 10. (a) DASA has a minimum of linear geometry in the gas phase. (b) In the aqueous solution, this minimum has a notable out-of-plane bending. The water molecules in green form hydrogen bonds with the DASA molecule.
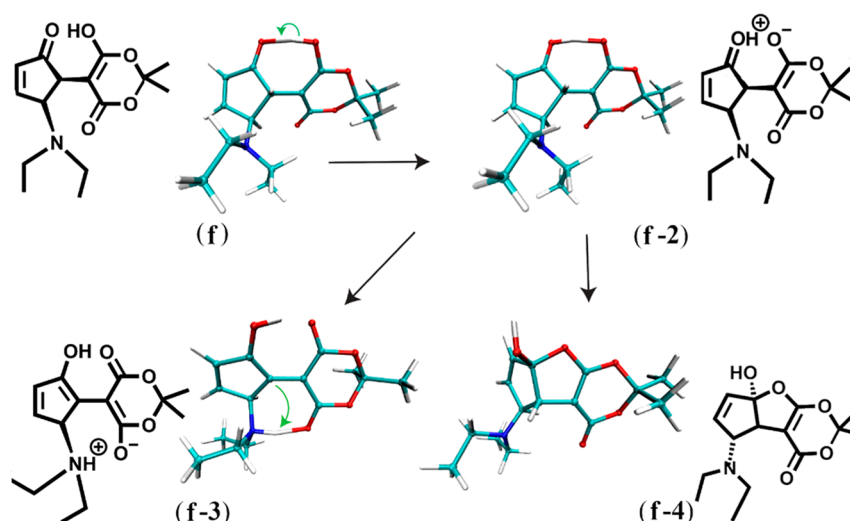
Figure 11. Species **f-3** and **f-4** of the DASA molecule were discovered only in the aqueous solution, and they reside in low-energy minima. They were discovered from **f** via the intermediate structure **f-2**, which was unstable in the gas phase. The green arrows indicate the movement of the hydrogen atoms. The transition barriers are provided in Table 2.
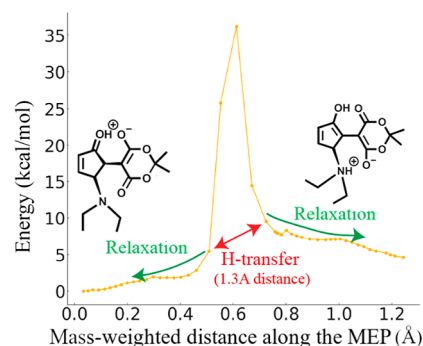
Table 2. Reaction Barriers along the Pathways Leading From f in Figure 11, Calculated via the NEB Approach Using the ωB97x-D3/6-31G*/MM Method in the Aqueous Solution

transitions	barrier height relative to **f** (kcal/mol)
f to **f-2**	2.6
f-2 to **f-3**	36.5
f-2 to **f-4**	18.3
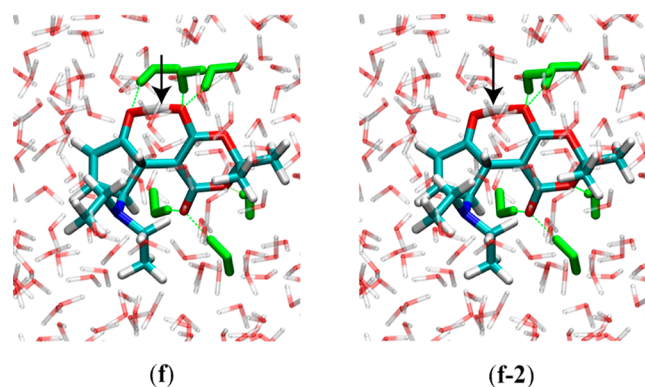
Figure 12. DASA molecule in the **f** and **f-2** minima form hydrogen bonds with water molecules, which stabilize the carbonyl group, enabling the hydrogen atom (highlighted by black arrows) to remain at its position near the hydroxyl group during the cyclization of the DASA's triene chain.

Figure 13. MEP for the **f-2** to **f-3** transition determined by NEB calculation. The barrier is 36.5 kcal/mol, which arises from the hydrogen transfer reaction. The hydrogen moves a distance of 1.3 Å between the NEB images indicated by the red arrows, while the rest of the molecular geometry remains largely unchanged between these two images.

■ CONCLUSION AND DISCUSSION

In summary, we developed a new version of *Crystal* algorithm that is well-suited for systematically exploring chemical reactions in systems with numerous degrees of freedom, positioning *Crystal* as a powerful complement to MD-based approaches. Large and complex molecular systems, such as bilirubin and Meldrum's acid DASA, especially when embedded in condensed-phase environments, are well beyond *Crystal*'s previous size limit for full-dimensional exploration. However, the new version of the *Crystal* algorithm can nonetheless successfully reproduce data from prior studies and uncover additional new reaction pathways. This was mainly achieved via the effective relaxation of implicit d.o.f.s in response to the explicit d.o.f.s. The multiple technical advancements greatly extend *Crystal*'s applicability to reaction discovery in large, complex and realistic molecular systems in the condensed phase, and enable it to compete with, and even surpass, conventional methods in revealing complex features of the PES.

While the selection of explicit d.o.f.s remains critical for the successful discovery of reactions, *Crystal* consistently exhibits

bonds. Among Z−Z isomers, the most stable configurations involve 4−6 hydrogen bonds. The differences between the lowest three configurations in Figure 15 are minor. The most stable E−Z/Z−E isomers show 2−4 hydrogen bonds, while the most stable E−E configurations have 1−2 hydrogen bonds. While the importance of rotations about dihedrals 1, 3, 4, 6 (defined in Figure 6) has been well-established in literature, the inclusion of dihedrals 2 and 5 (defined in Figure 6) in *Crystal*'s explicit d.o.f.s leads to a more thorough exploration.
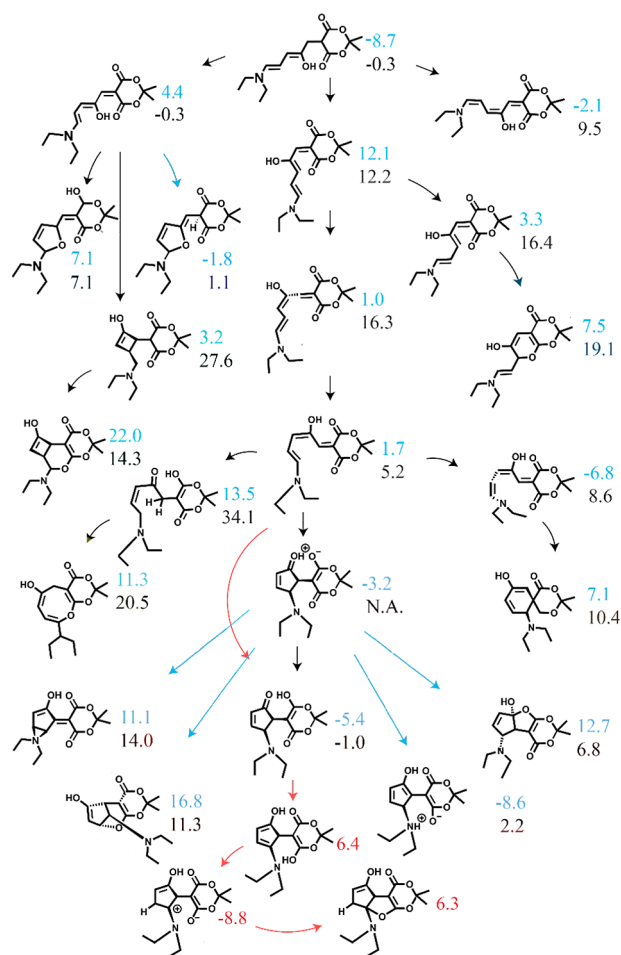
Figure 14. Summary of discovered geometries of Meldrum's acid DASA using *Crystal*. The arrows indicate the sequence in which *Crystal* identified the geometries. Blue arrows indicate geometries only discovered during the exploration in the aqueous solution. Red arrows indicate those discovered only during the exploration in implicit solvent. Black arrows indicate geometries found during both these explorations. Linear geometries (formed without cyclization reaction) were grouped by the isomeric form, while cyclic geometries (formed with at least one cyclization reaction) were grouped by the conformational isomer. For the lowest geometry in each group, the energy values are provided in kcal/mol. The blue numbers indicate energies in the explicit solvent. The black numbers indicate energies in the gas phase. The red numbers indicate the energies in implicit solvent. All energies are relative to the reference structures in Figure 10. The energies were obtained via refinement at the ωB97x-D3/def2-TZVP(-f)/COSMO level of theory for the implicit solvent calculations, and at the ωB97x-D3/6-31G* level of theory for all other cases.

the ability to identify unexpected reaction paths beyond the initially selected subspace. This is a very key point which must be emphasized. Since the d.o.f.s associated with these newly discovered reaction paths can, in principle, be recursively incorporated into the explicit space, the *Crystal* algorithm holds considerable potential for automation. However, *Crystal*'s success in identifying reactions beyond the initially chosen explicit lattice space likely relies on the coupling or overlap between these reactions and undiscovered pathways, as illustrated by the following observations.

In one instance, hydrogen hopping was observed in gas-phase DASA (species **c** in Figure 7) without explicitly
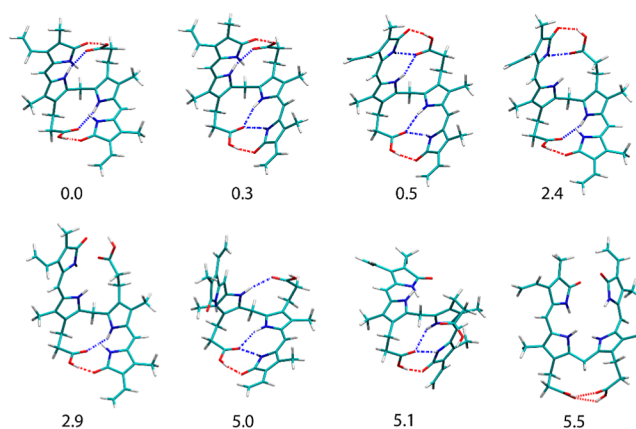
Figure 15. Unique hydrogen bond patterns identified by *Crystal* among the lowest-energy Z−Z isomers of bilirubin, with their energy values provided in kcal/mol. The calculations were done at the PM6-FOMO-CASCI level of theory. The E−Z and E−E isomers are provided in Figures S4 and S5.

including this hydrogen in *Crystal*'s lattice space. This occurred because the cyclization of the triene chain eliminated the energy barrier for hydrogen transfer from the hydroxyl to the carbonyl group, thereby enabling the hop. In another instance, including this hydrogen in *Crystal*'s lattice space revealed a previously unidentified double-hydrogen transfer (species **c-1**). The addition of the first hydrogen to *Crystal*'s explicit space facilitated the simultaneous hopping of the second hydrogen, as its return from the carbonyl to the hydroxyl group overlapped with the double-hydrogen transfer process.

The inherent complexity of this low-energy double-hydrogen transfer likely explains why it evaded detection by other approaches. The return of the first hydrogen from the acceptor to the triene oxygen is energetically unfavorable, as it lacks a stable energy minimum unless the second hydrogen transfers simultaneously in the opposite direction. The discovery of this pathway underscores the distinct advantage of *Crystal* in revealing intricate reaction mechanisms, which can complement existing powerful tools such as metadynamics.

Finally, the exploration of bilirubin revealed numerous hydrogen bond connectivity patterns that play an important role in stabilizing its isomeric forms. Explicitly involving a total of six degrees of freedom, this exploration covers a larger subspace than any other bilirubin study in the literature to our knowledge.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.5c00324.

> The Supporting Information contains a comprehensive discussion of the motivation, rationale and technical details of the development over the previous version of the *Crystal* algorithm, with benchmark of its computational efficiency. It also contains Tables S1−S3, summarizing the lowest-energy minima of the bilirubin in Z−Z, Z−E, E−Z, and E−E isomeric forms. Figures S4−S5 summarizes the geometries of lowest-energy minima of the bilirubin in the Z−E, E−Z, and E−E isomeric forms (PDF)

> Discovered minima of DASA and bilirubin (ZIP)

■ AUTHOR INFORMATION

Corresponding Authors

Ruibin Liang − *Department of Chemistry and Biochemistry, Texas Tech University, Lubbock, Texas 79409, United States;* orcid.org/0000-0001-8741-1520; Email: rliang@ttu.edu

Bill Poirier − *Department of Chemistry and Biochemistry, Texas Tech University, Lubbock, Texas 79409, United States;* orcid.org/0000-0001-8277-746X; Email: Bill.Poirier@ttu.edu

Authors

Ankit Pandey − *Department of Chemistry and Biochemistry, Texas Tech University, Lubbock, Texas 79409, United States*

Gustavo J. Costa − *Department of Chemistry and Biochemistry, Texas Tech University, Lubbock, Texas 79409, United States*

Mushfiq Alam − *Department of Chemistry and Biochemistry, Texas Tech University, Lubbock, Texas 79409, United States*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jctc.5c00324

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Wang, L.-P.; Titov, A.; McGibbon, R.; Liu, F.; Pande, V. S.; Martínez, T. J. Discovering chemistry with an ab initio nanoreactor. *Nat. Chem.* **2014**, *6* (12), 1044−1048.

(2) Maeda, S.; Harabuchi, Y.; Taketsugu, T.; Morokuma, K. Systematic Exploration of Minimum Energy Conical Intersection Structures near the Franck−Condon Region. *J. Phys. Chem. A* **2014**, *118* (51), 12050−12058.

(3) Maeda, S.; Taketsugu, T.; Ohno, K.; Morokuma, K. From Roaming Atoms to Hopping Surfaces: Mapping Out Global Reaction Routes in Photochemistry. *J. Am. Chem. Soc.* **2015**, *137* (10), 3433−3445.

(4) Aldaz, C.; Kammeraad, J. A.; Zimmerman, P. M. Discovery of conical intersection mediated photochemistry with growing string methods. *Phys. Chem. Chem. Phys.* **2018**, *20* (43), 27394−27405.

(5) Lindner, J. O.; Sultangaleeva, K.; Röhr, M. I. S.; Mitrić, R. metaFALCON: A Program Package for Automatic Sampling of Conical Intersection Seams Using Multistate Metadynamics. *J. Chem. Theory Comput.* **2019**, *15* (6), 3450−3460.

(6) Meisner, J.; Zhu, X.; Martínez, T. J. Computational Discovery of the Origins of Life. *ACS Cent. Sci.* **2019**, *5* (9), 1493−1495.

(7) Pieri, E.; Lahana, D.; Chang, A. M.; Aldaz, C. R.; Thompson, K. C.; Martínez, T. J. The non-adiabatic nanoreactor: towards the automated discovery of photochemistry. *Chem. Sci.* **2021**, *12* (21), 7294−7307.

(8) Sanchez, D. M.; Raucci, U.; Martínez, T. J. In Silico Discovery of Multistep Chemistry Initiated by a Conical Intersection: The Challenging Case of Donor−Acceptor Stenhouse Adducts. *J. Am. Chem. Soc.* **2021**, *143* (48), 20015−20021.

(9) Raucci, U.; Sanchez, D. M.; Martínez, T. J.; Parrinello, M. Enhanced Sampling Aided Design of Molecular Photoswitches. *J. Am. Chem. Soc.* **2022**, *144* (42), 19265−19271.

(10) Raucci, U.; Rizzi, V.; Parrinello, M. Discover, Sample, and Refine: Exploring Chemistry with Enhanced Sampling Techniques. *J. Phys. Chem. Lett.* **2022**, *13* (6), 1424−1430.

(11) Xu, R.; Meisner, J.; Chang, A. M.; Thompson, K. C.; Martínez, T. J. First principles reaction discovery: from the Schrodinger equation to experimental prediction for methane pyrolysis. *Chem. Sci.* **2023**, *14* (27), 7447−7464.

(12) Chang, A. M.; Meisner, J.; Xu, R.; Martínez, T. J. Efficient Acceleration of Reaction Discovery in the Ab Initio Nanoreactor: Phenyl Radical Oxidation Chemistry. *J. Phys. Chem. A* **2023**, *127* (45), 9580−9589.

(13) Pandey, A.; Poirier, B.; Liang, R. Development of Parallel On-the-Fly Crystal Algorithm for Global Exploration of Conical Intersection Seam Space. *J. Chem. Theory Comput.* **2024**, *20* (11), 4778−4789.

(14) Türtscher, P. L.; Reiher, M. Pathfinder−Navigating and Analyzing Chemical Reaction Networks with an Efficient Graph-Based Approach. *J. Chem. Inf. Model.* **2023**, *63* (1), 147−160.

(15) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100* (2), 020603.

(16) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99* (20), 12562−12566.

(17) Invernizzi, M.; Parrinello, M. Rethinking Metadynamics: From Bias Potentials to Probability Distributions. *J. Phys. Chem. Lett.* **2020**, *11* (7), 2731−2736.

(18) Zimmerman, P. M. Growing string method with interpolation and optimization in internal coordinates: method and examples. *J. Chem. Phys.* **2013**, *138* (18), 184102.

(19) Dewyer, A. L.; Zimmerman, P. M. Finding reaction mechanisms, intuitive or otherwise. *Org. Biomol. Chem.* **2017**, *15* (3), 501−504.

(20) Jafari, M.; Zimmerman, P. M. Uncovering reaction sequences on surfaces through graphical methods. *Phys. Chem. Chem. Phys.* **2018**, *20* (11), 7721−7729.

(21) Stepanenko, S.; Engels, B. Tabu Search Based Strategies for Conformational Search. *J. Phys. Chem. A* **2009**, *113* (43), 11699−11705.

(22) Grebner, C.; Becker, J.; Weber, D.; Engels, B. Tabu search based global optimization algorithms for problems in computational chemistry. *J. Chem. Inf.* **2012**, *4* (1), P10.

(23) Grebner, C.; Kästner, J.; Thiel, W.; Engels, B. A New Tabu-Search-Based Algorithm for Solvation of Proteins. *J. Chem. Theory Comput.* **2013**, *9* (1), 814−821.

(24) Maeda, S.; Taketsugu, T.; Morokuma, K.; Ohno, K. Anharmonic Downward Distortion Following for Automated Exploration of Quantum Chemical Potential Energy Surfaces. *Bull. Chem. Soc. Jpn.* **2014**, *87* (12), 1315−1334.

(25) Töpfer, K.; Vazquez-Salazar, L. I.; Meuwly, M. Asparagus: A toolkit for autonomous, user-guided construction of machine-learned potential energy surfaces. *Comput. Phys. Commun.* **2025**, *308*, 109446.

(26) Aarabi, M.; Pandey, A.; Poirier, B. "On-the-fly" Crystal: How to reliably and automatically characterize and construct potential energy surfaces. *J. Comput. Chem.* **2024**, *45* (15), 1261−1278.

(27) Pandey, A.; Poirier, B. Using phase-space Gaussians to compute the vibrational states of OCHCO+. *J. Chem. Phys.* **2019**, *151* (1), 014114.

(28) Pandey, A.; Poirier, B. An algorithm to find (and plug) "holes" in multi-dimensional surfaces. *J. Chem. Phys.* **2020**, *152* (21), 214102.

(29) Pandey, A.; Poirier, B. Plumbing Potentials for Molecules with Up To Tens of Atoms: How to Find Saddle Points and Fix Leaky Holes. *J. Phys. Chem. Lett.* **2020**, *11* (15), 6468−6474.

(30) Maeda, S.; Ohno, K.; Morokuma, K. Automated Global Mapping of Minimal Energy Points on Seams of Crossing by the Anharmonic Downward Distortion Following Method: A Case Study of H2CO. *J. Phys. Chem. A* **2009**, *113* (9), 1704−1710.

(31) Maeda, S.; Saito, R.; Morokuma, K. Finding Minimum Structures on the Seam of Crossing in Reactions of Type A + B → X: Exploration of Nonadiabatic Ignition Pathways of Unsaturated Hydrocarbons. *J. Phys. Chem. Lett.* **2011**, *2* (8), 852−857.

(32) Maeda, S.; Ohno, K.; Morokuma, K. Systematic exploration of the mechanism of chemical reactions: the global reaction route mapping (GRRM) strategy using the ADDF and AFIR methods. *Phys. Chem. Chem. Phys.* **2013**, *15* (11), 3683−3701.

(33) Harabuchi, Y.; Maeda, S.; Taketsugu, T.; Minezawa, N.; Morokuma, K. Automated Search for Minimum Energy Conical Intersection Geometries between the Lowest Two Singlet States S0/S1-MECIs by the Spin-Flip TDDFT Method. *J. Chem. Theory Comput.* **2013**, *9* (9), 4116−4123.

(34) Klán, P.; Wirz, J. Physicochemical Aspects of Photoswitching. In *Molecular Photoswitches*; Wiley, 2022; pp 1−18.

(35) Welleman, I. M.; Hoorens, M. W. H.; Feringa, B. L.; Boersma, H. H.; Szymański, W. Photoresponsive molecular tools for emerging applications of light in medicine. *Chem. Sci.* **2020**, *11* (43), 11672−11691.

(36) Pianowski, Z. L. Recent Implementations of Molecular Photoswitches into Smart Materials and Biological Systems. *Chem.—Eur. J.* **2019**, *25* (20), 5128−5144.

(37) Boase, N. R. B. Shining a Light on Bioorthogonal Photochemistry for Polymer Science. *Macromol. Rapid Commun.* **2020**, *41* (18), 2000305.

(38) Boelke, J.; Hecht, S. Designing Molecular Photoswitches for Soft Materials Applications. *Adv. Opt. Mater.* **2019**, *7* (16), 1900404.

(39) Goulet-Hanssens, A.; Eisenreich, F.; Hecht, S. Enlightening Materials with Photoswitches. *Adv. Mater.* **2020**, *32* (20), 1905966.

(40) Helmy, S.; Oh, S.; Leibfarth, F. A.; Hawker, C. J.; Read de Alaniz, J. Design and Synthesis of Donor−Acceptor Stenhouse Adducts: A Visible Light Photoswitch Derived from Furfural. *J. Org. Chem.* **2014**, *79* (23), 11316−11329.

(41) Helmy, S.; Leibfarth, F. A.; Oh, S.; Poelma, J. E.; Hawker, C. J.; Read de Alaniz, J. Photoswitching Using Visible Light: A New Class of Organic Photochromic Molecules. *J. Am. Chem. Soc.* **2014**, *136* (23), 8169−8172.

(42) Clerc, M.; Sandlass, S.; Rifaie-Graham, O.; Peterson, J. A.; Bruns, N.; Read de Alaniz, J.; Boesel, L. F. Visible light-responsive materials: the (photo)chemistry and applications of donor−acceptor Stenhouse adducts in polymer science. *Chem. Soc. Rev.* **2023**, *52* (23), 8245−8294.

(43) Stricker, F.; Sanchez, D. M.; Raucci, U.; Dolinski, N. D.; Zayas, M. S.; Meisner, J.; Hawker, C. J.; Martínez, T. J.; Read de Alaniz, J. A multi-stage single photochrome system for controlled photoswitching responses. *Nat. Chem.* **2022**, *14* (8), 942−948.

(44) Bersuker, I. *The Jahn-Teller Effect*; Cambridge University Press, 2006. DOI: DOI: .

(45) Mallo, N.; Foley, E. D.; Iranmanesh, H.; Kennedy, A. D. W.; Luis, E. T.; Ho, J.; Harper, J. B.; Beves, J. E. Structure−function relationships of donor−acceptor Stenhouse adduct photochromic switches. *Chem. Sci.* **2018**, *9* (43), 8242−8252.

(46) Chen, Y.; Gao, S.; Cheng, Y.; Tian, X.; Xuan, X.; Wang, H.; Yao, W.; Li, Z.; Zhu, G.; Wang, J. Isomerization of DASA Molecules in the Nanopores of Metal−Organic Frameworks: What Determines Its Reversibility? *Chem.—Eur. J.* **2023**, *29* (27), No. e202300222.

(47) Berraud-Pache, R.; Santamaría-Aranda, E.; de Souza, B.; Bistoni, G.; Neese, F.; Sampedro, D.; Izsák, R. Redesigning donor−acceptor Stenhouse adduct photoswitches through a joint experimental and computational study. *Chem. Sci.* **2021**, *12* (8), 2916−2924.

(48) Fiorentino, A.; Sachini, B.; Corra, S.; Credi, A.; Femoni, C.; Fraix, A.; Silvi, S. Acidochromism of donor−acceptor Stenhouse adducts in organic solvent. *Chem. Commun.* **2022**, *58* (80), 11236−11239.

(49) Wang, D.; Zhao, L.; Zhao, H.; Wu, J.; Wagner, M.; Sun, W.; Liu, X.; Miao, M.-s.; Zheng, Y. Inducing molecular isomerization assisted by water. *Commun. Chem.* **2019**, *2* (1), 118.

(50) Busch, A. W. U.; Montgomery, B. L. Interdependence of tetrapyrrole metabolism, the generation of oxidative stress and the mitigative oxidative stress response. *Redox Biol.* **2015**, *4*, 260−271.

(51) Kräutler, B. Phyllobilins − the abundant bilin-type tetrapyrrolic catabolites of the green plant pigment chlorophyll. *Chem. Soc. Rev.* **2014**, *43* (17), 6227−6238.

(52) Warren, M. J.; Tetrapyrroles, A. G. S. *Birth, Life and Death*; Springer Landes Bioscience, 2009.

(53) Mancuso, C. Bilirubin and brain: A pharmacological approach. *Neuropharmacology* **2017**, *118*, 113−123.

(54) Dohi, K.; Satoh, K.; Ohtaki, H.; Shioda, S.; Miyake, Y.; Shindo, M.; Aruga, T. Elevated Plasma Levels of Bilirubin in Patients with Neurotrauma Reflect its Pathophysiological Role in Free Radical Scavenging. *Vivo* **2005**, *19* (5), 855.

(55) Upadhyaya, H. P. Computational Characterization of "Dark" Intermediates in the Ultrafast Deactivation of Photoexcited Bilirubin. *J. Phys. Chem. A* **2018**, *122* (46), 9084−9092.

(56) Agati, G.; Fusi, F. New trends in photobiology recent advances in bilirubin photophysics. *J. Photochem. Photobiol., B* **1990**, *7* (1), 1−14.

(57) Alkén, J.; Håkansson, S.; Ekéus, C.; Gustafson, P.; Norman, M. Rates of Extreme Neonatal Hyperbilirubinemia and Kernicterus in Children and Adherence to National Guidelines for Screening, Diagnosis, and Treatment in Sweden. *JAMA Netw. Open* **2019**, *2* (3), No. e190858.

(58) Cremer, R. J.; Perryman, P. W.; Richards, D. H. INFLUENCE OF LIGHT ON THE HYPERBILIRUBINÆMIA OF INFANTS. *Lancet* **1958**, *271* (7030), 1094−1097.

(59) Carreira-Blanco, C.; Singer, P.; Diller, R.; Luis Pérez Lustres, J. Ultrafast deactivation of bilirubin: dark intermediates and two-photon isomerization. *Phys. Chem. Chem. Phys.* **2016**, *18* (10), 7148−7155.

(60) Lightner, D. A.; Wooldridge, T. A.; McDonagh, A. F. Photobilirubin: an early bilirubin photoproduct detected by absorbance difference spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.* **1979**, *76* (1), 29−32.

(61) McDonagh, A. F.; Palma, L. A.; Lightner, D. A. Blue Light and Bilirubin Excretion. *Science* **1980**, *208* (4440), 145−151.

(62) Lightner, D. A.; McDonagh, A. F. Molecular mechanisms of phototherapy for neonatal jaundice. *Acc. Chem. Res.* **1984**, *17* (12), 417−424.

(63) Invernizzi, M.; Piaggi, P. M.; Parrinello, M. Unified Approach to Enhanced Sampling. *Phys. Rev. X* **2020**, *10* (4), 041034.

(64) Ben-Nun, M.; Martínez, T. J. Ab Initio Quantum Molecular Dynamics. In *Advances in Chemical Physics*; John Wiley & Sons, Inc., 2002; pp 439−512.

(65) Liang, R.; Cotton, S. J.; Binder, R.; Hegger, R.; Burghardt, I.; Miller, W. H. The symmetrical quasi-classical approach to electronically nonadiabatic dynamics applied to ultrafast exciton migration processes in semiconducting polymers. *J. Chem. Phys.* **2018**, *149* (4), 044101.

(66) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.* **2019**, *15* (3), 1652−1671.

(67) Lin, Y.-S.; Li, G.-D.; Mao, S.-P.; Chai, J.-D. Long-Range Corrected Hybrid Density Functionals with Improved Dispersion Corrections. *J. Chem. Theory Comput.* **2013**, *9* (1), 263−272.

(68) Klamt, A.; Schüürmann, G. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc., Perkin Trans. 1* **1993**, *2* (5), 799−805.

(69) Zhu, X.; Thompson, K. C.; Martínez, T. J. Geodesic interpolation for reaction pathways. *J. Chem. Phys.* **2019**, *150* (16), 164103.

(70) Henkelman, G.; Uberuaga, B. P.; Jónsson, H. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys.* **2000**, *113* (22), 9901−9904.

(71) Titov, A. V.; Ufimtsev, I. S.; Luehr, N.; Martinez, T. J. Generating Efficient Quantum Chemistry Codes for Novel Architectures. *J. Chem. Theory Comput.* **2013**, *9* (1), 213−221.

(72) Ufimtsev, I. S.; Martinez, T. J. Quantum Chemistry on Graphical Processing Units. 3. Analytical Energy Gradients, Geometry Optimization, and First Principles Molecular Dynamics. *J. Chem. Theory Comput.* **2009**, *5* (10), 2619−2628.

(73) 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* (15), 154104.

(74) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32* (7), 1456−1465.

(75) Wang, L.-P.; Song, C. Geometry optimization made simple with translation and rotation coordinates. *J. Chem. Phys.* **2016**, *144* (21), 214108.

(76) Kästner, J.; Carr, J. M.; Keal, T. W.; Thiel, W.; Wander, A.; Sherwood, P. D. L.-F. I. N. D. DL-FIND: An Open-Source Geometry Optimizer for Atomistic Simulations. *J. Phys. Chem. A* **2009**, *113* (43), 11856−11865.

(77) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25* (9), 1157−1174.

(78) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph Model.* **2006**, *25* (2), 247−260.

(79) Wu, Y. J.; Tepper, H. L.; Voth, G. A. Flexible simple point-charge water model with improved liquid-state properties. *J. Chem. Phys.* **2006**, *124* (2), 024503.

(80) Neese, F.; Wennmohs, F.; Becker, U.; Riplinger, C. The ORCA quantum chemistry program package. *J. Chem. Phys.* **2020**, *152* (22), 224108.

(81) Metz, S.; Kästner, J.; Sokol, A. A.; Keal, T. W.; Sherwood, P. ChemShell—a modular software package for QM/MM simulations. *Wiley Interdiscip. Rev.:Comput. Mol. Sci.* **2014**, *4* (2), 101−110.

(82) Bakhtiiari, A.; Costa, G. J.; Liang, R. On the Simulation of Thermal Isomerization of Molecular Photoswitches in Biological Systems. *J. Chem. Theory Comput.* **2023**, *19* (18), 6484−6499.

(83) Stewart, J. J. P. Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *J. Mol. Model.* **2007**, *13* (12), 1173−1213.

(84) Slavíček, P.; Martínez, T. J. Ab initio floating occupation molecular orbital-complete active space configuration interaction: An efficient approximation to CASSCF. *J. Chem. Phys.* **2010**, *132* (23), 234102.

(85) Klajn, R. Spiropyran-based dynamic materials. *Chem. Soc. Rev.* **2014**, *43* (1), 148−184.

(86) Cruz, A.·. B.; de Brito, L. G.; Leal, P. V. B.; dos Santos Ramos, W. T.; Pereira, D. H. Intramolecular hydrogen bonds interactions in the isomers of the bilirubin molecule: DFT and QTAIM analysis. *J. Mol. Model.* **2023**, *29* (10), 318.