White Rose Research Online
eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Leeds.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/238213/

Version: Accepted Version

Article:

Goodfellow, A.S. and Nguyen, B.N. (Accepted: 2026) Graph-Based Internal Coordinate Analysis for Transition State Characterisation. Journal of Chemical Theory and Computation. ISSN: 1549-9618 (In Press)

https://doi.org/10.1021/acs.jctc.5c02073

This is an author produced version of an article accepted for publication in Journal of Chemical Theory and Computation, made available via the University of Leeds Research Outputs Policy under the terms of the Creative Commons Attribution License (CC-BY), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

Reuse
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

Graph-Based Internal Coordinate Analysis for Transition State Characterisation

Alister S. Goodfellow[a] * and Bao N. Nguyen[a]

[a]School of Chemistry, University of Leeds, Woodhouse Lane, Leeds, LS2 9JT, UK.

Abstract

We present `graphRC`, a method for rapid transition state (TS) mode analysis using internal coordinates derived from molecular graphs. The imaginary mode of a TS describes the direction of atomic motion at the saddle point, providing a local approximation to the reaction coordinate, while Intrinsic Reaction Coordinate (IRC) and Quick Reaction Coordinate (QRC) calculations trace the full pathway to adjacent minima. In all cases, displacements are expressed in Cartesian coordinates and do not directly describe changes in bonding. By translating these into internal coordinate changes (bonds, angles, dihedrals), `graphRC` provides chemical insight into the TS mode and reaction coordinate trajectories without prior knowledge of reactant and product structures. Molecular connectivity is determined using `xyzgraph`, a flexible graph builder validated across 4846 structures spanning 61 elements and 490 element-pair bond types with close agreement to DFT-derived bonding. Initial validation on 16 diverse TS achieved 100% identification of bond changes, rotations and inversions with zero false positives compared to IRC and QRC connectivity. Across 395 TS covering organic, organometallic and catalytic transformations, normal mode analysis alone detects the primary bond change in every case, with high agreement to IRC-derived connectivity. This enables programmatic TS verification at a fraction of the cost of formal reaction coordinate calculations, complementing more rigorous methods with rapid, interpretable analysis.

1 Introduction

Transition state (TS) validation is central to mechanistic computational chemistry but remains a major bottleneck for high-throughput computational workflows. Formal verification is performed using Intrinsic Reaction Coordinate (IRC) calculations which trace the minimum-energy pathway downhill from the transition state to adjacent minima (**Figure 1A**).[1–3] While IRC calculations are the gold standard, these are computationally expensive, typically requiring re-evaluation of the Hessian and very small displacement steps to accurately map out the Potential Energy Surface (PES). These calculations can be challenging to perform and often fail to converge on flat regions of the PES or with low-magnitude imaginary modes, making them impractical for large-scale studies.

Quick Reaction Coordinates (QRC), introduced by Goodman and Silva,[4] offer a quick alternative by geometrically displacing the TS along each direction of the imaginary mode, generating *pre-* and *post*-TS structures which are then optimised to adjacent local minima (**Figure 1A**). This approach is widely used to verify connectivity to adjacent

---

*Corresponding author: a.s.goodfellow@leeds.ac.uk

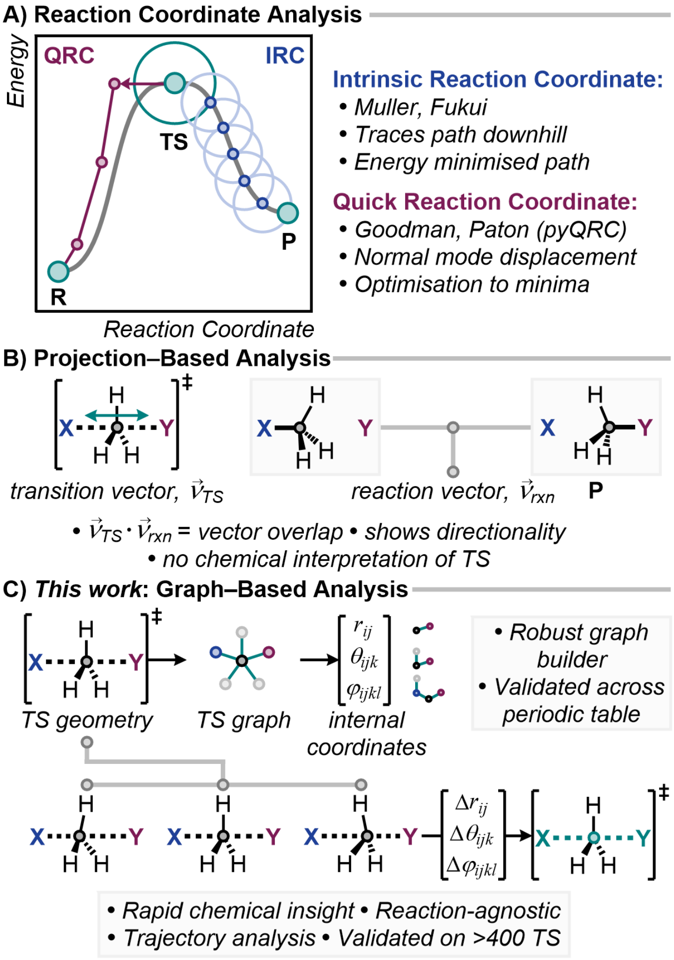


Figure 1: A) Reaction coordinate IRC and QRC transition state validation. **B)** Normal mode projection overlap. **C)** *This work*: TS-centric graph based internal coordinate analysis.

minima across organic and organometallic transformations assisted by the availability of normal mode displacement in various software packages, including `Avogadro`,[5] `Molden`,[6] `Gaussview`,[7] `ORCA`,[8] and tools such as `pyQRC`[9] and `CCLIB`.[10] QRC calculations avoid the expense of IRC calculations by requiring only a single standard geometry optimisation in each direction rather than constrained path-following. While this makes QRC simple and reliable across TSs, the cost still scales poorly for validating thousands of TSs, limiting its practicality for high-throughput workflows. The validation bottleneck is increasingly relevant as automated tools such as `autoDE`,[11] `AARON`,[12] and `TS-tools`[13] streamline TS searches, and multi-level workflows that locate TS at a lower level of theory (*e.g.* semi-empirical GFN2-xTB[14]) before refinement at DFT generate large numbers of candidate structures requiring validation. Double-ended methods such as QST2,[15] NEB,[16] or FSM[17] help direct the TS search but still require formal verification that the TS connects the intended reactant and product. Advances in machine learning have driven the production of large-scale datasets, but these often avoid exploring transition states because validation remains computationally complex (*e.g.* OMol25[18]). Some workflows instead adopt a low-cost approach to TS validation by projecting the imaginary mode onto a reaction vector that connects the reactant and product structures.[19,20] This checks whether the TS mode is broadly aligned with the expected transformation, but does not identify which bonds form or break, and may mask multistep rearrangements that partially overlap with the expected reaction vector (**Figure 1B**).

Quantum chemistry calculations naturally produce molecular structures as Cartesian coordinates, but chemically meaningful interpretation requires molecular graphs and internal coordinates. Molecular graphs and adjacency matrices are widely employed in cheminformatics and machine learning workflows, often using `RDKit`. Existing graph builders (`xyz2mol`[21] and single-metal extension `xyz2mol_tm`[22]) are integrated into `RDKit`,[23] using distance-based connectivity and Hückel calculations to reliably assign bond orders for equilibrium structures.[24] These methods are not designed for transition states and can fail to produce molecular graphs. Partially formed or broken bonds fall outside normal bonding ranges, formal charge assignments become ambiguous, standard valence rules are violated at atoms involved in bond changes, and multi-metal coordination patterns are left unresolved. Automated TS analysis requires a robust graph construction method that can handle non-equilibrium geometries to translate vibrational displacements into changes in chemically meaningful internal coordinates.

Here we present `graphRC`, a lightweight Python package for automated transition state analysis using internal coordinate tracking across both normal mode displacements and reaction trajectories (**Figure 1C**). Unlike projection-based methods that only consider correlation with a specified reaction pathway, `graphRC` combines robust graph construction with vibrational analysis to characterise the behaviour of the TS imaginary mode. The package accepts standard computational output files (including `ORCA` and `Gaussian`) containing TS geometries and Hessians for normal mode projection, or reaction coordinate trajectories from IRC or QRC calculations. Molecular graphs are constructed using `xyzgraph`,[25] which integrates flexible geometric validation and valence-charge optimisation, producing chemically accurate connectivity for both ground state and transition state geometries. Building on this framework, `graphRC` maps normal mode displacement and reaction coordinate trajectories into internal coordinates, enabling the rapid identification of bond, angle, and dihedral changes along the reaction coordinate without prior knowledge of the reactant or product structures. This reaction-agnostic approach delivers a chemically valid interpretation of the TS mode at a computational cost orders of magnitude lower than both IRC and QRC calculations. Structured outputs enable programmatic verification and complement reaction coordinate approaches, offering a straightforward integration into high-throughput automated workflows. The package is openly available on GitHub (`https://github.com/aligfellow/graphRC.git`), `pip`-installable and free from heavy dependencies. In the following sections we outline the methodology and validate the performance across diverse examples.

2 Methodology

Transition state analysis is performed in internal coordinates, using graph-based connectivity from `xyzgraph` (`https://github.com/aligfellow/xyzgraph.git`).[25] `graphRC` offers CLI and Python API options for analysing normal mode displacement and IRC or QRC trajectories. The workflow is outlined below.

2.1 Molecular Graph Construction

Molecular graph construction is performed using `xyzgraph`,[25] which employs a two-pass distance-based approach (**Figure 2**). The first pass establishes baseline connectivity using van der Waals radii[26] with thresholds varying by bond type (0.38–0.7 × the sum of vdW radii), applying tighter criteria for covalent bonds and looser ones for metal-ligand interactions. Bonds with interatomic distances well below the threshold (distance <0.6 × threshold) are accepted directly, while longer bonds undergo geometric validation to reject chemically invalid connectivity. Validation includes acute angle criteria (<15° for metals, <30° for non-metals) and ring diagonal checks using

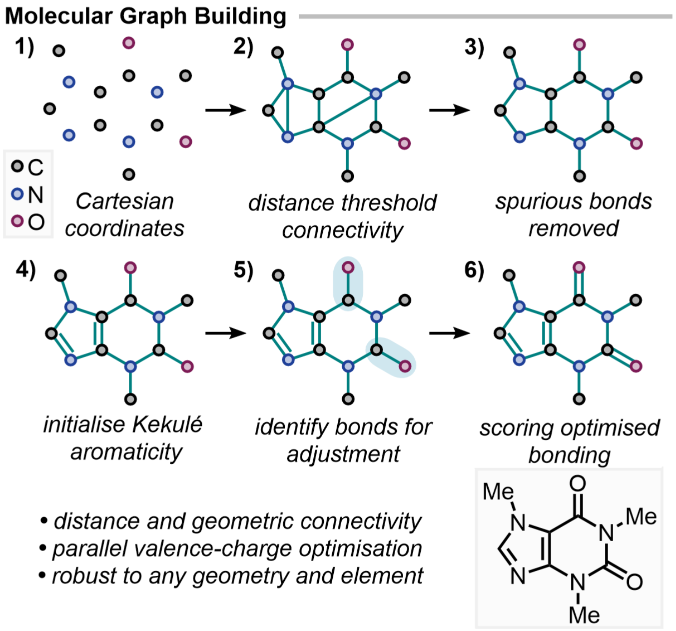


Figure 2: Molecular graph construction workflow using `xyzgraph`. Geometric validation filters spurious connectivity, followed by parallel valence-charge optimisation. The method applies across all elements and produces graph connectivity for any input geometry, allowing for conversion to a set of non-redundant internal coordinates.

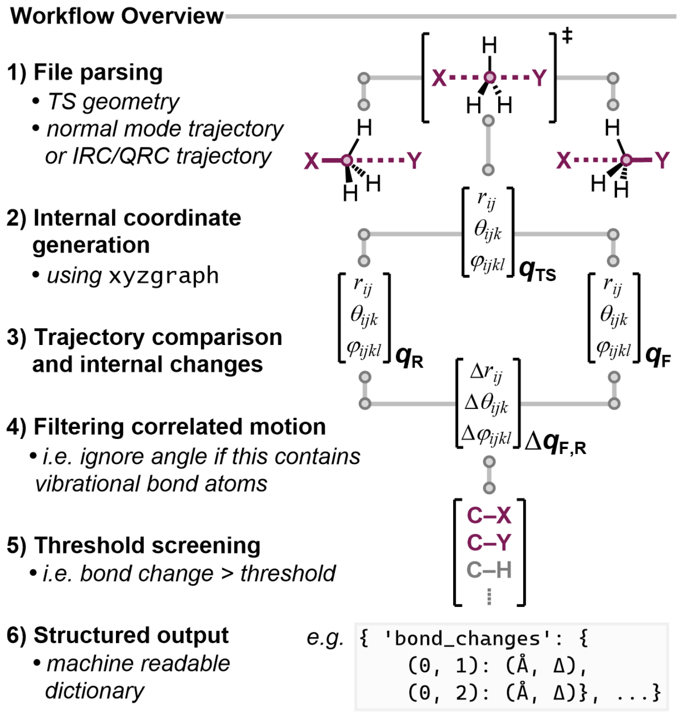
distance ratios to prevent the formation of artefactual three-membered and cross-linked rings. For transition state geometries, a global threshold scaling parameter uniformly extends all distance criteria in a second pass (1.4 × default bonding thresholds), enabling detection of elongated bonds while geometric validation maintains chemically reasonable connectivity.

Bond orders are optimised iteratively, first beginning with Kekulé initialisation for aromatic rings (alternating single and double bonds) to reduce the number of iterations in the optimisation process. At each step, bonds with the worst valence description are selected for bond order adjustment (single ↔ double ↔ triple). Changes that minimise formal charges and satisfy valence requirements are prioritised. Multiple candidate bonding patterns are explored simultaneously to avoid convergence to local minima. Metal coordination is handled by fixing metal-ligand bond orders to one throughout the optimisation. The bonding pattern of the ligand sphere is then optimised to assign dative or covalent character to the metal-ligand bonds. Finally, Kekulé aromaticity is converted to aromatic bond orders (1.5) for a uniform representation across aromatic rings.

2.2 Vibrational Analysis

1. File parsing: The workflow begins by either performing normal mode displacement on a transition state geometry or by reading an IRC (or QRC) trajectory. Computational output files (from packages such as `ORCA`[8] and `Gaussian`[7]) are parsed using `CCLIB`[10] to extract vibrational data and generate a vibrational trajectory.

2. Internal coordinate generation: Molecular graphs are constructed from the transition state Cartesian coordinates using `xyzgraph`. From the graph topology, a set of internal coordinates (bonds, angles, dihedrals) is generated. Optionally, graphs from reactant and product structures can be used to augment the internal coordinate set, ensuring that key bond coordinates are not missed from the extended TS graph.



```
{ 'bond_changes': {
    (0, 1): (Å, Δ),
    (0, 2): (Å, Δ)}, ...}
```

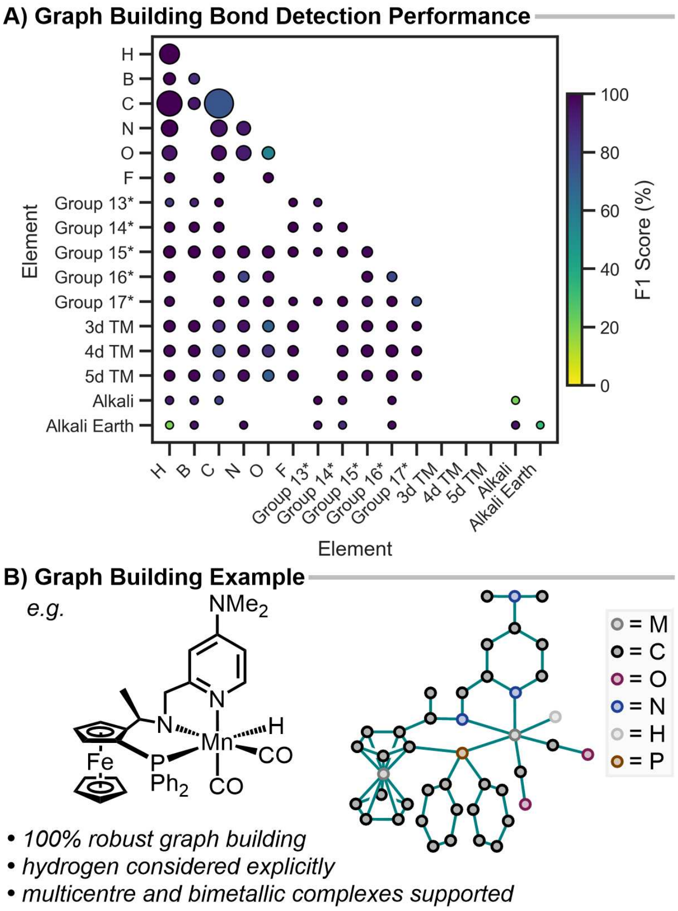
Figure 3: Overview of the workflow employed in this work. q is the internal coordinate vector evaluated at the TS and in the forward (F) and reverse (R) directions.

3. Trajectory comparison and internal changes: Two frames are selected from the trajectory, either the maximally geometrically diverse (largest pairwise RMSD) or the first and last frames. Internal coordinates are evaluated at each frame and the changes in bond lengths, angles and dihedrals are compared.

4. Filtering correlated motion: Internal coordinate changes are organised hierarchically to separate primary and coupled motion. Angles and dihedrals involving atoms with changing bonds are filtered out, as these variations are a geometric consequence of the bond change. Equivalent dihedrals that share a rotational axis are filtered to ensure independent rotational characterisation, retaining the dihedral defined by the heaviest atoms.

5. Threshold screening: Default thresholds are applied to detect meaningful changes: 0.4 Å for bonds, 10° for angles, 20° for dihedrals (initially assigned empirically and validated systematically, *vide infra*). If no changes are detected, thresholds are reduced by 50% to accommodate low-magnitude modes such as hindered rotations or heavy-atom motion. Correlated proton transfers are captured with a secondary reduced threshold, ensuring that each hydrogen-containing bond-breaking event has a corresponding bond-making event where appropriate.

6. Structured output: Results are returned as structured dictionaries containing internal coordinate changes, mode characterisation, graph objects and file paths. This structured output enables programmatic TS validation in high-throughput workflows. Displaced structures can also be generated for subsequent QRC analysis (*c.f.* `pyQRC`).



A) Graph Building Bond Detection Performance

B) Graph Building Example

- *100% robust graph building*
- *hydrogen considered explicitly*
- *multicentre and bimetallic complexes supported*

Figure 4: A) Element pair performance (F1 score %) of graph building relative to DFT-derived connectivity using a 0.2 bond order cut-off. Heavy *p*-group elements are grouped for clarity. An asterisk (*) denotes that these groupings exclude first row *p*-block elements (<Ne) which are shown individually. Marker size is proportional to the frequency of the element-pair, coloured by F1 score. **B)** Example of the molecular graph building for a bimetallic metal hydride complex. The graph is schematically drawn with nodes (atoms) as dots and detected edges (bonds) as lines. C–H hydrogens are hidden for clarity.

3 Results & Discussion

3.1 Graph Building

To ensure a reliable translation from Cartesian into internal coordinates, we first validated the molecular graph-building approach across diverse chemical space. Performance was evaluated using two computational datasets that provide QM-derived bond orders. The first was the `GMTKN55` dataset, developed by Goerigk and Grimme,[27] containing 2346 geometries spanning a broad range of organic and main group chemistries designed for functional benchmarking. Because we require a consistent graph-building procedure across the whole periodic table, we also extended the validation to include a stratified sample of 2500 organometallic complexes from the `tmQM` datasets by Balcells.[28]

Our method successfully generated valid molecular graphs for all of the 4846 structures, spanning 61 elements and including 490 element-pair bond types. Graph connectivity agrees closely with DFT-derived bond connectivity, achieving 98.4% detection even across exotic bonding environments (**Table S1**). Performance is comparable to

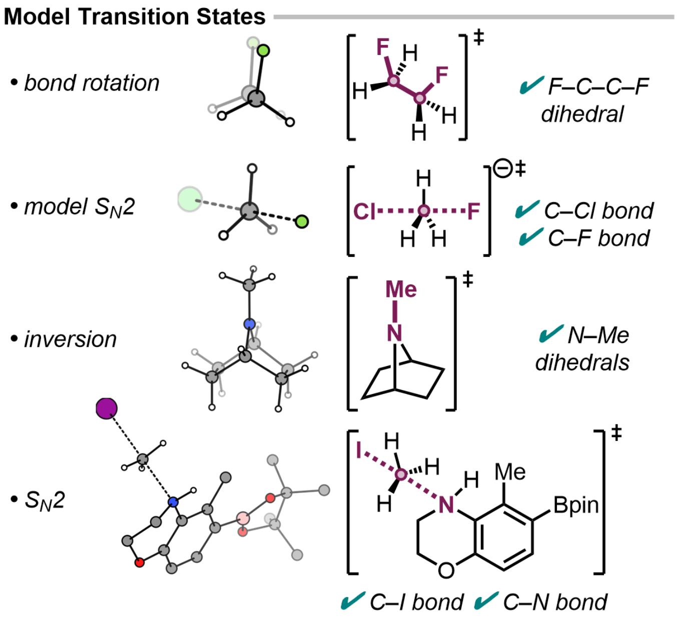


Figure 5: Automated `graphRC` analysis on model transition states. Visualisation of structures using references [29, 30].

established tools such as `RDKit` (`xyz2mol`) and `xyz2mol_tm`, with 99.8% agreement on `GMTKN55` and 97.1% on `tmQM`, slightly reduced due to detecting more M−L bonds with `xyzgraph`. These differences are strongly influenced by the chosen bond order cut-off for single bond definitions (**Figures S1–S4**). At a bond order cut-off of 0.2, we examined the bonding performance in more detail (**Figure 4A**). This shows an extremely high accuracy for a wide range of bonding types with some reduced accuracy for unusual bonding patterns of *s*-group metals. The comparatively low performance of C−C (F1 score ≈ 80%) is an artefact of the chosen bond order cut-off, and `xyzgraph` is consistent with the performance of `xyz2mol` (see SI).

Overall, these results demonstrate that the graph-building procedure is robust across the entire periodic table and provides a reliable foundation for internal coordinate analysis. A schematic example of the molecular graph built from a bimetallic hydridic complex with multidentate bonding is shown in **Figure 4B**, with further examples in **Figure S6**.

3.2 Development of Vibrational Analysis

Having established a robust internal coordinate approach, we next evaluate the vibrational analysis workflow using four simple and well-known transition state examples including dihedral rotation, pyramidal inversion and concerted S_N2 reactions (**Figure 5**). For these small systems, visual inspection of the normal mode displacement is sufficient for validation given the limited number of internal coordinates and well-defined transformations. Automated analysis reproduced these results with full accuracy, identifying the relevant bond, angle and dihedral internal coordinate changes. By incorporating elongated bonds into a set of chemically valid internal coordinates, the method captures connectivity at the non-equilibrium TS geometry, allowing for identification of the TS mode. Post-processing angle and dihedral changes characterises pyramidal inversions and bond rotations, including methyl rotations (output examples in SI).

To demonstrate generalisability, we extended the approach to more complex transition states (**Figure 6**) from

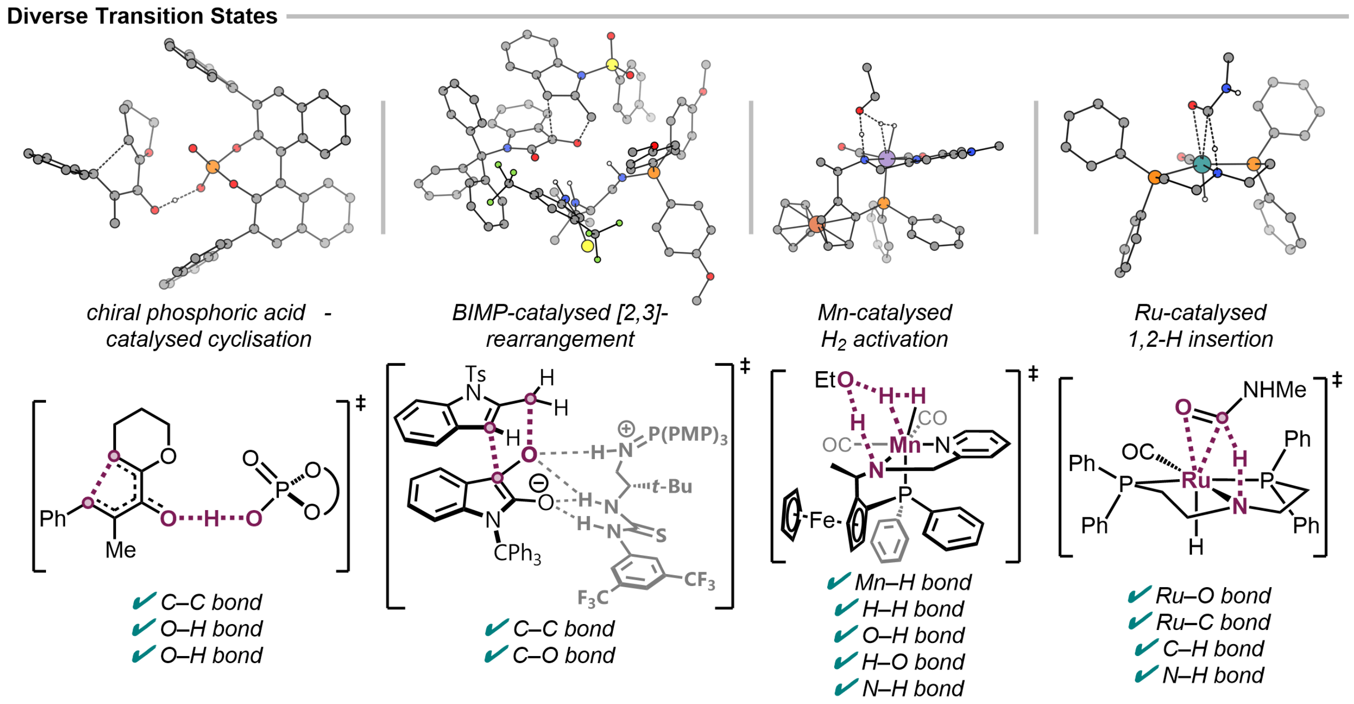


Figure 6: Examples of `graphRC` performance on IRC verified transition states. Each TS is fully characterised by the automated analysis with full detection of the internal coordinate changes and no false positive detections (see also **Figures S7–S10**).[31–34]

previously published mechanistic studies where IRC or QRC validation data was available. These examples span a wide range of chemical transformations and structural motifs, presenting unique challenges for both vibrational analysis and graph building with an order of magnitude larger internal coordinate representation than the model systems. For instance, the chiral phosphoric acid catalysed Nazarov cyclisation involves a proton transfer-initiated cyclisation, with three coupled bond changes in the TS. The BIMP-catalysed (bifunctional iminophosphorane organocatalyst) [2,3]-rearrangement features a 'loose' transition state with an asynchronous C–C bond formation, where the degree of asynchronicity is sensitive to the choice of density functional.[31,32,35]

Organometallic systems introduce further complexity for graph building. Unlike organic molecules, where bonding follows well-defined valence rules, transition metal complexes exhibit variable coordination numbers, oxidation states and binding modes. These include non-classical binding modes such as η^n-coordination where a ligand binds through n contiguous atoms, agostic interactions (weak $C-H\cdots M$ coordination) and multi-centre bonding including bridging hydrides. These arrangements lead to interatomic distances and angles outside standard bonding thresholds and violate valence limits. These challenges are compounded in transition states where metal-ligand bonds may be partially formed or broken.

The examples studied here test these challenges directly. The Mn-catalysed H_2 activation involves an η^2–H_2 complex with a ferrocene ligand, alongside five simultaneous bond changes in the TS.[33] The graph builder accurately represents the chemical structure without introducing spurious M–L bonds. The Ru-catalysed 1,2-hydrogen insertion TS involves 3-membered rings with an η^2–C=O arrangement and a proton transfer that occurs late in the reaction coordinate, with only a small displacement along the imaginary mode eigenvector (**Figure S10**).[34] In both examples, the graph builder accurately represents the coordination environment, achieving full accuracy for vibrational analysis.

Additional validation examples exhibit greater structural diversity (**Figures S11–S18**). Atropisomeric hindered

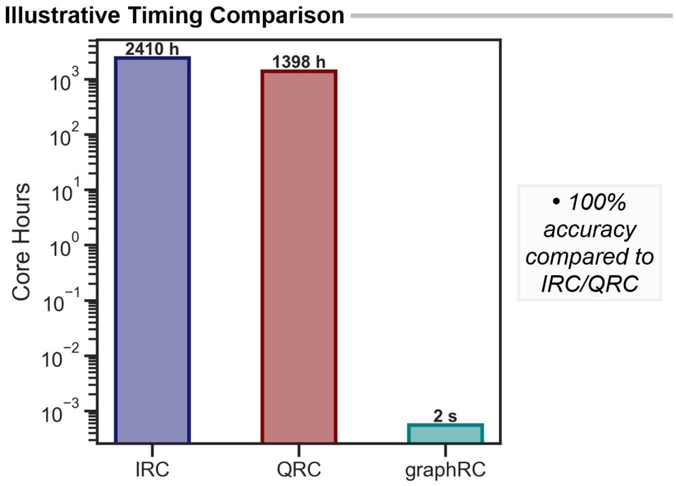


Figure 7: Illustrative comparison of computational time across IRC, QRC and graphRC methods for a BIMP catalysed [2, 3]-rearrangement.

rotations are extremely challenging to characterise using IRC calculations due to low magnitude imaginary modes.[36] As a result, normal mode displacement produces only a small Cartesian displacement, and we apply a secondary round of 50% relaxed thresholds which leads to a fully consistent characterisation of bond rotations in the TS. Spirocyclic transition state geometries present another unique challenge, with a 5-membered ring attached to a 4-membered ring. This was fully characterised without introducing spurious cross-linking that would lead to the wrong TS mode identification.[37]

These 16 TS examples are summarised in **Table 1**, covering proton transfers, C−C and C−X bond formations and cleavage, ring rearrangements, cyclisations, hindered bond rotations and pyramidal inversions across organic and organometallic transitions. Across all examples, the method achieved 100% detection of relevant internal coordinate changes with zero false positives, verified against both IRC and QRC connectivity.

3.3 Scalability and Threshold Optimisation

Following validation, we assess the computational efficiency, which is critical for high-throughput workflows. As an illustrative comparison, the 172 atom BIMP transition state required ∼151 hours of wall time for the IRC calculation on 16 cores (≈ 2410 core hours) while a QRC calculation required ∼44 hours on 32 cores (≈ 1398 core hours). In contrast, `graphRC` analysis of the optimised TS completed in roughly 2 seconds on a single core, approximately six orders of magnitude faster, while retaining full accuracy for bond detection. This difference illustrates the scalability of this approach: as IRC and QRC calculations become prohibitively expensive, the cost of `graphRC` remains negligible (**Figure 7**).

A bond detection threshold defines the minimum bond length change required to flag a bond as forming or breaking during the vibrational mode. The magnitude of these changes depends on the displacement amplitude used when projecting the imaginary mode eigenvector onto the TS geometry, as a larger displacement produces larger Cartesian changes. A fixed displacement amplitude is used throughout to produce chemically reasonable displaced structures, and the detection threshold was optimised at this amplitude. Scanning the bond change threshold from 0.1 Å to 0.5 Å across all 16 examples from **Table 1** revealed that 0.4 Å provides an optimal balance, achieving

Table 1: 16 transition state examples analysed in detail. These TSs were selected from a combination of previous mechanistic work and broader literature examples to provide a diverse and challenging set including simple model systems and complex organometallic transformations. Complexity shown as filled dots (●) out of three with the number of internal coordinate changes in the TS in parentheses. Accuracy reported as the F1 score (harmonic mean of precision and recall). VI = Visual Inspection, *pt* = proton transfer.

TS	Transformation	Class	Complexity	Features	Accuracy	Validation
dihedral rotation	F−C−C−F rotation	model	● ○ ○ (1)	low-barrier torsion	100%	VI
model S_N2	C−Cl break + C−F form	model	● ○ ○ (2)	concerted	100%	VI
pyramidal inversion	N inversion	model	● ○ ○ (1)	atom inversion	100%	VI
S_N2	C−I break + C−N form	organic	● ○ ○ (2)	concerted	100%	VI
Nazarov cyclisation[31]	C−C ring close + O−H−O *pt*	organic	● ● ○ (3)	proton transfer + ring close	100%	IRC
[2,3]-rearrangement[32]	C−O break + C−C form	organic	● ● ○ (2)	'loose' TS, asynchronous	100%	IRC
Mn-H_2 activation[33]	H−H break + N−H−O *pt*	organo-metallic	● ● ● (5)	η^2-H_2 bonding + solvent	100%	IRC
Mn-catalysed reduction[33]	Mn−H−C hydride transfer	organo-metallic	● ● ○ (2)	metal binding	100%	IRC
Aza-Morita-Baylis-Hillman[38]	C−H−N *pt*	organic	● ○ ○ (2)	proton transfer	100%	IRC
Ru-catalysed insertion[34]	M−L rearr. + N−H−C *pt*	organo-metallic	● ● ● (4)	M−L 3mem ring + rearrangement	100%	IRC
thia-Michael addition[36]	C−S form	organic	● ○ ○ (1)	sulfur nucleophile	100%	IRC
atropisomeric rotation[36]	hindered bond rotation	organic	● ● ○ (2)	low-frequency mode	100%	QRC
Ru methylamine elimination[34]	C−N break + N−H−N *pt*	organo-metallic	● ● ○ (3)	asynchronous	100%	QRC
cyclisation[36]	C−N ring close	organic	● ○ ○ (1)	ring closure	100%	QRC
spirocyclic Nuc-attack[37]	C−N form	organic	● ● ○ (1)	small rings	100%	QRC
spirocyclic ring opening[37]	C−O break	organic	● ● ○ (1)	small rings	100%	QRC

full detection of expected bond changes with zero false positives (**Figure 8**). Below this value, small-amplitude displacements are incorrectly identified as bond changes, and above it, true bond changes start to be missed. This choice prioritises precision over exhaustive detection, avoiding over-interpretation of Cartesian displacements for

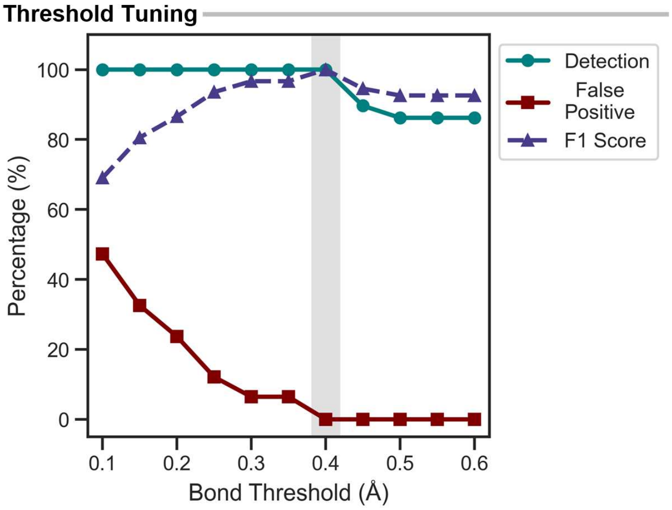


Figure 8: Threshold optimisation metrics including detection (% of expected bonds found), false positive (% of spurious detections), and F1 score (%) (balanced measure combining both metrics, expressed as a %). Optimal default threshold highlighted in grey (0.40).

reliable high-throughput analysis. While fully accurate for the examples above, this threshold may miss highly asynchronous bond coordinate changes where the displacement along the imaginary mode is small (*vide infra*).

3.4 Limitations

While the optimised threshold performs very well across the examples studied, highly asynchronous or late bond formation can still pose challenges (**Figure 9**). This reflects the inherent nature of normal mode analysis, which approximates the transition as a single displacement vector at the saddle point. While this linear projection captures the primary transformation accurately, it cannot fully describe the complex multidimensional surface of highly asynchronous transition states. To stress-test the methodology, we applied it to two literature examples of strongly asynchronous concerted transformations verified by IRC. In both, the automated analysis correctly identifies the primary internal coordinate change but struggles to fully capture the asynchronous changes.

In the epoxide ring expansion reported by Trujillo *et al.*,[39] the analysis correctly identifies the epoxide O−C ring opening but does not detect the strongly asynchronous C−O bond formation in this TS. This bond forms late along the reaction coordinate and has a bond length of 2.9 Å in the TS, beyond the default threshold for constructing the internal coordinate. Relaxing connectivity criteria allows this bond coordinate to be identified; however, this introduces false positive internal coordinate changes that reduce confidence and chemical interpretability. In such cases, we argue that a partial identification of the primary transformation is preferable to over-detection. Analysis of the IRC trajectory fully captures the TS connectivity, capturing both the epoxide ring opening and 5-membered ring closing in the TS mode. Similarly, in the Cu(II) open-shell transmetallation by Macgregor *et al.*,[40] the method detects the primary C−B bond cleavage but misses the Cu−C and Cu−O bond changes unless the threshold is reduced to 0.3 Å.

These cases highlight the limitation of normal mode analysis, which cannot fully capture asynchronous multidimensional behaviour. For application to rapid mechanistic screening and automated workflows, prioritising the primary transformation is preferable to an exhaustive detection that may introduce artefacts. Where asynchronous

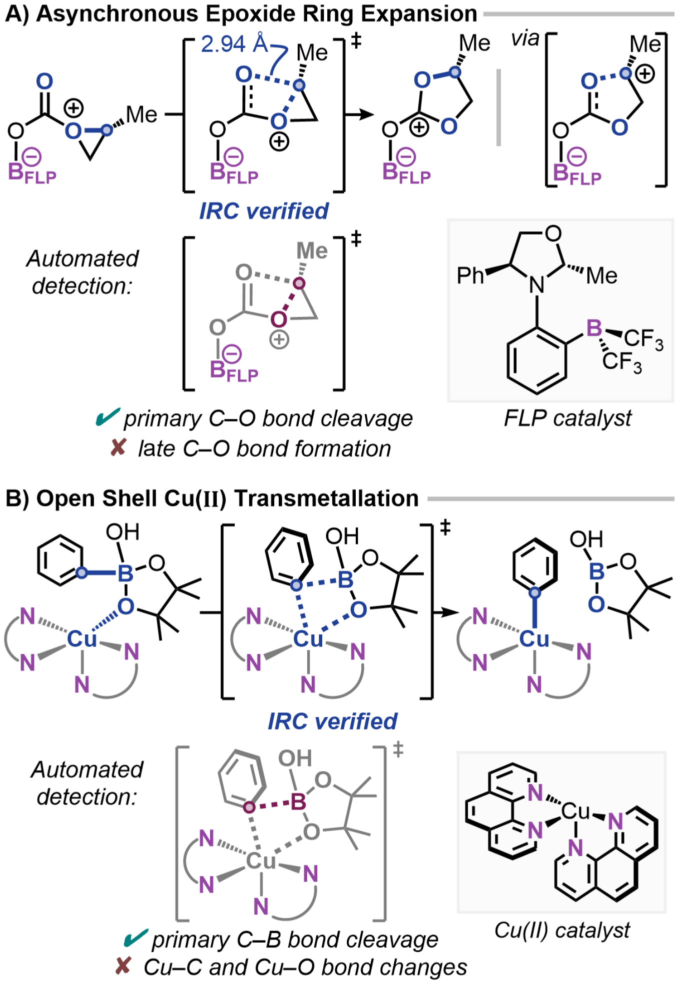


Figure 9: Examples that push the limits of the normal mode displacement approach compared to IRC connectivity. **A)** Epoxide ring expansion.[39] **B)** Open shell Cu(II) transmetallation.[40]

transitions are present, a short QRC path can complement this approach, enabling interpretable internal coordinate analysis across IRC and QRC pathways.

3.5 High-Throughput Validation

To further assess the robustness of this approach, we extend the analysis across 395 transition states from reference [41, 42], which provides GFN2-xTB[14] IRC trajectories for transition states mined from the supporting information of published studies. These 395 TS contain 952 vibrational bonds, including 35 elements involved in 89 unique bond types (**Figure 10**).

Using only normal mode displacement, the method achieves an F1 score of 88.4% and correctly identifies the primary bond change in 100.0% of transition states compared to full IRC calculations. Across the bond changes identified by IRC, 81.8% are detected while maintaining a high precision of 96.1% (false positive rate 3.9%). This diverse dataset covers a variety of organic, organometallic and catalytic examples, including proton transfers, Diels-Alder, various M−L rearrangements and ring opening and closing (20 random examples shown in **Figures S19–S22**). In more detail, element-pair performance is shown in **Figure 10**. Common organic transformations involving O−H, N−H, C−H, C−C, C−O and C−N dominate and are identified with high accuracy. Organometallic reactivity is

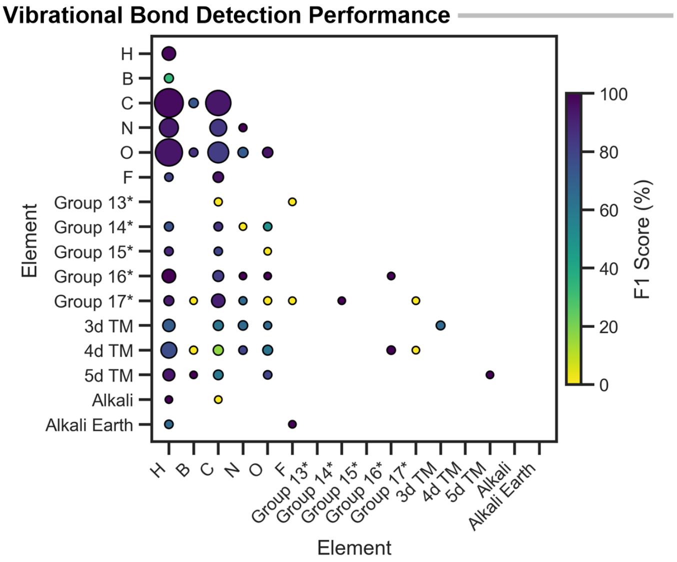
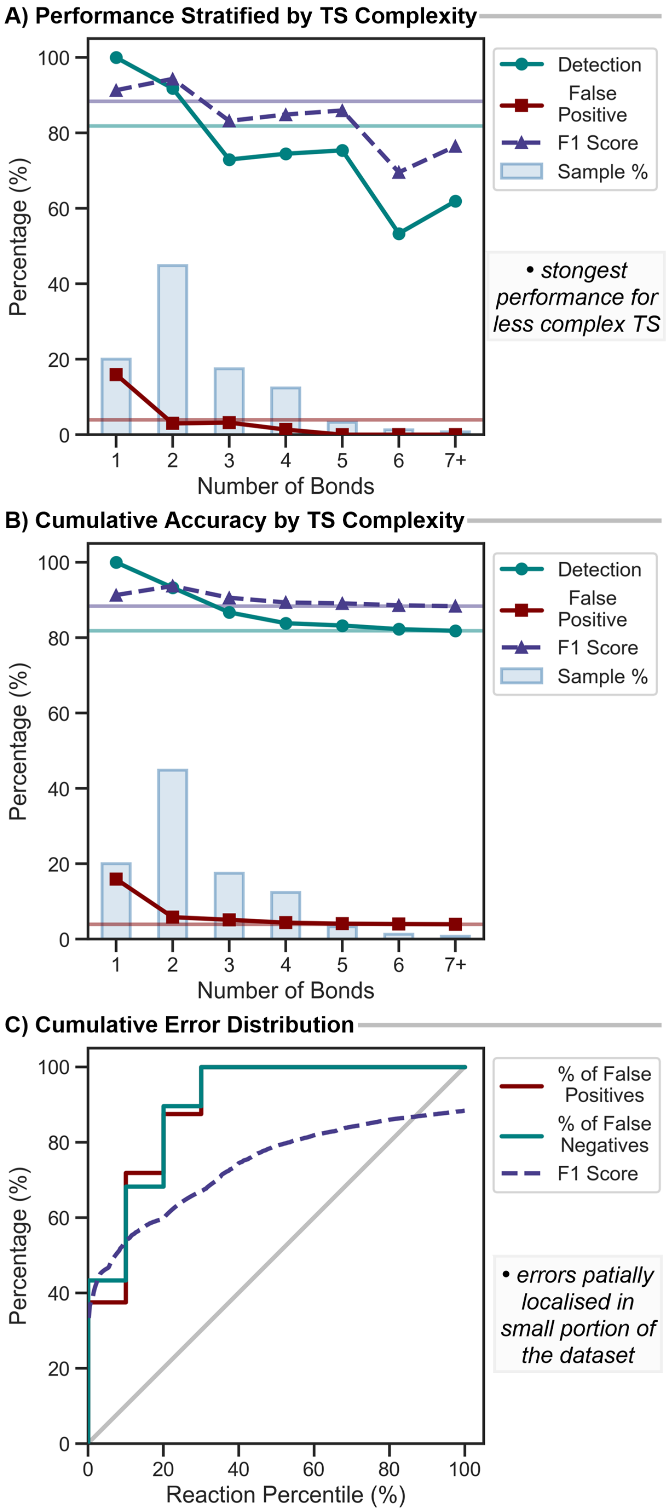


Figure 10: Element pair performance (F1 score %) of automated vibrational analysis relative to GFN2-xTB derived IRC connectivity. Heavy *p*-group elements are grouped for clarity. An asterisk (*) denotes that these groupings exclude first row *p*-block elements (<Ne) which are shown individually. Marker size is proportional to the frequency of the element-pair, coloured by F1 score.

also represented, including M−H activation and M−L coordination changes. A reduced performance is observed for boron and some heavier elements; however, this may partly reflect limitations of the GFN2-xTB level of theory. Overall, these 395 TS cover a wide range of relevant chemical reactivity with fully validated IRC bond changes. Extended analysis under relaxed TS criteria increases elemental coverage but includes TS that are not strictly IRC-validated, reducing confidence in the underlying IRC-based ground truth (**Figure S25**).

Stratifying by the number of bond changes reveals that detection accuracy decreases with increasing complexity (**Figure 11A**, **Table S2**). 100.0% of single-bond changes are detected, though detection drops for more complex transformations with >6 bond changes. False positives remain low overall and are negligible for TS with >2 bond changes, though they are slightly more frequent for TS with one bond change event. This behaviour reflects the limitation of normal mode displacement, which uses a single vector projection to approximate a multidimensional transformation. Despite the reduced accuracy for the most complex examples with >6 bond changes, the overall performance is dominated by less complex TS (**Figure 11B**). The most common category, with two bond changes (44.8% of the dataset), achieves 91.8% detection with 97% precision. These metrics confirm that the method maintains high accuracy across the majority of transition states while prioritising precision over exhaustive detection.

Additionally, we suspect that this reduced performance may be partially related to the GFN2-xTB level of theory used for TS identification and IRC path evaluations. GFN2-xTB can favour more concerted transformations compared to higher levels of theory. To investigate, we examined the error distribution across the dataset. Ranking samples by F1 score revealed that the lowest 10% of TS accounted for 42.4% of all errors, including 37.5% of false positives and 43.4% of false negatives (**Figure 11C**). Inspection of the six TS with F1 scores <50% (**Figures S23–S24**[43–48]) showed a mixture of highly asynchronous, complex, multi-bond rearrangements which reflect two related limitations. Normal mode displacement cannot fully capture these transformations (*vide supra*) and the concertedness of these reactions may be exaggerated by the GFN2-xTB level of theory used for TS and IRC evaluation. Exclusion of the 19 worst-performing cases (bottom 5%) achieved yielded 89.1% detection with only 2.4% false positives and an F1 score



A) Performance Stratified by TS Complexity

• stongest performance for less complex TS

B) Cumulative Accuracy by TS Complexity

C) Cumulative Error Distribution

• errors patially localised in small portion of the dataset

Figure 11: Element-pair performance (F1 score %) compared to GFN2-xTB derived IRC connectivity. **A)** Performance metrics stratified by the number of bond changes in each transition state. Average shown with horizontal lines. Bars show the distribution of TS complexity. **B)** Cumulative performance across increasing TS complexity. **C)** Cumulative error distribution across dataset percentiles ranked by F1 score. Stepped lines represent cumulative percentages of false positives and false negatives. F1 score is shown as a rolling average. Identity line shown in grey.

of 91.5%. These 19 cases alone account for a quarter (24.9%) of the total errors, highlighting that the overall metrics understate the performance of the method for typical cases.

Overall, these results confirm that the approach is appropriate for transition state analysis with strong performance across a wide range of transformations in complex bonding scenarios. Notably, all 411 examples achieved 100.0% detection of the primary bond transformation, with 96.1% precision across 984 bond change events in 395 high-throughput IRC validated examples. This implementation offers a rapid tool for mechanistic studies, characterising transition states, facilitating QRC calculations and formal analysis of QRC and IRC trajectories. The structured output of the code supports integration into high-throughput campaigns and can be incorporated into feedback loops to refine geometries when secondary imaginary modes, such as methyl rotations, are present. Combined, these features enable scalable automated TS analysis that reduces the computational cost by orders of magnitude compared to IRC pathways while retaining a high accuracy across diverse reaction classes.

4 Conclusion

`graphRC` offers a fast, reliable, and generalisable approach to transition state (TS) analysis, validated across a broad variety of organic and organometallic transformations. Graph-based internal coordinate construction through `xyzgraph` enables the accurate identification of bond formation, cleavage, rotations, and inversions, while avoiding false positives in complex or low-magnitude vibrational modes. We have demonstrated high accuracy in identifying key internal coordinate changes, and structured outputs make it particularly suited to automated workflows. The practical applications are threefold:

1. **Mechanistic analysis**: Rapid verification of the TS mode prior to formal IRC or QRC calculations. Analysis of non-equilibrium structures with multiple imaginary frequencies. Generation of displaced geometries for eliminating small imaginary modes (*c.f.* `pyQRC`[9]).
2. **High-throughput TS campaigns**: A structured output provides machine-readable internal coordinate changes that can be integrated into automated pipelines that make use of atom-mapping or adjacency matrices for programmatic TS verification. These features support screening workflows that terminate upon the successful detection of targeted bond changes, and flexible displacement along vibrational modes supports the elimination of small imaginary frequencies, streamlining TS workflows.
3. **Reaction coordinate analysis**: Applied to IRC or QRC trajectories, the analysis translates Cartesian changes into chemically intuitive internal coordinate changes that reflect the underlying reaction mechanism. This can be integrated into high-throughput workflows by performing short, partial optimisations of displaced TS structures for enhanced accuracy compared to normal mode displacement.

Importantly, the tool prioritises accuracy through careful graph building logic, avoiding spurious bonds that would introduce meaningless internal coordinate changes. Detection of highly asynchronous transformations remains a challenge due to the limitations of using normal mode projections; however, a conservative approach ensures that reported changes are chemically meaningful and reliable. We emphasise the importance of IRC and QRC calculations for the true characterisation of transition states.[49] `graphRC` can also be applied to these trajectories, providing a formal internal coordinate description of the reaction coordinate. This tool is designed to supplement these methods, offering rapid mechanistic analysis and supporting high-throughput scenarios where IRC calculations are unfeasible.

Data Availability

Supporting information is available, including: graph building analysis, output examples, IRC and QRC pathways, high-throughput validation workflow and analysis. The `graphRC` software package is available on GitHub (`https://github.com/aligfellow/graphRC.git`), v1.3.5 used in this work. Threshold analysis can be reproduced using scripts in the repository. Further output and usage examples are documented in the GitHub repository. The `xyzgraph` software package is also available on GitHub (`https://github.com/aligfellow/xyzgraph.git`), v1.4.8 used in this work. IRC and QRC data generated in this work is available on Zenodo (`https://doi.org/10.5281/zenodo.17877172`).

Author Contributions

CRediT: **ASG**: Conceptualisation, Investigation, Methodology, Formal analysis, Software, Writing – original draft, Writing – review & editing. **BNN**: Writing – review & editing, Funding acquisition.

Conflicts of Interest

There are no conflicts to declare.

Acknowledgements

The authors acknowledge funding from the EPSRC (EP/X021033/1). This work was undertaken on the Aire HPC system at the University of Leeds, UK.

References

(1) K. Müller, *Angew. Chem. Int. Ed.*, 1980, **19**, 1–13.

(2) K. Fukui, *Acc. Chem. Res.*, 1981, **14**, 363–368.

(3) C. Gonzalez and H. Bernhard Schlegel, *J. Chem. Phys.*, 1989, **90**, 2154–2161.

(4) J. M. Goodman and M. A. Silva, *Tetrahedron Lett.*, 2003, **44**, 8233–8236.

(5) M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeerschd, E. Zurek and G. R. Hutchison, *J. Cheminformatics*, 2012, **4**, 1–17.

(6) G. Schaftenaar and J. H. Noordik, *J. Comput. Aid. Mol. Des.*, 2000, **14**, 123–134.

(7) M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian 16, Revision C.01*, Gaussian Inc., Wallingford CT, 2019.

(8) F. Neese, *WIREs Comput. Mol. Sci.*, 2025, **15**, e70019.

(9) R. S. Paton and G. Luchini, *pyQRC: version 1.0.3*, 2021, `www.doi.org/10.5281/zenodo.4784549`.

(10) E. Berquist, A. Dumi, S. Upadhyay, O. D. Abarbanel, M. Cho, S. Gaur, V. H. Cano Gil, G. R. Hutchison, O. S. Lee, A. S. Rosen, S. Schamnad, F. S. Schneider, C. Steinmann, M. Stolyarchuk, J. E. Vandezande, W. Zak and K. M. Langner, *J. Chem. Phys.*, 2024, **161**, 42501.

(11) T. A. Young, J. J. Silcock, A. J. Sterling and F. Duarte, *Angew. Chem. Int. Ed.*, 2021, **60**, 4266–4274.

(12) Y. Guan, V. M. Ingman, B. J. Rooks and S. E. Wheeler, *J. Chem. Theory Comput.*, 2018, **14**, 5249–5261.

(13) T. Stuyver, *J. Comput. Chem.*, 2024, **45**, 2308–2317.

(14) C. Bannwarth, S. Ehlert and S. Grimme, *J. Chem. Theory Comput*, 2019, **15**, 1652–1671.

(15) C. Peng and H. Bernhard Schlegel, *Isr. J. Chem.*, 1993, **33**, 449–454.

(16) V. Ásgeirsson, B. O. Birgisson, R. Bjornsson, U. Becker, F. Neese, C. Riplinger and H. Jónsson, *J. Chem. Theory Comput.*, 2021, **17**, 4929–4945.

(17) J. Marks and J. Gomes, *J. Chem. Theory Comput.*, 2025, **21**, 12110–12120.

(18) D. S. Levine, M. Shuaibi, E. W. C. Spotte-Smith, M. G. Taylor, M. R. Hasyim, K. Michel, I. Batatia, G. Csányi, M. Dzamba, P. Eastman, N. C. Frey, X. Fu, V. Gharakhanyan, A. S. Krishnapriyan, J. A. Rackers, S. Raja, A. Rizvi, A. S. Rosen, Z. Ulissi, S. Vargas, C. L. Zitnick, S. M. Blau and B. M. Wood, *The Open Molecules 2025 (OMol25) Dataset, Evaluations, and Models*, 2025, `https://arxiv.org/abs/2505.08762`.

(19) L. D. Jacobson, A. D. Bochevarov, M. A. Watson, T. F. Hughes, D. Rinaldo, S. Ehrlich, T. B. Steinbrecher, S. Vaitheeswaran, D. M. Philipp, M. D. Halls and R. A. Friesner, *J. Chem. Theory Comput.*, 2017, **13**, 5780–5797.

(20) M. A. Short, C. A. Tovee, C. E. Willans and B. N. Nguyen, *Catal. Sci. Technol.*, 2023, **13**, 2407–2420.

(21) J. H. Jensen, *xyz2mol*, 2021, `https://github.com/jensengroup/xyz2mol.git`.

(22) M. H. Rasmussen, M. Strandgaard, J. Seumer, L. K. Hemmingsen, A. Frei, D. Balcells and J. H. Jensen, *J. Cheminformatics*, 2025, **17**, 1–13.

(23) G. Landrum, *RDKit: Open-source cheminformatics*, `https://www.rdkit.org`.

(24) Y. Kim and W. Y. Kim, *B. Kor. Chem. Soc.*, 2015, **36**, 1769–1777.

(25) A. S. Goodfellow, *xyzgraph: Molecular Graph Construction from Cartesian Coordinates*, 2025, `https://github.com/aligfellow/xyzgraph.git`.

(26) J. Charry and A. Tkatchenko, *J. Chem. Theory Comput.*, 2024, **20**, 7469–7478.

(27) L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi and S. Grimme, *Phys. Chem. Chem. Phys.*, 2017, **19**, 32184–32215.

(28) D. Balcells and B. B. Skjelstad, *J. Chem. Inf. Model.*, 2020, **60**, 6135–6146.

(29) K. Briling, *v: A simple X11 molecular viewer*, 2025, `https://github.com/briling/v.git`.

(30) K. Briling, *xyz2svg: A lightweight script to make vector images of molecules*, 2025, `https://github.com/briling/xyz2svg.git`.

(31) Y. P. Chin and E. H. Krenske, *J. Org. Chem.*, 2021, **87**, 1710–1722.

(32) T. Kang, J. O'Yang, K. Kasten, S. S. Allsop, T. Lewis-Atwell, E. H. Farrar, M. Juhl, D. B. Cordes, A. P. McKay, M. N. Grayson and A. D. Smith, *Nat. Chem.*, 2026, 1–10.

(33) A. S. Goodfellow, M. L. Clarke and M. Bühl, *Chem. Eur. J.*, 2025, **31**, e202501063.

(34) J. Luk, A. S. Goodfellow, N. D. More, M. Bühl and A. Kumar, *Chem. Sci.*, 2024, **15**, 16594–16604.

(35) B. Biswas, S. C. Collins and D. A. Singleton, *J. Am. Chem. Soc.*, 2014, **136**, 3740–3743.

(36) A. J. Nimmo, A. S. Goodfellow, J. T. Guntley, A. P. Mckay, D. B. Cordes, M. Bühl and A. D. Smith, *Chem. Sci.*, 2025, **16**, 10494.

(37) A. Conboy, A. S. Goodfellow, K. Kasten, J. Dunne, D. B. Cordes, M. Bühl and A. D. Smith, *Chem. Sci.*, 2024, 52–69.

(38) D. Roy, C. Patel and R. B. Sunoj, *J. Org. Chem*, 2009, **74**, 6936–6943.

(39) M. Ferrer, I. Iribarren, T. Renningholtz, I. Alkorta and C. Trujillo, *Beilstein J. Org. Chem.*, 2024, **20**, 2668–2681.

(40) M. J. Andrews, A. Carpentier, A. M. Slawin, D. B. Cordes, S. A. Macgregor and A. J. Watson, *ACS Catal.*, 2023, **13**, 11117–11126.

(41) Z. Li, I. Y. Yang and B. M. Savoie, *ChemRxiv*, 2025, DOI: `10.26434/chemrxiv-2025-ccgfs`.

(42) Q. Zhao and B. M. Savoie, *Nat. Comput. Sci.*, 2021, **1**, 479–490.

(43) S. Mirzaei, A. A. Taherpour and H. Khalilian, *ChemistrySelect*, 2018, **3**, 6042–6049.

(44) B. Barcs, L. László, L. Lászlókollár, T. Tamás, K. Kégl and J. J. Szentágothai, *Organometallics*, 2012, **31**, 8082–8097.

(45) Q. Feng, S. Li, Z. Li, Q. Yan, X. Lin, L. Song, X. Zhang, Y.-D. Wu and J. Sun, *J. Am. Chem. Soc*, 2022, **144**, 14846–14855.

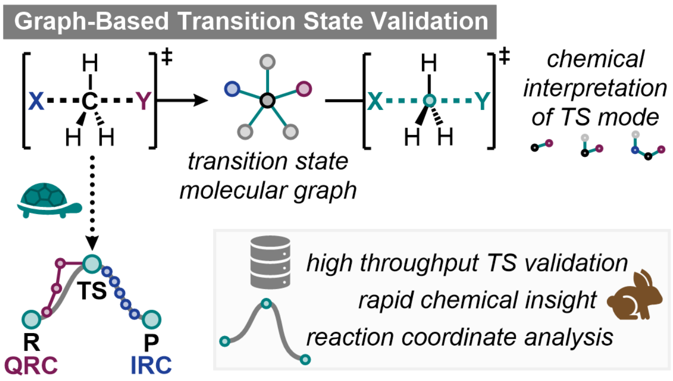
(46) C. X. Cui, J. R. He, L. B. Qu, C. X. Li, J. L. Peng and F. Maseras, *Chem. Eur. J.*, 2024, **30**, e202402572.

(47) A. Boddeda, M. M. Hossain, M. Saeed Mirzaei, S. V. Lindeman, S. Mirzaei and R. Rathore, *Org. Chem. Front*, 2020, **7**, 3215.

(48) D. Šakić, M. Hanževački, D. M. Smith and V. Vrček, *Org. Biomol. Chem*, 2015, **13**, 11752.

(49) M. Bursch, J.-M. Mewes, A. Hansen and S. Grimme, *Angew. Chem. Int. Ed.*, 2022, **61**, e202205735.

ToC Image


Graph-Based Transition State Validation
X
C
Y
H
H H
‡
transition state
molecular graph
X
Y
H
H H
‡
chemical
interpretation
of TS mode
TS
R
P
QRC
IRC
high throughput TS validation
rapid chemical insight
reaction coordinate analysis