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

Ankit Pandey[1], Gustavo J. Costa[1], Mushfiq Alam[1], Bill Poirier[1,2]*, and Ruibin Liang[1]*

[1]Department of Chemistry and Biochemistry, Texas Tech University, Lubbock, TX, 79409, USA

[2]Department of Chemistry, University of Vermont, Burlington, VT, 05405, USA

AUTHOR INFORMATION

Corresponding Author

*Ruibin Liang

Email address: rliang@ttu.edu

*Bill Poirier

Email address: Bill.Poirier@uvm.edu

Abstract

The parallel on-the-fly *Crystal* algorithm is a new, efficient global search algorithm that has found utility in exploring the single-state potential energy surfaces and conical intersection seam spaces of a wide range of molecules in the gas phase. Despite its 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, complex molecular photoswitches in the condensed phase. This is achieved by improved search algorithms that facilitate the exploration of a comparatively large *Crystal* configurational subspace while relaxing the other degrees of freedom. The new *Crystal* algorithm is applied to bilirubin and a next-generation class of molecular photoswitches in the vacuum and the aqueous solution environment in a quantum mechanics/molecular mechanics (QM/MM) setting. We designed an automatic and systematic workflow to discover previously unreported minima and low-energy reaction pathways and analyze them with a focus on the low-energy spectrum. Our findings reveal the algorithm's effectiveness in quickly exploring the configuration space, uncovering new product minima that are kinetically accessible, which provides new insights into the intricate chemical reactivities of these molecules. The results underscore the promising potential of parallelized global exploration methods in reaction discovery in biomolecular systems.

Introduction

Reaction discovery through molecular modeling is an emerging field that has seen significant progress in recent years.[1-13] Computational predictions of reaction products and pathways in complex molecular systems have become useful for interpreting and guiding new experiments. Central to the reaction discovery simulation is the efficient exploration of the low-energy regions

of the potential energy surfaces (PES). To achieve this goal, a class of approaches utilized trajectories to dynamically sample the configuration space of the 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,[14-16] or changing the global temperature and pressure of the system, such as the nano-reactors.[7] Another class of approaches does not involve running trajectories, and they can overcome reaction barriers and locate new product minima by evaluating the energies, gradients and Hessians at existing static geometries, such as the growing string method.[17-19] Among these algorithms, some methods were designed to attempt a global search of the PES, such as the anharmonic downward distortion approach[20] and the *Crystal* algorithm.[13, 21-24] In principle, the advantage of the global algorithm is the exhaustive search in the configurational space and identify all minima and transition states. The global nature of the algorithm enables the identification of new products and low-energy pathways that are usually not sampled by trajectory-based approaches. This is because the search results of global algorithms do not depend so sensitively on the choice of initial structure and the set of selected collective variables as trajectory-based enhanced sampling simulations usually do.

The *Crystal* algorithm[13, 21-24] is based on the simple idea of performing an exhaustive search among all combinations of discretized moves of all selected reaction coordinates while satisfying a maximal energy limit. The moves are performed in a high-dimensional lattice space spanned by these reaction coordinates, with equally-spaced lattice points in each dimension. Notably, the number of lattice points with energies below a certain preset value only scales *polynomially* with respect to the 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.

Recently, the *Crystal* algorithm has been under major developments. One major progress was its coupling with on-the-fly electronic structure calculations of potential energies[21], thus eliminating the need to obtain a predetermined global PES of the system prior to the search. Also, its efficiency has been greatly increased by an improved search algorithm and the implementation of a massively parallelizable version of the code[13]. It has also been generalized to search for low-energy conformations in the conical intersection seam space[13] in addition to single-state PES, which facilitates photochemical reaction discovery. All this progress has made the *Crystal* algorithm applicable to reaction discovery in realistic systems with increasing sizes. Although other global search algorithms have been recently developed, such as the anharmonic-downward-distortion based approaches developed by Morokuma et.al[2, 25-28], *Crystal* holds the unique advantages that it does not require Hessian evaluations and that it is massively parallelizable.

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 rapidly increases very quickly as the number of d.o.f.s spanning the lattice space grows. Currently, the *Crystal* algorithm cannot efficiently search in the 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. in the *Crystal* lattice space, even with relatively inexpensive quantum chemistry methods such as the semiempirical methods. This is largely due to the existence of an enormous number of low-lying

lattice points in the high-dimensional space, as well as the slower speed at which the PES of larger systems is evaluated on-the-fly using electronic structure calculations.

In this paper, we develop the *Crystal* algorithm to significantly enhance its capability of reaction discovery in large and complex molecular systems, including condensed phase systems. The central new idea behind the development is to allow the d.o.f.s outside the *Crystal*'s configuration subspace (lattice space) to relax in response to the changes of the d.o.f.s spanning the subspace. In other words, this new approach enables the exhaustive sampling of all d.o.f.s that are essential to the chemical events while letting the rest of the system adjust to the motions of these d.o.f.s. There are several key technical advantages of this new approach over the previous-generation one, which fixed the values of d.o.f.s outside the *Crystal*'s subspace. The first is that the new algorithm enables the use of larger steps of *Crystal* moves, i.e., larger lattice spacing, making the conformational exploration more efficient while effectively avoiding steric clashes caused by d.o.f.s outside 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 of *Crystal* moves. This advantage is critical for retaining the capability of the *Crystal* algorithm to faithfully discover kinetically relevant reaction pathways below a certain limit of energy barrier. 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 the condensed phase molecular systems.

All the above-mentioned advances greatly expand the applicability of *Crystal* towards reaction discovery in larger molecular systems of importance in biomedical and material sciences. To test

the effectiveness of this new method, we applied it to the 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[29]. They can be covalently or non-covalently embedded in large molecular systems and make their physical and chemical properties photoresponsive. Their scope of applications has been growing in the fields of photopharmacology to photoresponsive materials[30-34], where light can be used to control the activity of drugs in target biomolecules and change 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[35-37]. DASAs are "multi-step" photoswitches, with multiple isomers that can photoswitch into one another reversibly, thus opening the possibility of complex molecular manipulation[38]. The photoswitching of DASAs is mediated by conical intersections[8, 39] that facilitate molecular de-excitation to different photoisomers through various pathways. Understanding these pathways and identifying "unproductive" pathways that cause fatigue-induced damage, i.e., loss of photoisomer yield, requires the knowledge of the relevant stable and meta-stable states that can be accessed by the molecule[8, 40, 41]. 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, Rauci et al.[9] employed metadynamics to discover many previously unknown conformers of DASA on the ground state.

The Meldrum's acid DASA derivative is unstable in aqueous environments, so it is typically studied in non-polar solvents. Computational studies of DASAs have been performed in various

environments, including the gas phase[42], toluene[9] and trifluoromethanesulfonic acid [43]. However, there is growing interest in water-soluble DASA molecules. Water-assisted isomerization has been explored in ref [44]. Studying the Meldrum's acid DASA derivative in water can reveal key interactions between water molecules and DASA compounds. This may aid in the understanding of solubility and dissociation mechanisms, designing stable water-insoluble DASAs, and predicting water's effect on photoisomerization. These insights could guide the development and tailoring of DASA derivatives with desirable properties.

Another class of biologically significant molecules in which photoactivity plays a crucial role is tetrapyrrole, including chlorophyll, porphyrins, phycobilins, and bilirubin.[45-48] Among these, bilirubin plays an important role in human biology, exhibiting neuroprotective effects at physiological concentrations.[49] 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 liver makes it water-soluble, facilitating excretion.[50, 51] However, immature liver function in newborns can hamper this, leading to kernicterus.[52] Light therapy is used to treat this condition by photo-isomerizing bilirubin, making it water-soluble and easier to eliminate.[53] In its native Z,Z isomer form, bilirubin is stabilized by six intramolecular hydrogen bonds. Photo-isomerization disrupts some of these bonds, allowing water to form hydrogen bonds with bilirubin. The isomerization of bilirubin is known to proceed via dark intermediates.[54] The reverse-isomerization of its photo-isomers is also slow in the dark. A smaller portion of bilirubin is irreversibly converted to water-soluble lumirubin, aiding its removal from

the body.[55-57] Bilirubin is also structurally closely related to several key chromophores in a variety of natural photoreceptors, such as the 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 derivative in gas phase, implicit solvent, and condensed phase, as well as bilirubin in the gas phase. The results demonstrate the excellent performance of this algorithm for reaction discovery in large and complex biomolecular systems of practical interest in biomedical and material science.

Methods

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 semiempirical methods.[13, 21] 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 **Fig. 1**. For all lattice points in the list, neighbors that have not yet been evaluated are added to an evaluation queue. Secondary jobs manage the evaluation process 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 reduced the computational cost and made it scale polynomially instead of exponentially with respect to the dimensionally of the subspace.

(a)	(b)	(c)
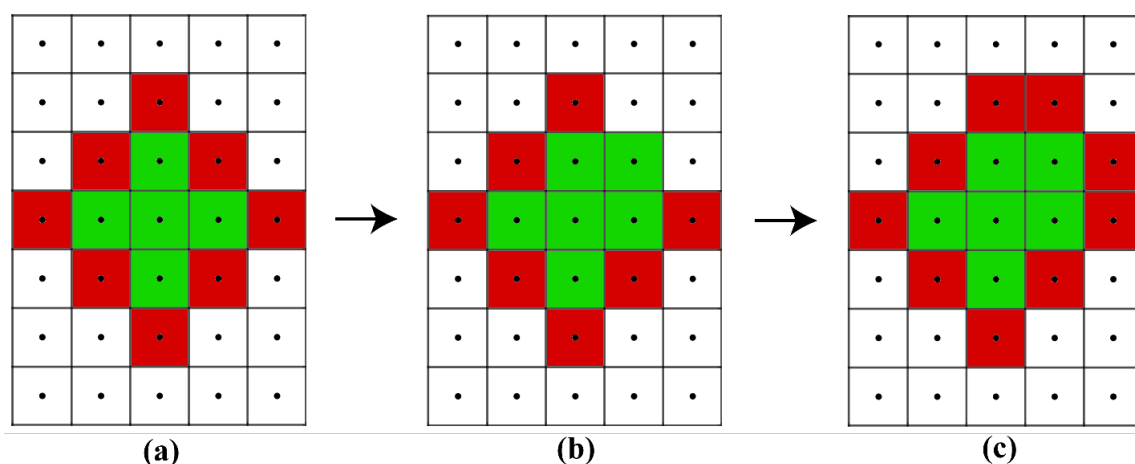
Fig. 1. Overview of the *Crystal* algorithm. (a) Starting configuration (green lattice point) for the search represented in the lattice space. The red lattice points denote those that are either queued

for evaluation or are currently being evaluated by parallel secondary jobs. Once one of these points is evaluated (b), its unexamined neighbors (c) are added to the evaluation queue. The secondary jobs work independently and start evaluating the next point in the queue after they finish their current task.

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, 21-24], *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, which could have been rejected using a small enough lattice spacing. This can lead to the discovery of reaction pathways that are kinetically inaccessible and many unwanted minima with little physical relevance. 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 the computational efficiency.

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 improved 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 timescale separation that has been extensively exploited in enhanced sampling simulations with collective variables (CV) and trajectory dynamics (such as umbrella sampling). 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.

We hypothesize that adopting such a scheme in *Crystal* can 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 makes the algorithm more tolerant to the use of large lattice spacing, which further reduces the computational cost.

In order 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 and 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 (**Fig. 2**). For instance, if a single *Crystal* move dictates that a torsion in the explicit d.o.f.s be changed from 0 to 60 degrees, it is realized by a relaxed scan of the system with that torsion gradually increased at 15-degrees intervals towards 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. Furthermore, the likelihood of missing minima is decreased, as relaxing the d.o.f.s. outside the lattice space helps guide molecular configurations toward their minimum-energy configurations. Importantly, the computational cost of this additional step scales linearly instead of exponentially with respect to the number of intermediate points in the relaxed scan for each explicit d.o.f. A flowchart illustrating the overall scheme of the new *Crystal* algorithm is provided in **Fig. 3**.
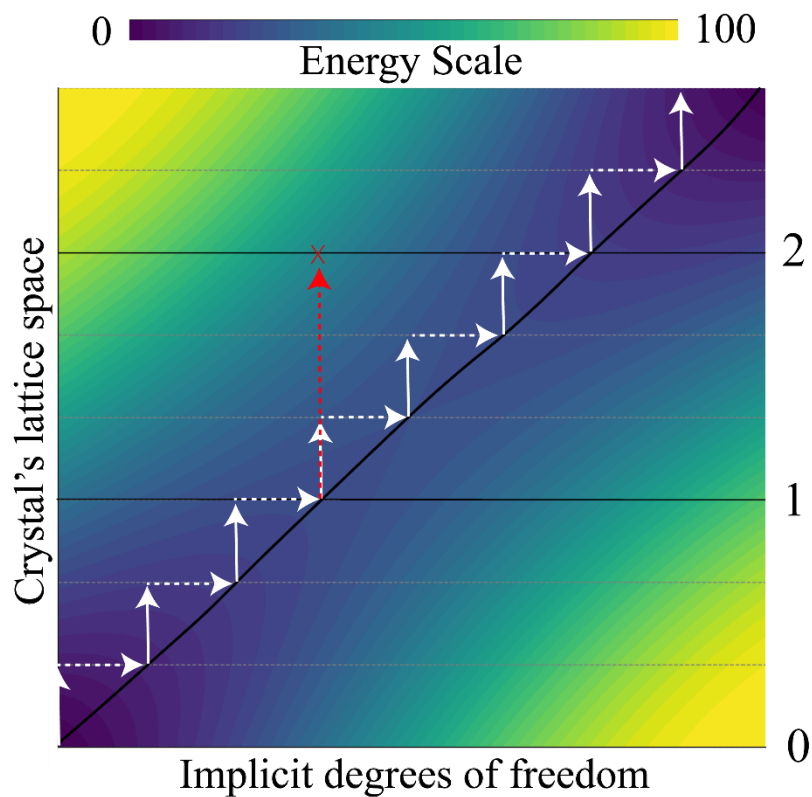
Fig 2. Attempting large *Crystal* moves in the explicit d.o.f.s (e.g., from 1 to 2, red arrow) can result in high energy geometries and/or inadvertent stepping over hidden high energy barriers. To prevent this, d.o.f.s outside the lattice space (implicit d.o.f.s) are optimized gradually (dotted white arrows) during the relaxed scan of the explicit d.o.f.s at small intervals (solid arrows).
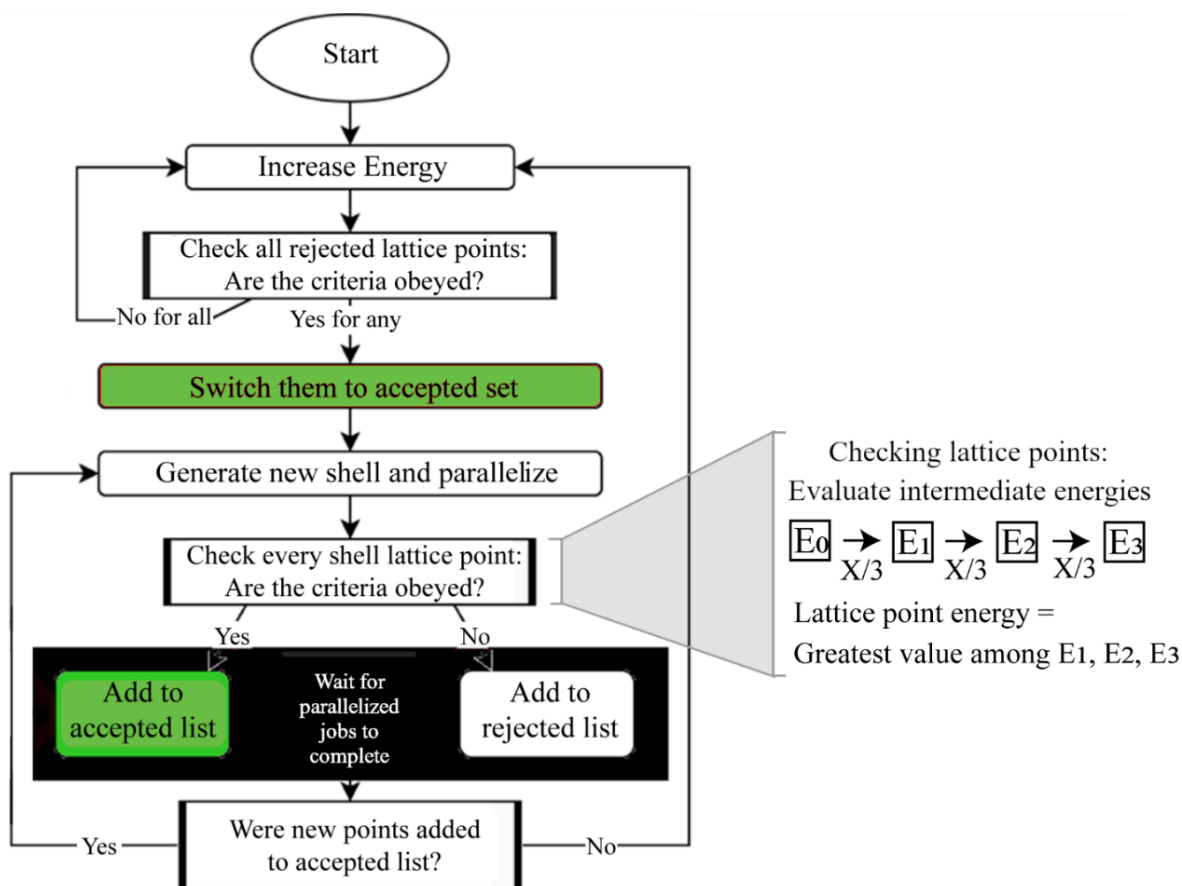
Fig 3. Flowchart of the parallel on-the-fly *Crystal* algorithm. **Left:** Overview of the general *Crystal* algorithm (see ref [13, 21] for further details). **Right:** The methodology for lattice point energy evaluation has been adapted to handle large lattice spacing. In this example, a relaxed scan through three intermediate small steps is performed during a big *Crystal* move step X. At each of the small steps (X/3), the implicit d.o.f are optimized, and the energy is evaluated. *Crystal* records the highest energy from these evaluations as the barrier for transitioning from the previous lattice point to the new lattice point reached by 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.

The adaptations described above were applied to investigate Meldrum's acid 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. As shown in the Results, this new version of *Crystal* algorithm, even with relatively low maximal energy thresholds, enables the discovery of reaction pathways that are overlooked by existing methods[9].

Meldrum's acid in implicit solvent

Our workflow starts with a *Crystal*-based exploration phase, during which the PES is scanned using large lattice spacings in combination 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. *Crystal* is then restarted from geometries with newly discovered bond connectivity using finer lattice spacings for more detailed exploration. During the exploration phase, we employ the semiempirical electronic structure method GFN2-xTB. This method can effectively optimize geometries of molecular systems with hundreds of atoms, providing reasonable accuracy at a significantly faster speed than DFT.[58, 59] In the exploration phase, all explicit d.o.f.s used lattice spacings of 60 degrees, with the implicit d.o.f.s. optimized after every 20-degree interval in the relaxed scan. In ref [13], this choice of 60-degree 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 degrees, 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 **Fig 4**.
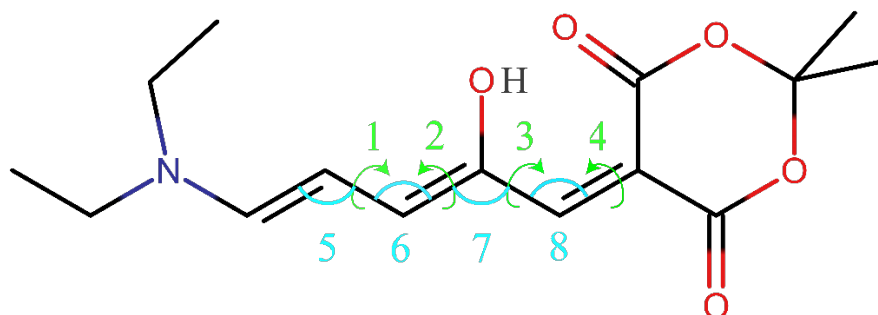
Fig 4. The 8-dimensional configuration space of explicit d.o.f.s explored for Meldrum's acid DASA by *Crystal*. Torsional d.o.f.s are labeled as 1–4 (green), while angular d.o.f.s are labeled as 5–8 (cyan). All remaining implicit d.o.f.s were optimized.

The refinement stage involves optimizing the low-energy lattice points identified by *Crystal* using a more accurate electronic structure method such as ωB97x-D3/def2-TZVP(-f)/COSMO($\varepsilon = 2.38$). The geometries are classified based on the *cis/trans* isomerization of the linear triene chain and any changes in bond connectivity. Structures with new bond connectivity are used as starting points for a finer set of calculations, with angular changes of 45 degrees and the implicit d.o.f.s relaxed every 15 degrees. The hydrogen atom marked in the **Fig. 4** was observed to hop to other atoms in the first round of *Crystal* simulation, despite not being explicitly included among the explored d.o.f.s. This hopping behavior, commonly reported in the literature, is known to occur during the cyclization of the triene chain. During subsequent *Crystal* restarts with finer lattice spacings, 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[60], followed by optimization using the NEB[61] approach with frozen endpoints method. The NEB calculations were performed at ωB97x-D3/def2-TZVP(-f)/COSMO($\varepsilon = 2.38$) level of theory.

All energy and gradient evaluations were conducted using the TeraChem software[62-65], 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[66] integrated into the TeraChem package. The NEB calculations were performed using the DL-FIND optimizer[67] interfaced with TeraChem package.
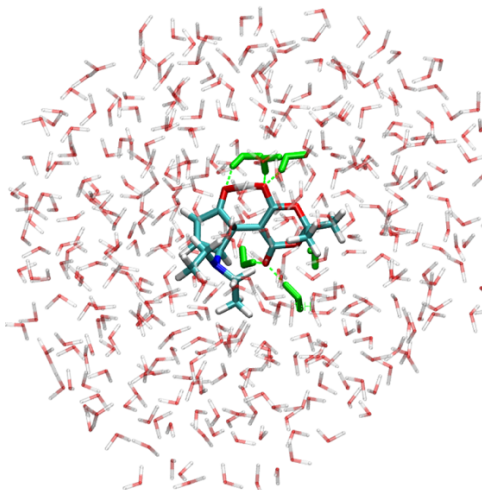
Meldrum's acid DASA in the aqueous solution

The Meldrum's acid DASA molecule 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[68, 69] was used to parameterize the force fields of the DASA molecule. The SPC/Fw water model[70] 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 1 fs timestep. A 10 Å cutoff was applied for nonbonded interactions, and the coordinates of Meldrum's acid 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 (**Fig. 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, with the QM region including the DASA molecule only and treated with BHHLP/6-31G* level of theory, and the MM region including 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.

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 atoms of the DASA molecule 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[71], which was interfaced with the Chemshell[72] software package. All constrained optimizations and relaxed scans were performed with the DL-FIND optimizer[67] 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 endpoint 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.

Fig 5. Meldrum's acid DASA in aqueous solution. 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.

Bilirubin
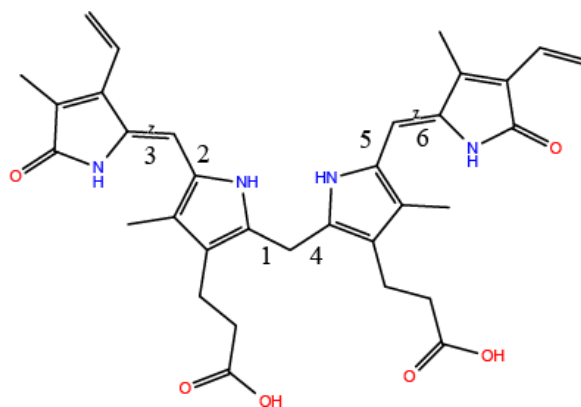
Fig 6. The bilirubin molecule with the six explicitly explored d.o.f.s labeled as indices.

For bilirubin, double-bond isomerizations and their relationship with hydrogen bonds are considered critical due to their biological significance. However, studying double-bond isomerization typically requires multi-reference electronic structure methods due to the near-degenerate frontier molecular orbitals near the transition state[73], which could cause singularity issues for single-referenced methods. For this reason, we used the semi-empirical PM6 method[74]

in combination with the FOMO-CASCI method[75] for *Crystal's* exploration phase as well as for the subsequent geometry refinement. Finer lattice-spacing restarts were not performed.

Energy calculations of bilirubin were performed with the MOPAC software package, interfaced with the standalone GeomeTRIC software package[66] for optimizations. The explicitly d.o.f.s are labeled in **Fig. 6**. A lattice spacing of 60 degrees was used for all torsions, with implicit d.o.f.s minimizations carried out at every 20-degree interval in the relaxed scan. The energy cutoff for *Crystal* was set at 0.06 Hartrees (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 forms do not readily interconvert without light or catalytic intervention. 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 **Fig. 6**—differs by 45 degrees. 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 degrees.

Results

Meldrum's acid in implicit solvent

Fig. 7 illustrates the set of conformational isomers discovered during *Crystal*'s exploration phase. Species **a**, **b, c,** and **d** were all reported by Raucci et al.[9] previously. As observed from our first-round of *Crystal* simulation, the labeled hydrogen atom in **Fig. 7** is prone to hopping between

heavy atoms. To better account for this motion, three additional degrees of freedom 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 (**Fig. 8**), which has not been reported in the literature before.

a)   b)   c)   d)
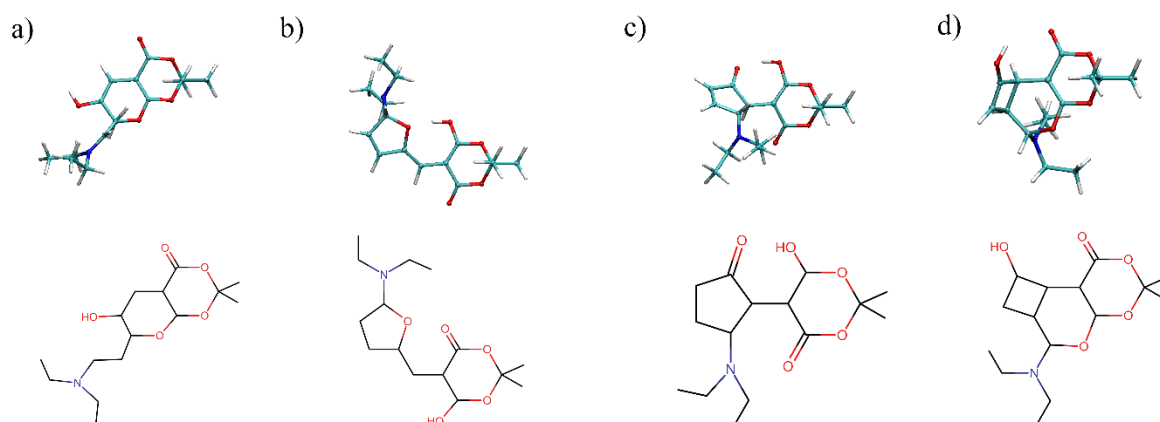
Fig 7. The four conformational isomers identified by *Crystal* during the exploration phase, when starting from Meldrum's acid DASA in its linear form (**Fig 4**). The hydrogen atom prone to hopping between heavy atoms is labeled in the chemical structures at the bottom.

(c)   (c-2)

(e)   (c-3)
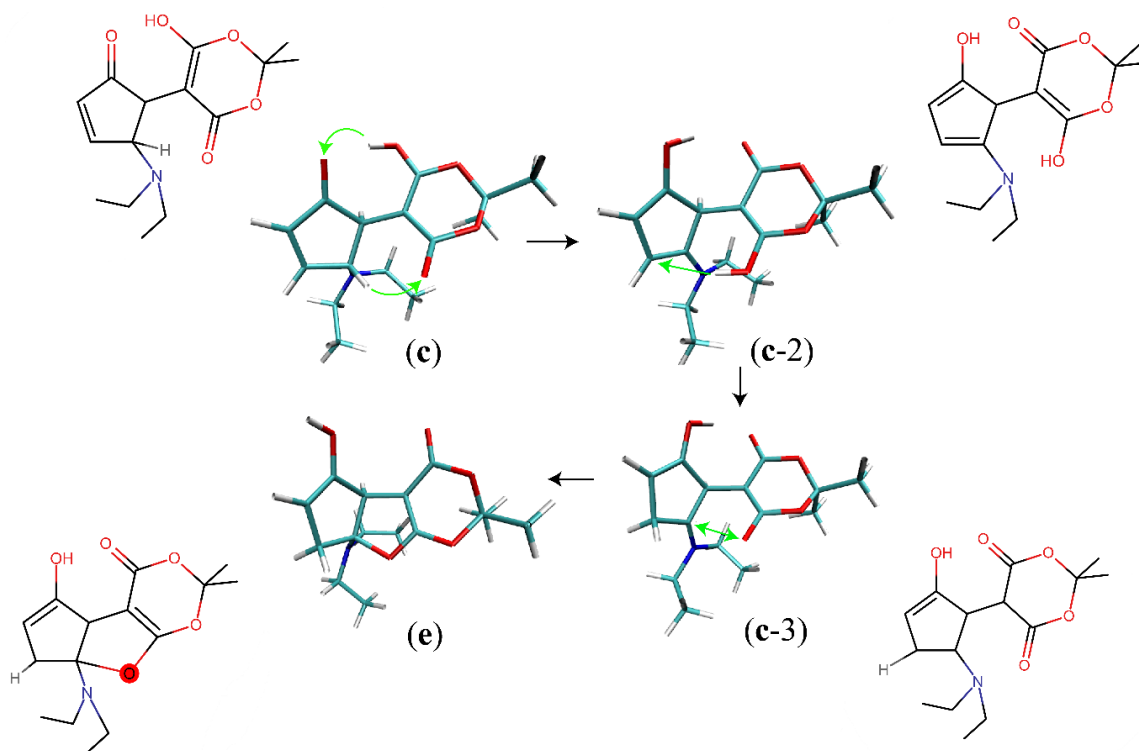
Fig 8. Restarting *Crystal* simulation from species **c** produced species **c-2**, **c-3**, and **e**. The two hydrogen atoms involved in the transfers are explicitly labeled in red and black. Species **c-2** forms via double-hydrogen hopping (green arrows), with additional hydrogen transfers giving rise to species **c-3** and **e**. All transformations exhibit transition barriers below 25 kcal/mol relative to species **c**.
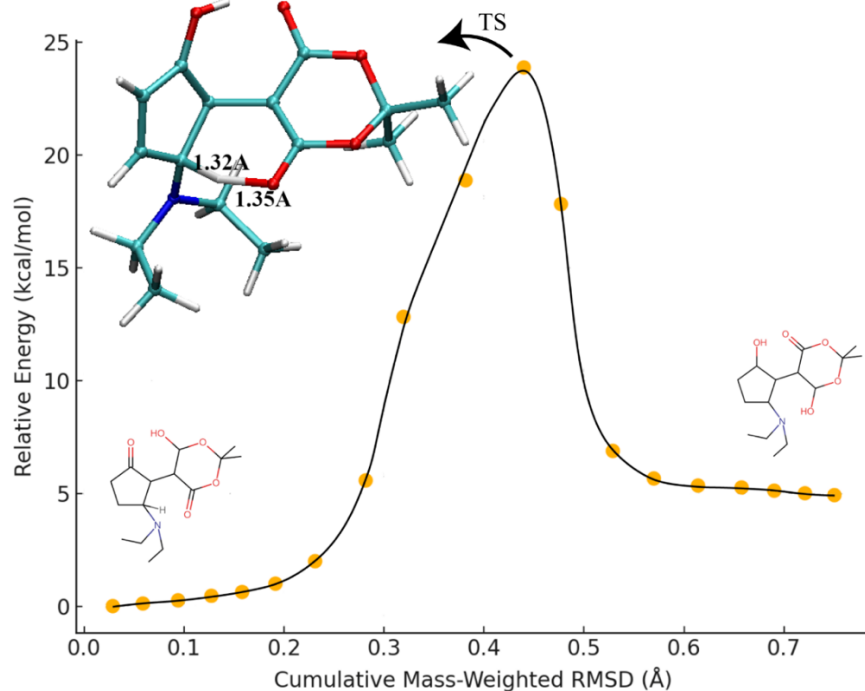
Fig 9. The reaction pathway for the double-hydrogen hopping, showing the transition state (TS) between species **c** (left) and **c-2** (right). At the TS, the first hydrogen transfer (O-to-O near the top) is complete, while the second transfer (C-to-O, marked by the distance values) is midway between the two carbons. Notably, no local minimum exists to support the first hydrogen transfer without the simultaneous second hydrogen transfer.

Table 1. Transition barriers along the double-hydrogen hopping pathway calculated via the NEB approach, utilizing the ωB97x-D3/def2-TZVP(-f)/COSMO($\varepsilon = 2.38$) 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

The transition state (TS) along the MEP is shown in **Fig 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 pathway could be an important consideration when working with DASA molecules that have a similar acceptor ring to the Meldrum's acid.

Restarting the *Crystal* from species **a**, **b**, and **d** with 45-degree lattice spacings reproduced all the geometries reported in ref [9], demonstrating the power of this method to perform comprehensive reaction discovery in large systems with complex reactivity.

Meldrum's acid in the aqueous solution

Our results indicate that the aqueous solution environment can influence the geometries of the DASA molecule corresponding to the same minima. For example, the linear geometries in the gas-phase minima, when solvated in aqueous solution, exhibit significant out-of-plane curvature, as shown in **Fig 10**.

Gas-phase linear minimum   Condensed-phase linear minimum
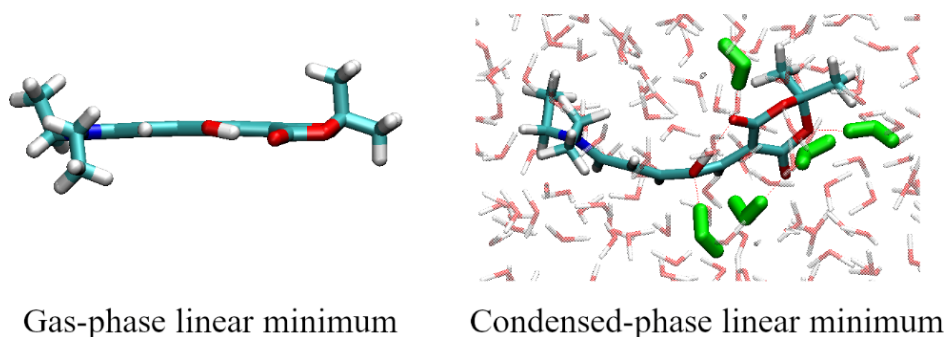
Fig 10. Linear geometries of Meldrum's acid DASA in the gas-phase minima were observed to have notable out-of-plane bending. The green color highlights the water molecules forming hydrogen bonds with DASA.

Unique to the aqueous solution environment, *Crystal* identified a stable cyclic structure where the hydrogen atom remains with the hydroxyl group immediately after ring cyclization (**Fig. 11**). The structure at this minimum, **f-2**, was determined to be only 0.5 kcal/mol deep in the MEP.

Additionally, multiple series of isomers were observed to arise from this pathway, suggesting that this geometry can be an intermediate to other, more stable configurations. Two such configurations with relatively low energy barriers are shown in **Fig. 11**. Notably, a previously unreported zwitterionic species, **f-3**, was discovered, where the nitrogen atom receives hydrogen from a carbon atom rather than from the hydroxyl group.

(f)   (f-2)

(f-3)   (f-4)
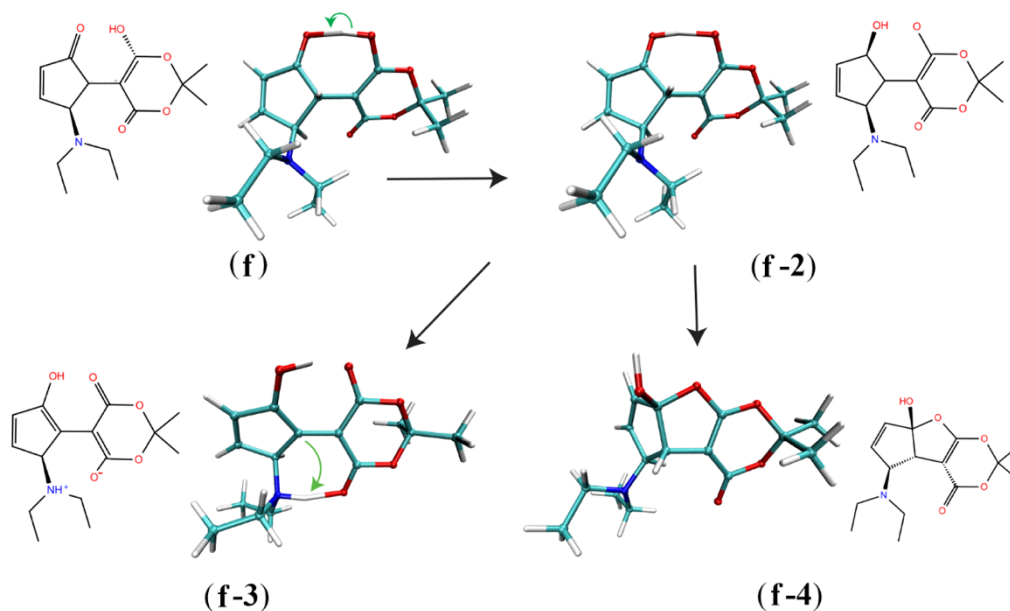
Fig 11. Low-energy species in the aqueous solution **f-3** and **f-4** were discovered to form via the intermediate structure **f-2**, which was unstable in the gas phase. The green arrows indicate the movement of the hydrogen atoms. The transition barriers are provided in **Table 2**.

Table 2. Reaction barriers along the pathways leading from **f in Fig. 11**, calculated via the NEB approach using the ωB97x-D3/6-31G*/SPC/Fw 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

In aqueous environments, cyclic isomers are known to be thermodynamically favored[44, 76]. According to ref 44, 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 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 enables the retention of the hydrogen within the hydroxyl group during cyclization. This stabilization could potentially lower the cyclization barrier.

It has been reported that increasing the temperature promotes the reversal from cyclic to linear isomeric forms. [44] This observation aligns with two factors: first, heating reduces the ability of coordinated water molecules to assist in proton transfer, and second, it disrupts hydrogen bonds that stabilize the **f-2** intermediate. In **Fig. 12**, we illustrate the hydrogen bonds observed for species **f** and **f-2**.

For the **f-2** to **f-3** transition, the MEP is illustrated in **Fig. 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.
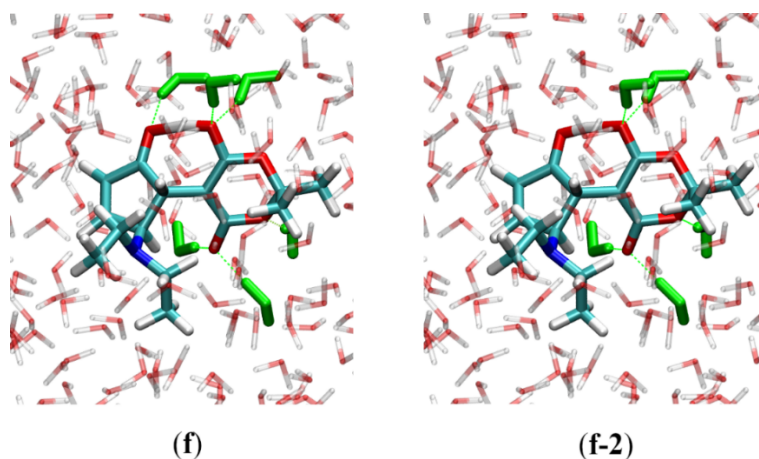
Fig. 12. Hydrogen bonds with water stabilize the carbonyl group, enabling the hydrogen to remain at its hydroxyl position during the cyclization of the DASA's triene chain.
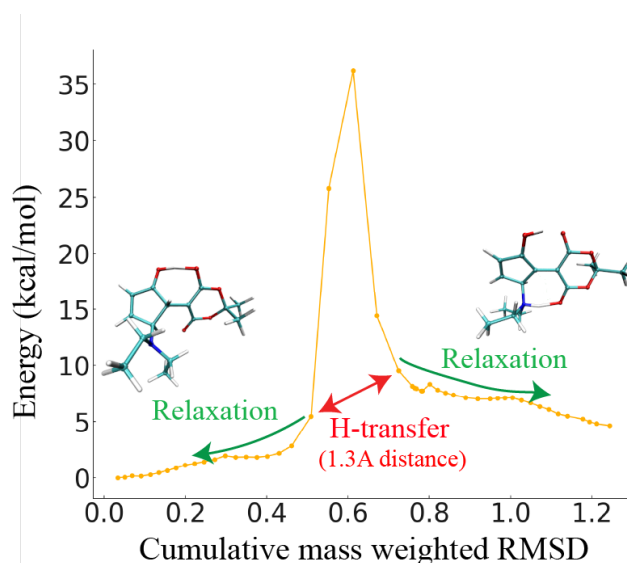
Fig 13. The **f-2** to **f-3** barrier was determined to be 36.5 kcal/mol using the NEB method. This transition barrier is dominated by hydrogen transfer, which may involve quantum tunneling. The hydrogen moves a distance of 1.3 Å between the NEB images indicated by the red arrows, while the rest of the molecular geometry remains largely unchanged between these two images.

Fig 14 contains the geometries and energies identified by *Crystal* in the aqueous solution, and the corresponding gas phase energies. The latter was evaluated by removing the water molecules and

followed by reoptimization of structures and pathways. All calculations were performed using the ωB97x-D3/6-31G* level of theory.
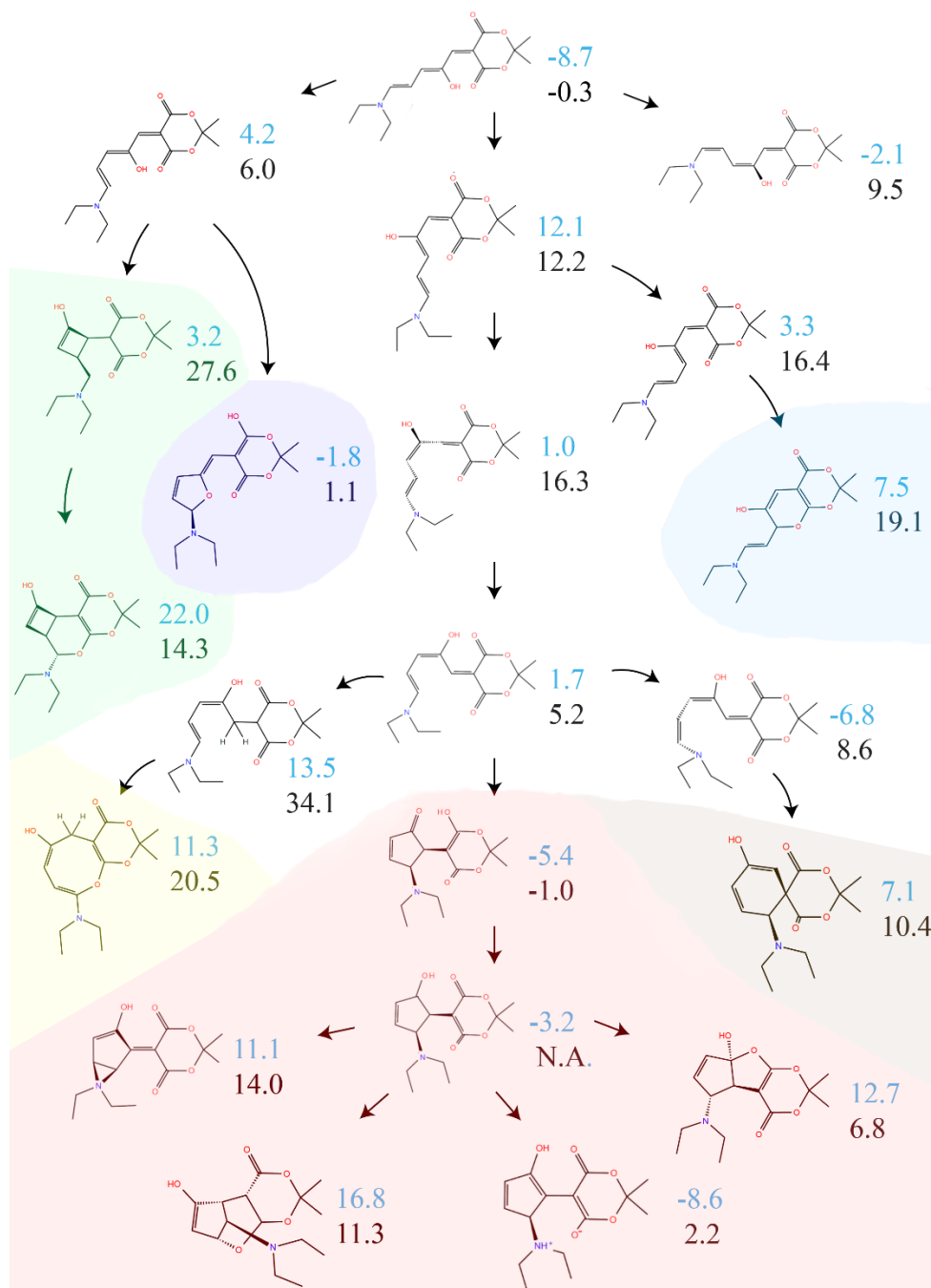

Fig 14. *Crystal*'s discovery of Meldrum's acid DASA configurations. The arrows indicate the sequence in which *Crystal* identified the geometries. Linear geometries were grouped by the isomeric form, while cyclic geometries were grouped by the conformational isomer. The energy

values correspond to the lowest-energy geometry within each group. For each group, the top number represents the condensed-phase energy, and the bottom number represents the gas-phase energy, both relative to the reference structure in **Fig. 10**.

Bilirubin in the gas phase

The *Crystal* simulation of bilirubin exhaustively explores six of its key torsional degrees of freedom simultaneously using a "multireference-like" version of a semiempirical method, i.e., PM6/FOMO-CASCI. The *Crystal* algorithm completely captures both double-bond isomerizations and the associated changes in the H-bond patterns concurrently, generating the Z-Z, Z-E/E-Z, and E-Z isomeric forms. 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. The 10 lowest-energy unique geometries identified for each isomeric form are summarized in **Tables 3-5**. The unique hydrogen bonding patterns observed among these geometries are illustrated in the **Figs. 15-17**.

For example, in ref 77, 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 hydrogen bonds. Additionally, carbons were treated as hydrogen donors in the latter study, which was not the case in ours. A closer comparison of the H-bond distance values reported in ref 77 shows that our lowest-energy structures exhibit significantly more hydrogen bonds within a 2.5 Å cutoff compared to those in ref 77. However, some of our higher-energy isomers closely match their descriptions.

1.1 kcal/mol	1.4 kcal/mol	6.1 kcal/mol
6.5 kcal/mol	6.5 kcal/mol	6.6 kcal/mol
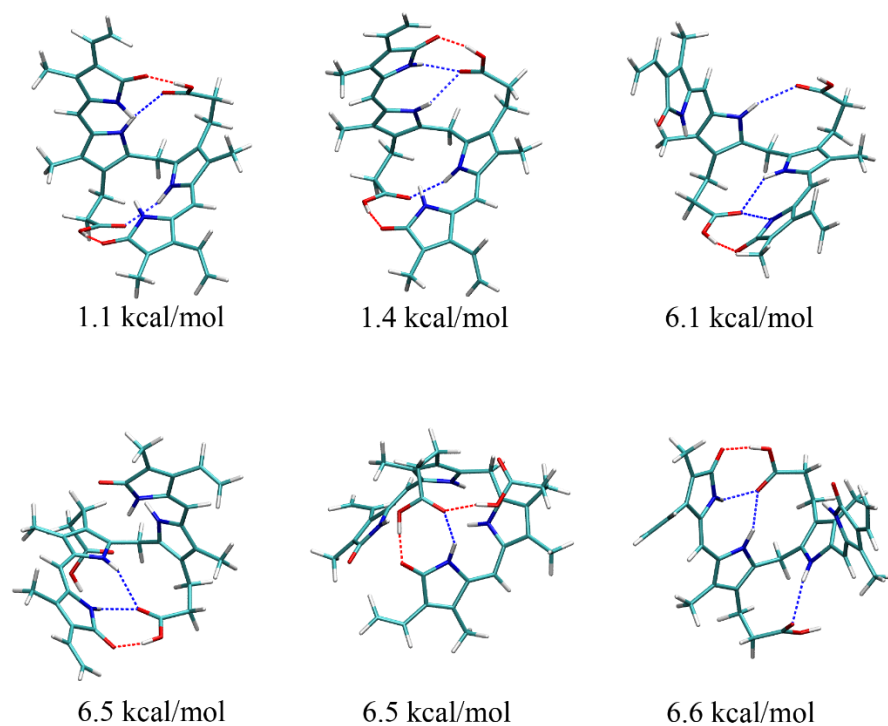
Fig 15. The unique hydrogen bond patterns identified by *Crystal* among the lowest-energy Z-Z isomers of bilirubin.

Table 3. The lowest-energy Z-Z isomers of bilirubin, arranged in ascending order of energy.

Geometry	H bonds	Energy (kcal/mol)	Dihedral 1	Dihedral 2	Dihedral 4	Dihedral 5
1	6	0.0	54.4	54.0	19.8	20.6
2	4	1.1	61.2	-28.0	61.1	-28.0
3	5	1.4	56.2	18.6	57.7	-25.3
4	5	1.5	58.16	-24.8	57.4	19.3
5	5	1.6	55.6	16.3	57.3	-24.0
6	4	6.1	60.7	-128.4	46.71	25.1
7	4	6.4	46.7	25.0	60.8	-129.1

8	3	6.5	-39.8	-26.2	-55.6	37.5
9	3	6.5	15.7	-44.1	-75.5	121.0
10	2	6.6	-51.9	90.7	-21.6	45.09

| 5.7 kcal/mol | 6.5 kcal/mol | 7.3 kcal/mol |

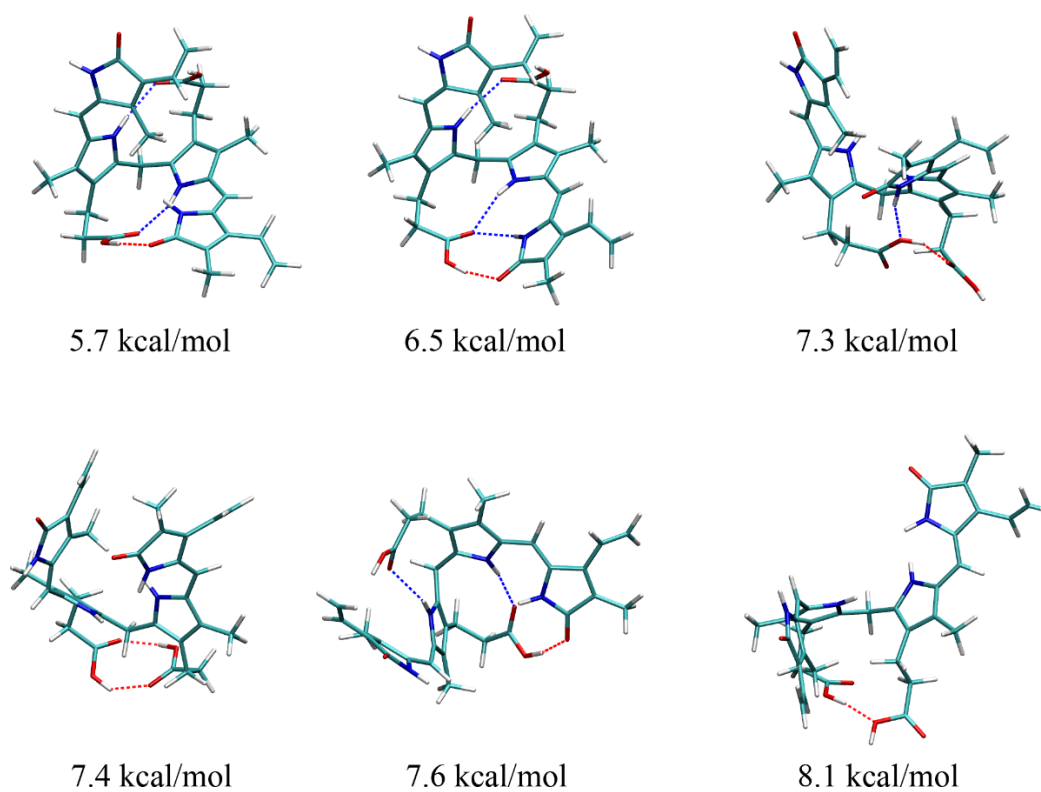| 7.4 kcal/mol | 7.6 kcal/mol | 8.1 kcal/mol |

Fig 16. The unique hydrogen bond patterns identified by *Crystal* among the lowest-energy E-Z/Z-E isomers of bilirubin.

Table 4. The lowest-energy E-Z / Z-E isomers of bilirubin, arranged in ascending order of energy.

Geometry	H bonds	Energy (kcal/mol)	Dihedral 1	Dihedral 2	Dihedral 4	Dihedral 5
1	3	5.7	61.6	-48.8	64.9	-32.4
2	4	6.5	59.5	-41.2	57.1	19.5

3	4	7.2	59.2	61.9	44.7	24.7
4	2	7.3	80.4	-92.0	-2.8	40.0
5	2	7.6	-41.8	92.4	-24.5	45.1
6	3	8.1	59.3	44.3	56.1	-25.6
7	1	8.9	15.9	-94.1	-107.2	46.9
8	2	9.3	62.3	-93.0	30.6	-128.1
9	2	9.9	57.3	-47.4	67.2	-124.4
10	1	10.0	19.4	54.0	-98.1	35.6

10.9 kcal/mol   12.8 kcal/mol   13.2 kcal/mol   14.7 kcal/mol
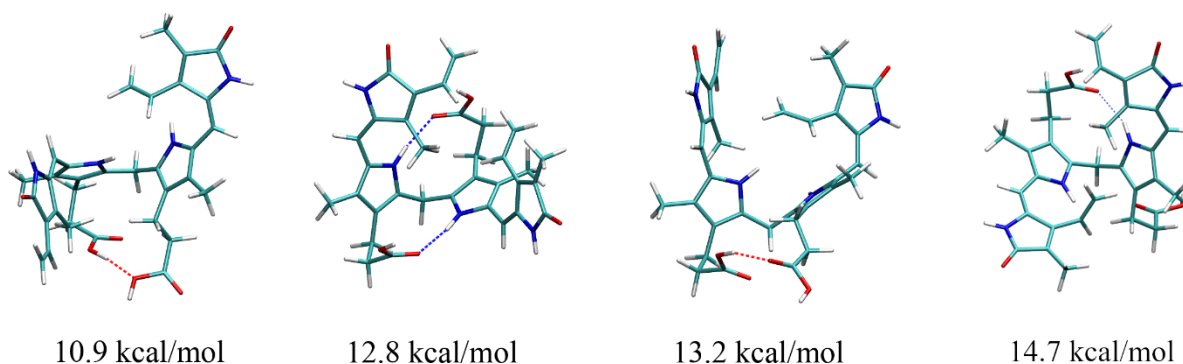
Fig 17. The unique hydrogen bond patterns identified by Crystal among the lowest-energy E-E isomers of bilirubin.

Table 5. The lowest-energy E-E isomers of bilirubin, arranged in ascending order of energy

Geometry	H bonds	Energy (kcal/mol)	Dihedral 1	Dihedral 2	Dihedral 4	Dihedral 5
1	1	10.9	15.5	-94.1	-107.2	46.9
2	2	12.8	65.4	-44.9	64.7	-129.2

3	2	12.8	61.5	-42.9	69.2	-38.2
4	1	13.3	21.9	62.8	-106.4	101.8
5	2	13.6	52.6	-43.0	69.9	42.3
6	2	13.6	59	-62.4	53.19	-121.0
7	2	13.8	59.5	-45.3	66.4	120.3
8	1	14.0	16.3	-94.2	-107.7	46.7
9	0	14.5	16.26	-94.2	-107.7	46.7
10	1	14.7	-78.8	51.9	-60.8	18.9

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 it 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 size limit for full-dimensional exploration. However, the new version of *Crystal* algorithm can nonetheless successfully reproduce data from prior studies and uncover additional new reaction pathways. This was mainly achieved via the relaxation of d.o.f.s outside *Crystal*'s lattice space in response to the d.o.f.s. explicitly sampled by *Crystal*. The multiple technical advancements greatly extend *Crystal*'s applicability to reaction discovery in large, complex and realistic molecular systems and enable it to compete with or even surpass conventional methods in revealing complex features of the PES.

While the selection of explicit d.o.f.s remain critical for the successful discovery of reactions, *Crystal* consistently exhibits the ability to identify unexpected reaction paths beyond the initially defined explicit d.o.f.s. 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 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 **Fig. 7**) without explicitly 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 global 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.

AUTHOR INFORMATION

Corresponding Author

Ruibin Liang

ORCID: 0000-0001-8741-1520

Bill Poirier

ORCID: 0000-0001-8277-746X

Notes

The authors declare no competing financial interests.

ACKNOWLEDGMENT

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. DOI: 10.1038/nchem.2099.

(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. DOI: 10.1021/jp507698m.

(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. DOI: 10.1021/ja512394y.

(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, 10.1039/C8CP04703K. DOI: 10.1039/C8CP04703K.

(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. DOI: 10.1021/acs.jctc.9b00029.

(6) Meisner, J.; Zhu, X.; Martínez, T. J. Computational Discovery of the Origins of Life. *ACS Central Science* **2019**, *5* (9), 1493-1495. DOI: 10.1021/acscentsci.9b00832.

(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, 10.1039/D1SC00775K. DOI: 10.1039/D1SC00775K.

(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. DOI: 10.1021/jacs.1c06648.

(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. DOI: 10.1021/jacs.2c04419.

(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. DOI: 10.1021/acs.jpclett.1c03993.

(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, 10.1039/D3SC01202F. DOI: 10.1039/D3SC01202F.

(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. DOI: 10.1021/acs.jpca.3c05484.

(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. DOI: 10.1021/acs.jctc.4c00292.

(14) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100* (2), 020603. DOI: 10.1103/PhysRevLett.100.020603.

(15) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (20), 12562-12566. DOI: 10.1073/pnas.202427399 (acccessed 2025/01/09).

(16) Invernizzi, M.; Parrinello, M. Rethinking Metadynamics: From Bias Potentials to Probability Distributions. *J. Phys. Chem. Lett.* **2020**, *11* (7), 2731-2736. DOI: 10.1021/acs.jpclett.0c00497.

(17) Zimmerman, P. M. Growing string method with interpolation and optimization in internal coordinates: method and examples. *J Chem Phys* **2013**, *138* (18), 184102. DOI: 10.1063/1.4804162 From NLM PubMed-not-MEDLINE.

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

(19) Jafari, M.; Zimmerman, P. M. Uncovering reaction sequences on surfaces through graphical methods. *Phys. Chem. Chem. Phys.* **2018**, *20* (11), 7721-7729, 10.1039/C8CP00044A. DOI: 10.1039/C8CP00044A.

(20) 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. DOI: 10.1246/bcsj.20140189 (acccessed 1/9/2025).

(21) 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. DOI: https://doi.org/10.1002/jcc.27324 (acccessed 2024/05/04).

(22) Pandey, A.; Poirier, B. Using phase-space Gaussians to compute the vibrational states of OCHCO+. *J. Chem. Phys.* **2019**, *151* (1), 014114. DOI: 10.1063/1.5096770 (acccessed 5/5/2024).

(23) Pandey, A.; Poirier, B. An algorithm to find (and plug) "holes" in multi-dimensional surfaces. *J. Chem. Phys.* **2020**, *152* (21), 214102. DOI: 10.1063/5.0005681 (acccessed 2022/03/09).

(24) 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. DOI: 10.1021/acs.jpclett.0c01435.

(25) 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. DOI: 10.1021/jp810898u.

(26) 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. DOI: 10.1021/jz200262m.

(27) 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, 10.1039/C3CP44063J. DOI: 10.1039/C3CP44063J.

(28) 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. DOI: 10.1021/ct400512u.

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

(30) 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, 10.1039/D0SC04187D. DOI: 10.1039/D0SC04187D.

(31) Pianowski, Z. L. Recent Implementations of Molecular Photoswitches into Smart Materials and Biological Systems. *Chem. Eur. J* **2019**, *25* (20), 5128-5144. DOI: 10.1002/chem.201805814 (acccessed 2019/08/24).

(32) Boase, N. R. B. Shining a Light on Bioorthogonal Photochemistry for Polymer Science. *Macromolecular Rapid Communications* **2020**, *41* (18), 2000305. DOI: https://doi.org/10.1002/marc.202000305 (acccessed 2024/05/04).

(33) Boelke, J.; Hecht, S. Designing Molecular Photoswitches for Soft Materials Applications. *Advanced Optical Materials* **2019**, *7* (16), 1900404. DOI: https://doi.org/10.1002/adom.201900404 (acccessed 2024/05/04).

(34) Goulet-Hanssens, A.; Eisenreich, F.; Hecht, S. Enlightening Materials with Photoswitches. *Advanced Materials* **2020**, *32* (20), 1905966. DOI: https://doi.org/10.1002/adma.201905966 (acccessed 2024/05/04).

(35) 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. *The Journal of Organic Chemistry* **2014**, *79* (23), 11316-11329. DOI: 10.1021/jo502206g.

(36) 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. DOI: 10.1021/ja503016b.

(37) 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, 10.1039/D3CS00508A. DOI: 10.1039/D3CS00508A.

(38) 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. DOI: 10.1038/s41557-022-00947-8.

(39) Bersuker, I. *The Jahn-Teller Effect*; Cambridge University Press, 2006. DOI: DOI: 10.1017/CBO9780511524769.

(40) 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, 10.1039/C8SC03218A. DOI: 10.1039/C8SC03218A.

(41) 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), e202300222. DOI: https://doi.org/10.1002/chem.202300222 (acccessed 2024/05/04).

(42) 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, 10.1039/D0SC06575G. DOI: 10.1039/D0SC06575G.

(43) 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, 10.1039/D2CC03761K. DOI: 10.1039/D2CC03761K.

(44) 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. *Communications Chemistry* **2019**, *2* (1), 118. DOI: 10.1038/s42004-019-0221-5.

(45) Busch, A. W. U.; Montgomery, B. L. Interdependence of tetrapyrrole metabolism, the generation of oxidative stress and the mitigative oxidative stress response. *Redox Biology* **2015**, *4*, 260-271. DOI: https://doi.org/10.1016/j.redox.2015.01.010.

(46) Kräutler, B. Phyllobilins – the abundant bilin-type tetrapyrrolic catabolites of the green plant pigment chlorophyll. *Chem. Soc. Rev.* **2014**, *43* (17), 6227-6238, 10.1039/C4CS00079J. DOI: 10.1039/C4CS00079J.

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

(48) Mancuso, C. Bilirubin and brain: A pharmacological approach. *Neuropharmacology* **2017**, *118*, 113-123. DOI: https://doi.org/10.1016/j.neuropharm.2017.03.013.

(49) 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. *In Vivo* **2005**, *19* (5), 855.

(50) Upadhyaya, H. P. Computational Characterization of "Dark" Intermediates in the Ultrafast Deactivation of Photoexcited Bilirubin. *J. Phys. Chem. A* **2018**, *122* (46), 9084-9092. DOI: 10.1021/acs.jpca.8b09392.

(51) Agati, G.; Fusi, F. New trends in photobiology recent advances in bilirubin photophysics. *Journal of Photochemistry and Photobiology B: Biology* **1990**, *7* (1), 1-14. DOI: https://doi.org/10.1016/1011-1344(90)85138-M.

(52) 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 Network Open* **2019**, *2* (3), e190858-e190858. DOI: 10.1001/jamanetworkopen.2019.0858 (acccessed 1/13/2025).

(53) Cremer, R. J.; Perryman, P. W.; Richards, D. H. INFLUENCE OF LIGHT ON THE HYPERBILIRUBINÆMIA OF INFANTS. *The Lancet* **1958**, *271* (7030), 1094-1097. DOI: https://doi.org/10.1016/S0140-6736(58)91849-X.

(54) 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, 10.1039/C5CP06971H. DOI: 10.1039/C5CP06971H.

(55) 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. DOI: 10.1073/pnas.76.1.29 (acccessed 2025/01/13).

(56) McDonagh, A. F.; Palma, L. A.; Lightner, D. A. Blue Light and Bilirubin Excretion. *Science* **1980**, *208* (4440), 145-151. DOI: 10.1126/science.7361112 (acccessed 2025/01/13).

(57) Lightner, D. A.; McDonagh, A. F. Molecular mechanisms of phototherapy for neonatal jaundice. *Acc. Chem. Res.* **1984**, *17* (12), 417-424. DOI: 10.1021/ar00108a002.

(58) Grimme, S.; Bannwarth, C.; Shushkov, P. A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parametrized for All spd-Block Elements (Z = 1–86). *J. Chem. Theory Comput.* **2017**, *13* (5), 1989-2009. DOI: 10.1021/acs.jctc.7b00118.

(59) 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. DOI: 10.1021/acs.jctc.8b01176.

(60) Zhu, X.; Thompson, K. C.; Martínez, T. J. Geodesic interpolation for reaction pathways. *J. Chem. Phys.* **2019**, *150* (16), 164103. DOI: 10.1063/1.5090303 (acccessed 12/7/2024).

(61) 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. DOI: 10.1063/1.1329672 (acccessed 2020/10/15).

(62) 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. DOI: 10.1021/ct300321a.

(63) 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. DOI: 10.1021/ct9003004.

(64) 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. DOI: 10.1063/1.3382344 (acccessed 12/7/2024).

(65) 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. DOI: https://doi.org/10.1002/jcc.21759 (acccessed 2024/12/07).

(66) Wang, L.-P.; Song, C. Geometry optimization made simple with translation and rotation coordinates. *J. Chem. Phys.* **2016**, *144* (21), 214108. DOI: 10.1063/1.4952956 (acccessed 4/23/2023).

(67) Kästner, J.; Carr, J. M.; Keal, T. W.; Thiel, W.; Wander, A.; Sherwood, P. DL-FIND: An Open-Source Geometry Optimizer for Atomistic Simulations. *J. Phys. Chem. A* **2009**, *113* (43), 11856-11865. DOI: 10.1021/jp9028968.

(68) 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. DOI: 10.1002/jcc.20035.

(69) 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. DOI: https://doi.org/10.1016/j.jmgm.2005.12.005.

(70) 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). DOI: Artn 024503
Doi 10.1063/1.2136877.

(71) Neese, F.; Wennmohs, F.; Becker, U.; Riplinger, C. The ORCA quantum chemistry program package. *J. Chem. Phys.* **2020**, *152* (22), 224108. DOI: 10.1063/5.0004608 (acccessed 4/23/2023).

(72) Metz, S.; Kästner, J.; Sokol, A. A.; Keal, T. W.; Sherwood, P. ChemShell—a modular software package for QM/MM simulations. *WIREs Computational Molecular Science* **2014**, *4* (2), 101-110, https://doi.org/10.1002/wcms.1163. DOI: https://doi.org/10.1002/wcms.1163 (acccessed 2023/03/07).

(73) 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. DOI: 10.1021/acs.jctc.3c00451.

(74) Stewart, J. J. P. Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *Journal of Molecular Modeling* **2007**, *13* (12), 1173-1213. DOI: 10.1007/s00894-007-0233-4.

(75) 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. DOI: 10.1063/1.3436501 (acccessed 2022/12/30).

(76) Klajn, R. Spiropyran-based dynamic materials. *Chem. Soc. Rev.* **2014**, *43* (1), 148-184, 10.1039/C3CS60181A. DOI: 10.1039/C3CS60181A.

(77) Cruz, Á. 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. *Journal of Molecular Modeling* **2023**, *29* (10), 318. DOI: 10.1007/s00894-023-05720-3.