

Automated chromophore identification to support prediction of light-mediated reactivity



David J. Ponting, Rachel Hemingway and Martin A. Ott

david.ponting@lhasalimited.org

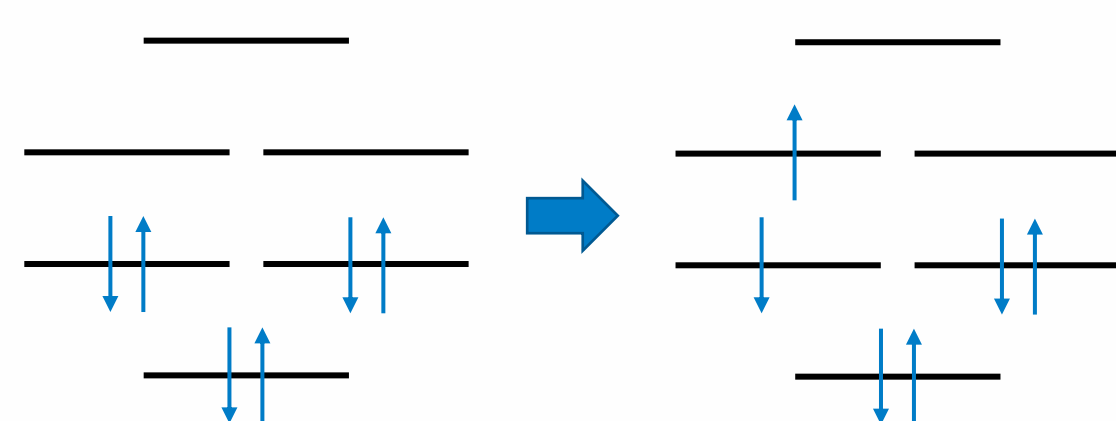
Granary Wharf House, 2 Canal Wharf, Leeds, LS11 5PS

What is a chromophore?

Many reactions in organic chemistry require initiation by absorption of a photon, exciting an electron to form a transient reactive species. Due to the quantisation of the energy levels of electrons within small molecules, the photon must have a specific energy defined by the energies of the molecular orbitals in the chromophore. This is principally defined by the size of the chromophore and the atom types contained therein.

Excitation is however not guaranteed:

- Spin-forbidden vs spin-allowed transitions
- Physical overlap of relevant orbitals



Prediction of the full absorption spectrum is therefore complex, but the chromophore – the system of connected orbitals, typically p-orbitals such as a conjugated π -system in an aromatic compound that can absorb light – can be determined solely from the molecular structure, with each component making the following contributions:

Feature	Atoms	Bonds	Electrons
Double bond	2	1	2
Triple bond	2	1	2
Aromatic bond	System counted as Kekulé structure		
Single bond between contributing features	0	1	0
Atom with exclusively single bonds and a lone pair (including e.g. C ⁻)	1	0	2
Atom with exclusively single bonds and an empty p-orbital (including e.g. C ⁺)	1	0	0
Singly bonded triplet carbene	1	0	1*
Singly bonded singlet carbene	1	0	2 (or 0)
Singly bonded radical	1	0	1*

* While radicals can be of higher multiplicity than doublet, and triplet carbenes are so by definition, at most one electron is likely to be in an orbital that is able to contribute.

The specific size of the chromophore corresponds inversely to the energy of the photons it can absorb, because the frontier orbitals become closer together the more p-orbitals are engaged in the system.

Once the photon is absorbed, light-initiated reactions take several principal forms:

1. Homolytic cleavage of a bond within/adjacent to the chromophore
2. Intra-molecular energy transfer resulting in reactivity at a different site
3. Cyclisations with modified selectivity (according to the Woodward-Hoffmann rules)

For the first and third of these, the atom whose reactivity is investigated should be in a chromophore; for the second, however, the chromophore need only be nearby, leading to additional complication. It is therefore important to know both the size and location within the molecule of any given chromophore.

Pattern-walking

The size of a chromophore is determined by the following steps and associated with each atom in the molecule as a property tag:

- Divide molecular graph by removing features that cannot contribute
 - Atoms where all valency orbitals are in single bonds
 - Single bonds now left dangling by the first step
- Measure size of substructure using the table above
- Associate with each atom in the substructure

The proximity of a chromophore to a potentially reactive atom – since a nearby chromophore can transfer energy to a reaction site - can be defined by pattern-walking. The algorithm by which this is performed can be defined as follows:

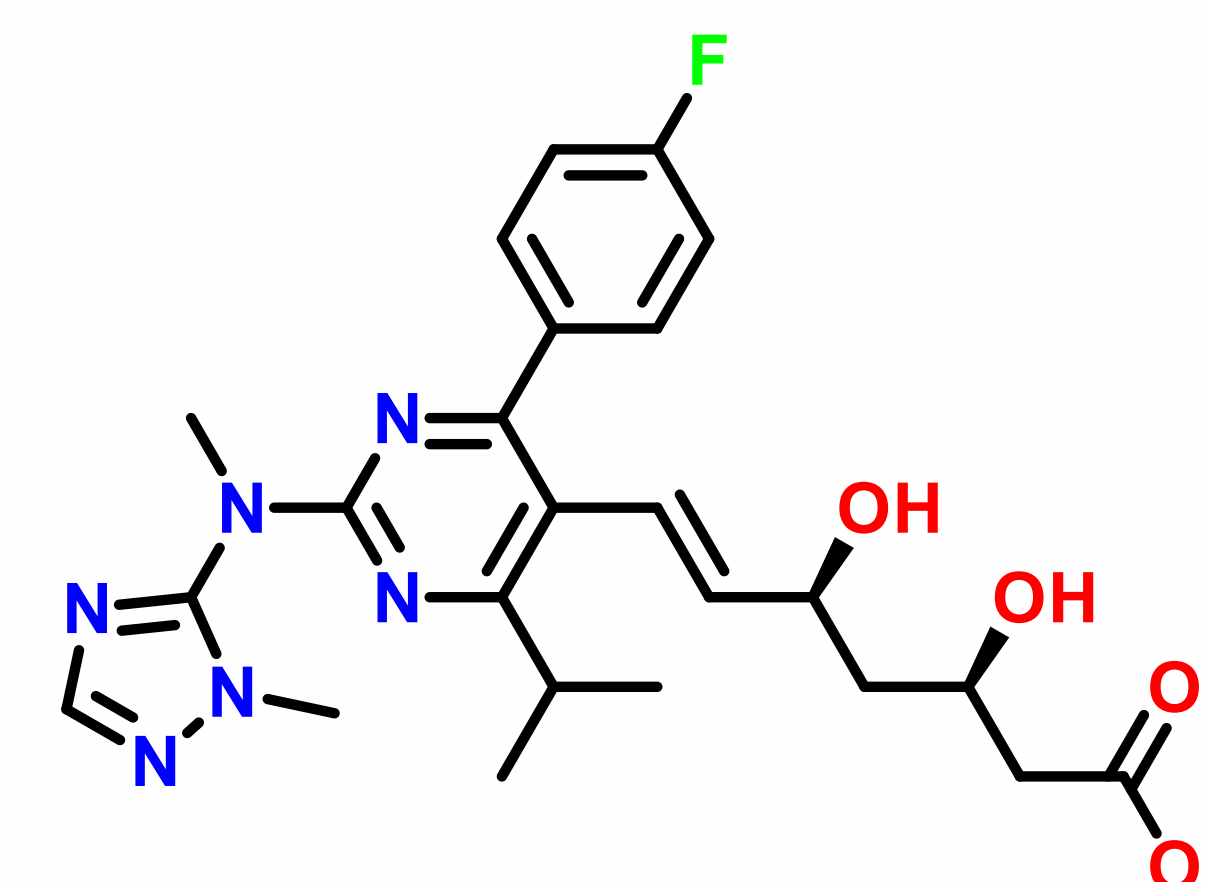
- Start at query atom, recording the atom as visited
- For the atom at the far end of each bond, query size and record visited status
- Repeat for unvisited atoms until entire structure visited or proximity limit reached
- Interrupt if a sufficiently large chromophore identified

References

1. Ahmad *et al* (2008), *J. Med. Chem.*, **51**, 2722-2733
2. Tattersall and Ruan (2006), poster, 232nd ACS National Meeting, San Francisco
3. Parenty *et al* (2013), *Mol. Pharm.*, **10**, 2962-2974
4. Zheng *et al* (2020), *Org. Lett.*, **22**, 6510-6515, Vogt *et al* (2009), *Synthesis*, 4268-4273

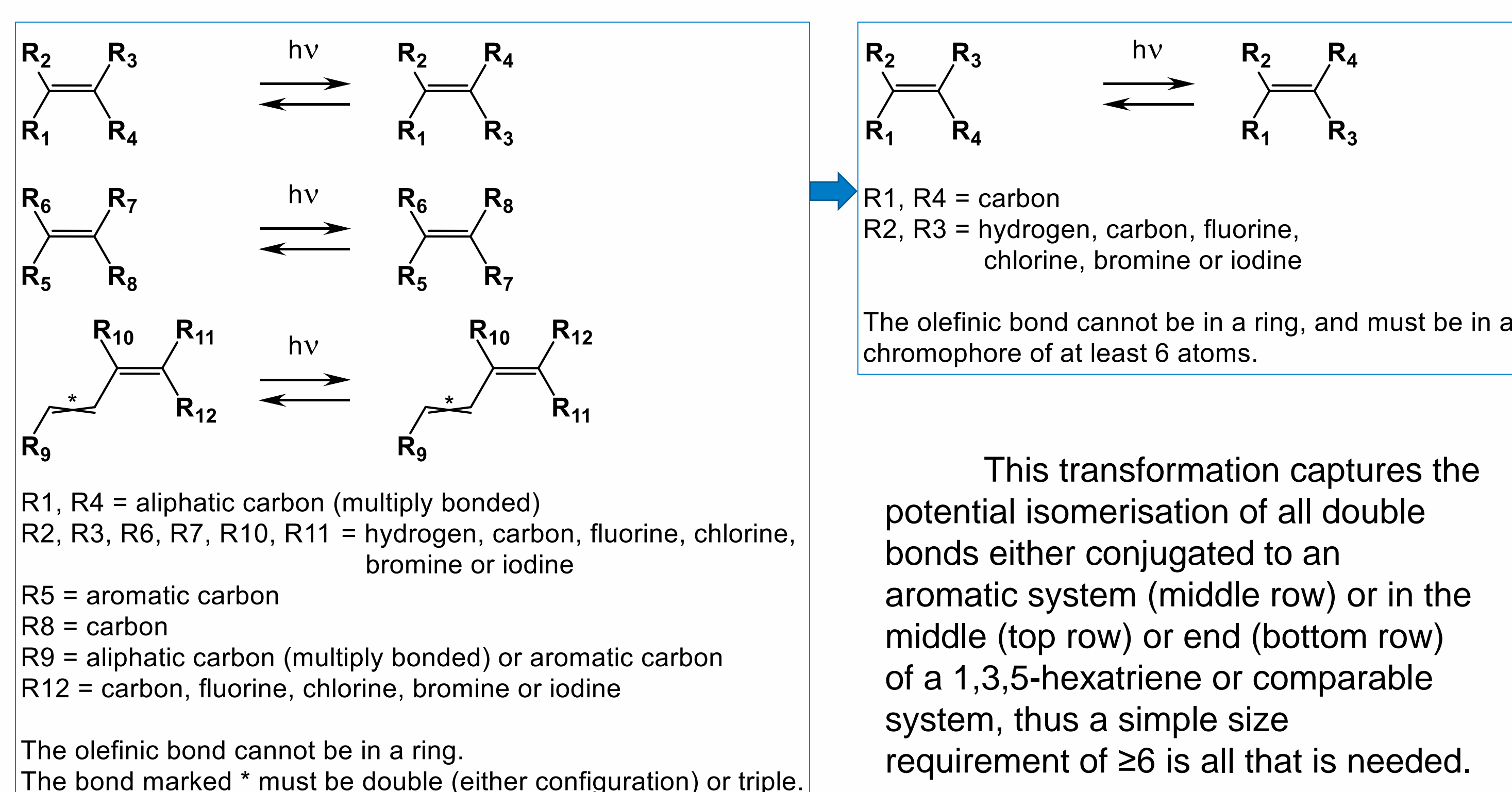
API Example

A published API structure (BMS-644950¹) has been reported to undergo photochemical degradation². The forced-degradation prediction software Zeneth³ (v 8.0, Lhasa Limited) also predicts photochemical degradations for this structure, some of which are significantly over-predicted.



Double bond isomerisation

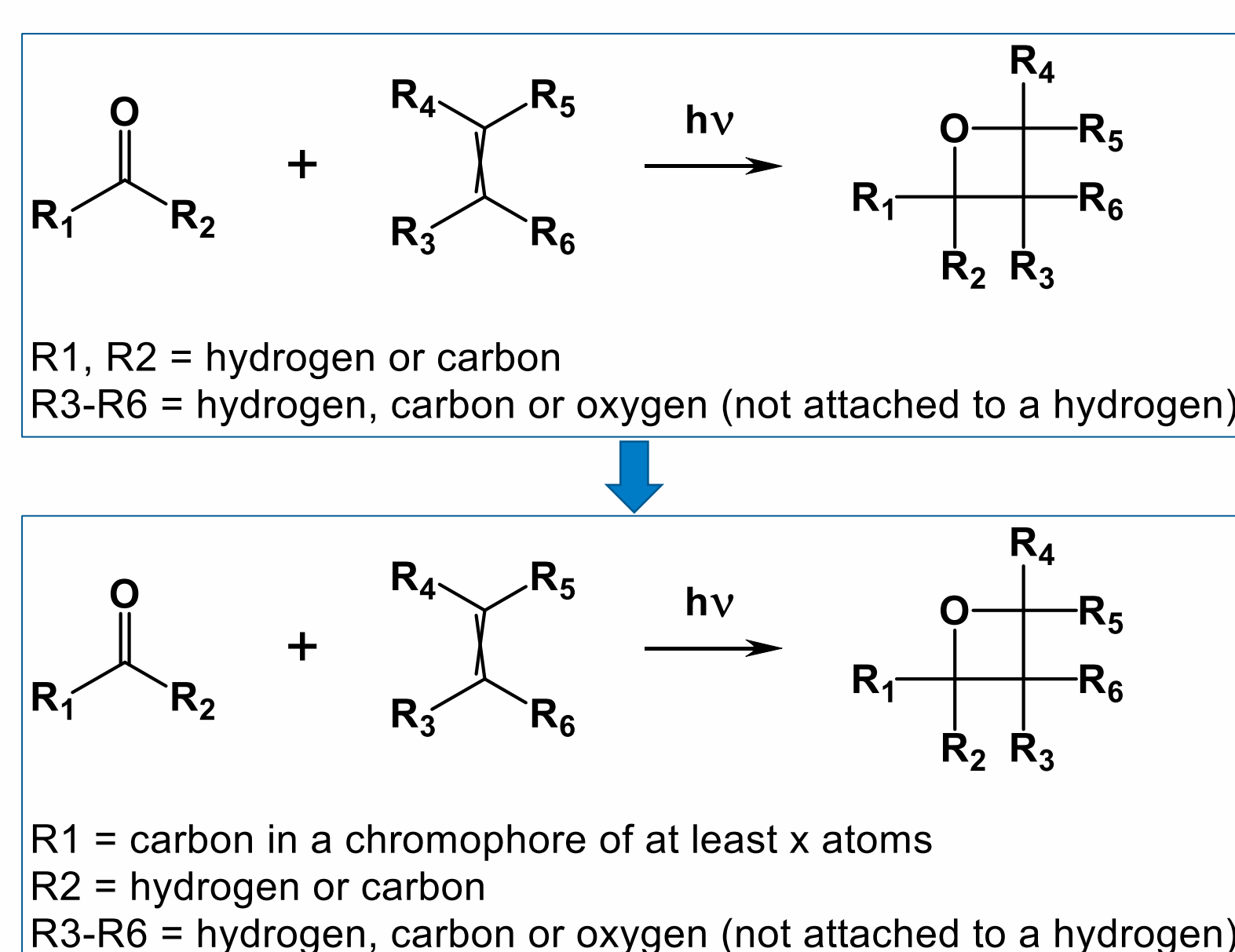
The Zeneth transformation that performs this has been restricted to a set of specific patterns in an attempt to capture the chromophore requirements – which results in a complicated transformation, challenging both to interpret and maintain. With a chromophore predictor, this complicated transformation could be reduced to a far simpler transformation.



Currently (Zeneth 8.0), this transformation creates 1 first-generation and 11 second-generation products for BMS-644950 which are not observed – investigation of this and other molecules could refine the chromophore size requirement and reduce overprediction.

Paternò-Büchi Reaction

For this transformation, a requirement for photoactivation and the ketone to be part of a chromophore is known⁴, but the specific nature of that chromophore is not incorporated in the pattern, since the pattern would end up like the above.



As a result, the pattern is overly general, and will activate for all suitable ketone + alkene pairs. It generates 8 second-generation products (after oxidation of the side-chain alcohols to ketones) for BMS-644950, none of which are observed. Adding a requirement for a chromophore would remove 4 of these, and investigation of chromophore size requirements may remove the other 4.

Conclusions

Without a chromophore predictor, there are two choices for photochemical patterns:

- Very specific patterns with high complexity and specific scope
- Very general patterns which activate for many substrates and over-predict

A chromophore predictor allows us to thread the needle between these, producing better predictive performance.

The chromophore predictor has been implemented into our internal cheminformatics toolkit; no timescale for release is currently available.