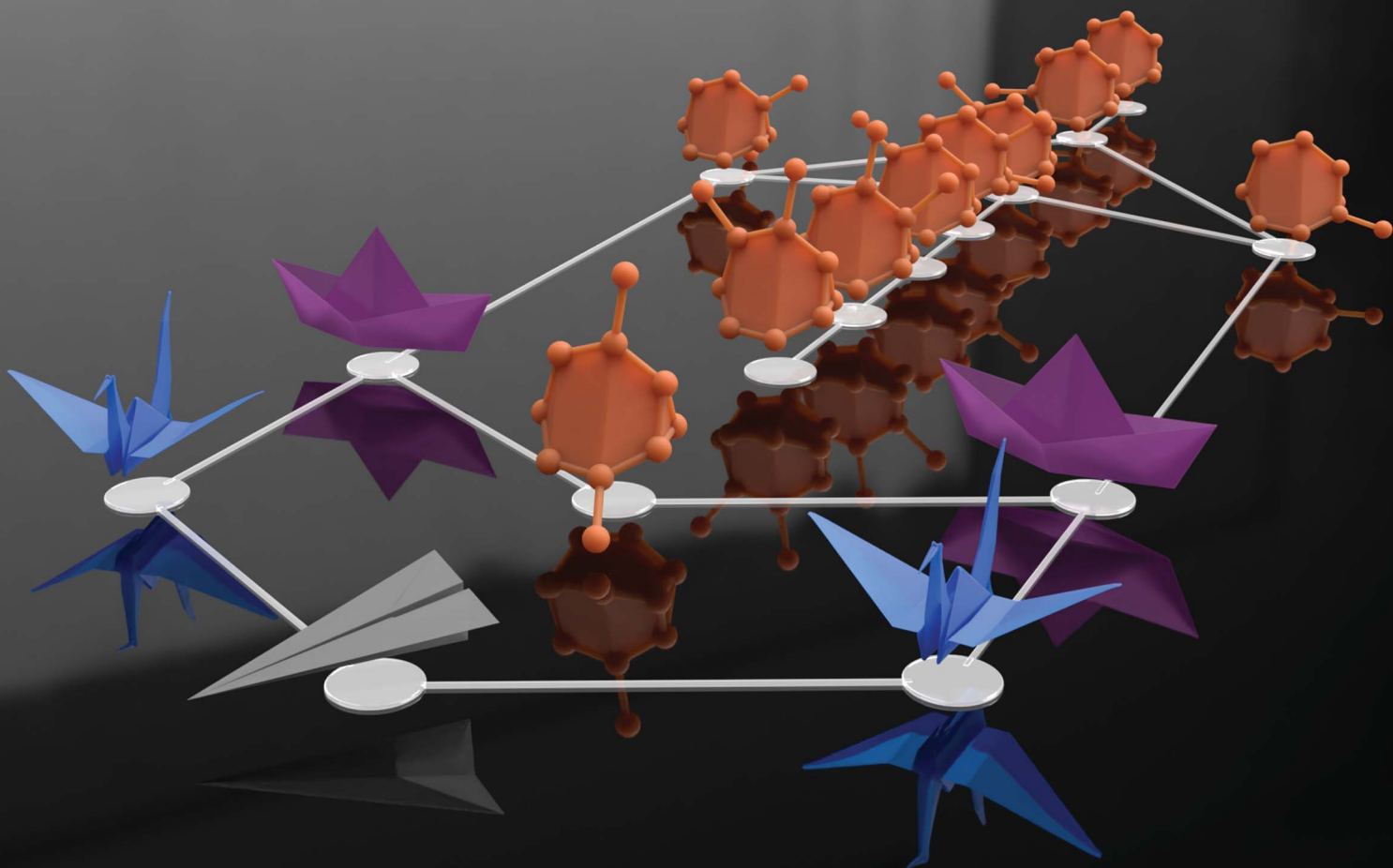


Chemical Science

Volume 15
Number 36
28 September 2024
Pages 14533-14990

rsc.li/chemical-science



ISSN 2041-6539

EDGE ARTICLES

Paul R. McGonigal *et al.*

A guide to bullvalene stereodynamics

Correlated shapeshifting and configurational isomerization

Cite this: *Chem. Sci.*, 2024, 15, 14608 All publication charges for this article have been paid for by the Royal Society of ChemistryReceived 6th June 2024
Accepted 23rd August 2024

DOI: 10.1039/d4sc03700f

rsc.li/chemical-science

A guide to bullvalene stereodynamics†

Robert A. Ives,[‡] William Maturi,[‡] Matthew T. Gill,[‡] Conor Rankine^{*,a} and Paul R. McGonigal^{*,ab}

Here, we analyze the stereodynamic properties of bullvalenes using principal moments of inertia and exit vector plots to draw comparisons with commonly used ring systems in medicinal chemistry. To aid analyses, we first classify (i) the four elementary rearrangement steps available to substituted bullvalenes, which (ii) can be described by applying positional descriptors (α , β , γ , and δ) to the substituents. We also (iii) derive an intuitive equation to calculate the number of isomers for a given bullvalene system. Using DFT-modelled structures for di-, tri-, and tetrasubstituted bullvalenes, generated using a newly developed computational tool (*bullviso*), we show that their 3D shapes and the exit vectors available from the bullvalene scaffold make them comparable to other bioisosteres currently used to replace planar aromatic ring systems in drug discovery. Unlike conventional ring systems, the shapeshifting valence isomerism of bullvalenes gives rise to numerous shapes and substituent relationships attainable as a concentration-independent dynamic covalent library from a single compound. We visualize this property by applying population weightings to the principal moments of inertia and exit vector analyses to reflect the relative thermodynamic stabilities of the available isomers.

Introduction

The C₁₀H₁₀ cage bullvalene (BV, Fig. 1a) fluctuates between 1 209 600 degenerate isomers through rapid and reversible strain-promoted Cope rearrangements.¹ There are now several useful synthetic methods available to prepare substituted derivatives of BV,^{1c,2} which fluctuate between nondegenerate constitutional isomers (Fig. 1) with distinct shapes. This 'shapeshifting' property of substituted BVs and other fluxional molecules has presented opportunities for their inclusion as dynamic structural units in functional molecules³ and materials, such as chemical sensors,^{2d,4} fluorophores,⁵ metal complexes,⁶ components of electromechanical systems,⁷ rigid-rod polymers,⁸ and antibiotics.⁹

In general, building blocks with rigid three-dimensional (3D) structures are key components of materials, such as metal-organic frameworks¹⁰ and other porous networks.¹¹ In the context of medicinal chemistry research—particularly for fragment-based drug discovery¹² (FBDD) libraries—it has been acknowledged that drug candidates based on flat or rod-like molecules offer limited shape diversity.¹³ Consequently, there is a desire for a greater number of diverse 3D fragments to be

included within these libraries to cover more chemical space and, ultimately, to provide better candidates for drug development.¹⁴ One way that this objective can be achieved is by including fragments that possess diverse aliphatic ring systems, particularly as the core ring system is considered the key factor in shape diversity.¹³

The BV ring system has the seemingly contradictory characteristics of, on one hand, being highly dynamic through its reversible Cope rearrangements, while on the other hand, being a rigidly 3D structure. Its tricyclic hydrocarbon skeleton is a shape-persistent structure with substituents projecting outwards at well-defined angles. Therefore, designing effective materials based on BV derivatives requires understanding of their stereodynamics, *i.e.*, their overall 3D shapes, the relative orientations that are accessible to appended substituents, and the relative energies of the isomers at equilibrium.

Here, we quantify the 3D shape diversity accessible from BVs and demonstrate that their fluxional behavior enables them to reversibly access diverse areas of chemical space. We provide a concise guide to the rearrangement processes of substituted BVs and apply computational modelling to categorize and quantify their stereodynamics. To do so, we perform population-weighted principal moment of inertia (PMI)¹⁵ and exit vector (EV)¹⁶ analyses. Much of the analysis can be automated using a new computational tool we have developed, *bullviso*, that generates all the isomers of a given substituted BV, and the input files needed to compute relative energy levels of each isomer. To demonstrate its utility, we apply *bullviso* to examine di-, tri- and tetramethyl BVs. The analysis illustrates that shapeshifting networks of BV derivatives dynamically

^aDepartment of Chemistry, University of York, Heslington, York, YO10 5DD, UK. E-mail: paul.mcgonigal@york.ac.uk

^bDepartment of Chemistry, Durham University, Lower Mountjoy, Stockton Road, Durham, DH1 3LE, UK

† Electronic supplementary information (ESI) available. See DOI: <https://doi.org/10.1039/d4sc03700f>

‡ These authors contributed equally.



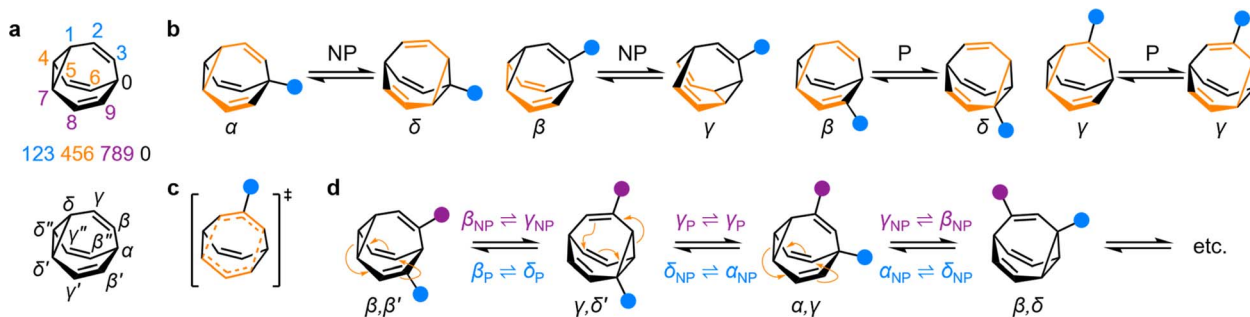


Fig. 1 (a) The BV isomer barcode labelling system, top, and relative positional labels, bottom. (b) The possible exchange processes following one Cope rearrangement step, enumerated for each BV position. Full isomerization requires sequential steps that include (P = participating) and exclude (NP = non-participating) the substituent in the rearranging 1,5-hexadiene motif (shown in orange).^{1b} (c) Illustration of the higher symmetry in the transition state for $\gamma_P \rightleftharpoons \gamma_P$ Cope rearrangement relative to the ground state. (d) The partial isomer network of a hetero-disubstituted BV showing the positional exchange arising from three sequential Cope rearrangement steps.

sample many different areas of chemical space from a single starting compound. They do so by positioning their substituents at a range of angles, extending beyond those typically found in *cis*-disubstituted rings, while maintaining a higher degree of sphericity than most common ring systems found in biologically active molecules.

Results and discussion

Given the large number of possible BV isomers, a method for naming them is essential for any discussion involving their interconversion. Bode and coworkers developed an elegant barcode labelling system that can be parsed by computer algorithms to construct a full network map for interconversion of all the non-degenerate isomers of a substituted BV – a task that is otherwise impractical to do by hand. Each digit of the barcode represents one of the carbon atoms in the BV structure (Fig. 1a) and each type of substituent is given a numeral, enabling each isomer to be described with a unique numeric code (see examples in Fig. 2).¹⁷ Importantly, this system enables isomer information to be coded unambiguously. But while this naming system is ideally suited to comprehensively describing the positions of all substituents in any given isomer, its high level of detail is not always needed. Indeed, perhaps because the length

of the barcodes and their unfamiliar appearance compared to typical nomenclature for organic structures, there has been a tendency in the literature to name bullvalene isomers with individual labels (such as numbers or letters, *e.g.*, isomer A, isomer B, *etc.*) that lack structural information. Therefore, we suggest that Greek letter locants be used as relative positional labels for succinctly discussing distinct positions in the BV structure and their relationships to one another through Cope rearrangement steps. The threefold rotational symmetry of the parent BV scaffold reduces its number of chemically inequivalent positions to four, which are labelled as α – δ (Fig. 1a) starting from α as the apical position, *i.e.*, the unique sp^3 -C, to δ as the cyclopropyl position. The olefin positions, which are typically the energetically favored sites for any non-hydrogen substituents,² are labelled as β and γ . Prime and double prime labels can be used distinguish the same locants on different ‘arms’ of the BV.¹⁸ We find that this labelling system is a useful shorthand that is complementary to the more detailed barcode system. Where appropriate, both systems are used below.

Elementary types of positional exchange

Each Cope rearrangement occurs on one of the three 1,5-hexadiene ‘faces’, involving two arms of the BV. The positions of the BV that are part of the 1,5-hexadiene motif undergoing a given Cope rearrangement can be referred to as participating sites (P), while the other four positions are non-participating sites (NP). Whether a substituent is located on a P or a NP site determines its resulting position following isomerization (Fig. 1b). The α position is never part of the 1,5-hexadiene motif, so it is always considered to be a NP site. On the other hand, for each of the other sites (β , γ , and δ), there exist two potential positional outcomes following a single rearrangement step.

Applying the α – δ and P/NP labels, it becomes clear that there are just four elementary types of positional exchange that occur during any Cope rearrangement step, which are illustrated in Fig. 1b using a monosubstituted BV as a model. Firstly, the sole possible outcome for the α position is migration to a newly formed δ position on the NP arm. A substituent at a β position will migrate to a γ position when on a NP arm, whereas P rearrangement exchanges the β position with a δ site. Finally,

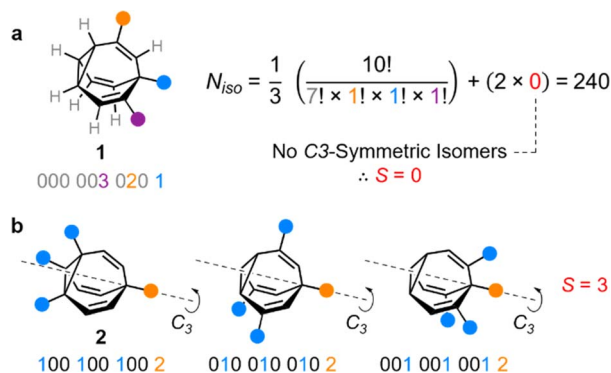


Fig. 2 Structural formulae and isomer barcodes of (a) one of the 240 unique isomers of BV 1, showing the calculation of N_{iso} , and (b) the three C_3 -symmetric isomers of BV 2 that are accounted for by a correction factor of $S = 3$.



a γ substituent remains at a γ position on a P arm following the Cope rearrangement on account of symmetry in the transition state (Fig. 1c). As the Cope rearrangement is reversible, the reciprocal of each of these exchange processes must also occur. Overall, therefore, the elementary types of positional exchange that govern the outcome of any BV rearrangement are (i) $\alpha_{\text{NP}} \rightleftharpoons \delta_{\text{NP}}$, (ii) $\beta_{\text{NP}} \rightleftharpoons \gamma_{\text{NP}}$, (iii) $\beta_{\text{P}} \rightleftharpoons \delta_{\text{P}}$, and (iv) $\gamma_{\text{P}} \rightleftharpoons \gamma_{\text{P}}$. These elementary types of positional exchange apply equally to every BV substituent in every isomer, regardless of the total number of substituents or their relative positions. For example, the functional groups of a heterodisubstituted BV migrate relative to one another during sequential Cope rearrangement steps, allowing them to switch from being on separate arms to occupying the same arm and back again (Fig. 1d).

Calculating the number of unique bullvalene permutations

Deriving the total number of unique nondegenerate isomers for a given substitution pattern is one of the most important considerations for the construction of BV interconversion networks. Bode reported a MATLAB code to calculate the number of unique BV isomers.¹⁷ However, taking account of the symmetry present in BVs, it is possible to perform a simple 'back-of-the-envelope' calculation to determine the number of isomers for a given BV system using eqn (1):

$$N_{\text{iso}} = \frac{1}{3} \left(\frac{10!}{\prod_{a=1}^A N_a!} + 2S \right) \quad (1)$$

where N_{iso} is the number of unique nondegenerate isomers (enantiomers are considered to be distinct from one another), N_a is the number of occurrences of a given type of substituent, $\prod_{a=1}^A N_a!$ the product of multiplying together the factorial of this term for each unique type of substituent (including hydrogen substituents), and S is a correction factor that accounts for the C_3 symmetry of BV and has a value of 0, 1, 3, or 6. In the same manner as the isomer barcode system, eqn (1) treats each substituent as a number within a group of ten numbers.

Combinatorics is used to derive the total number of ways in which these substituents can be ordered. The $1/3$ multiplier adjusts for the fact that the parent BV has C_3 symmetry, offsetting triple counting in the $10!/\prod_{a=1}^A N_a!$ term that arises for structures related by rotation. A further correction factor, $2S$, readjusts for the BV isomers that have three identically substituted arms and so are represented just once each in the $10!/\prod_{a=1}^A N_a!$ term. S is the number of ways that the substituents can be arranged around the BV core to retain its C_3 symmetry, or in other words, where all three arms of the BV possess identical substitution patterns.

Taking heterotrisubstituted BV **1** as a worked example (Fig. 2a), there are seven hydrogen substituents and three distinct non-hydrogen substituents (colored circles). Therefore, the product operation in the denominator of the equation is $7! \times 1! \times 1! \times 1!$. The substituents on **1** cannot be arranged in any pattern that gives C_3 symmetry, hence $S = 0$ and applying eqn (1) gives $N_{\text{iso}} = 240$.

Taking structure **2** (Fig. 2b) as a second worked example, there are six hydrogens, a substituent type that occurs three

times (blue circles), and another substituent type that occurs just once (orange circle), so the product operation is $6! \times 3! \times 1!$. However, three of the substituents of **2** are the same and thus there are three different ways in which they can be arranged such that the BV possesses C_3 symmetry, which are shown in Fig. 2b. These C_3 -symmetric isomers arise when the unique substituent is at the α position and the three identical substituents occupy the same position on each of the three BV arms, *i.e.*, either all β , all γ , or all δ . Therefore, $S = 3$ for **2** and applying eqn (1) gives $N_{\text{iso}} = 282$.

There are 42 possible variations of substituted BVs having different numbers and identities of substituents. Based on all of these possible substituent patterns, a comprehensive reference table is provided in the ESI (Table S1)[†] where this method and eqn (1) have been applied to generate N_{iso} and S , as well as the numbers of achiral and chiral isomers.

Generating bullvalene isomers using *bullviso*

We have developed a Python3 code, *bullviso*,¹⁹ which interfaces with RDKit²⁰ to generate the Cartesian coordinates of substituted BVs. It is publicly available under the GNU Public License (GPLv3) on GitLab. The *bullviso* code generates exhaustively all possible isomer barcodes for a substituted BV and filters out the non-unique isomer barcodes according to the protocol outlined by Bode.¹⁷ It also outputs the connectivity between isomers, which can be used to generate interconversion network diagrams.²¹ Cartesian coordinates sampling the constitutional isomers of the substituted BV are generated by grafting given substituents (supplied as SMILES strings) onto a BV to produce a unique structure corresponding to each isomer barcode. The *bullviso* code generates up to N configurational and conformational isomers according to the experimental-torsion distance geometry with 'basic knowledge' embedding approach (ETKDGv3)²² implemented in RDKit. These configurational and conformational isomers are then (pre-)optimized using the Universal Force Field²³ and the M lowest-energy isomers are outputted. Cartesian coordinates can be written in xyz file format or, alternatively, as pre-prepared inputs for computational chemistry packages, *e.g.*, Gaussian²⁴ or Orca,²⁵ to enable subsequent optimization and analysis at higher levels of theory.

Principal moments of inertia analysis

The PMI analysis developed by Schwartz¹⁵ has been used as a straightforward and quantitative method to assess the shape diversity of potential pharmaceutical building blocks.^{12,13} Typical PMI plots are constructed (i) for the lowest energy conformer of any given compound to compare either the inherent shapes of an array of molecules²⁶ or (ii) for a range of conformers to gain insight into conformational diversity of a limited number of molecules.¹⁴ We selected di-, tri-, and tetrasubstituted BVs (Fig. 3a) as targets to examine shape diversity present in dynamic BV networks arising from rapid constitutional isomerism. To enable us to focus our analysis on the dynamic shape that is inherent to the functionalized BV scaffold itself rather than any potential conformational processes in the



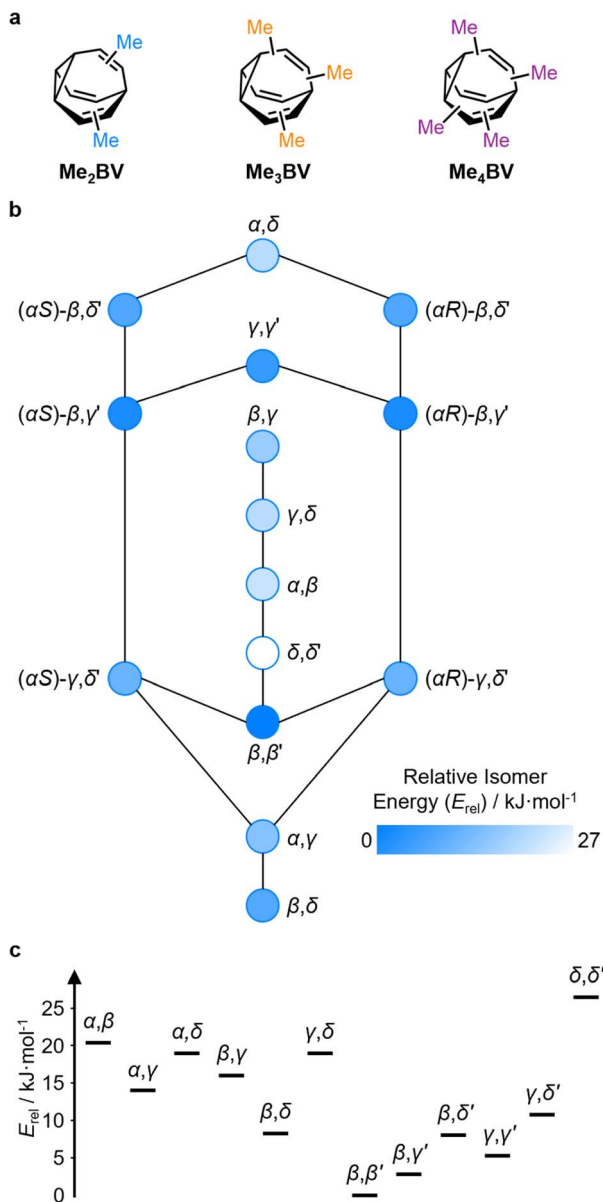


Fig. 3 (a) Structural formulae of methyl-substituted BVs. (b) The population-weighted isomer interconversion network calculated for Me_2BV (PBE0-D3/def2-SV(P)). The diagram has a mirror plane with achiral isomers down the middle and enantiomeric pairs of chiral structures on either side. Chiral structures are labelled with an R/S descriptor according to the stereogenic α position. (c) A graph of the relative energies of Me_2BV isomers. Pairs of enantiomers are iso-energetic, so are represented just once.

attached substituents, we chose to investigate the methyl-substituted derivatives, *i.e.*, dimethyl- (Me_2BV), trimethyl- (Me_3BV), and tetramethylbullvalene (Me_4BV).

We first generated all possible isomers of the methyl-substituted BVs using *bullviso*, then optimized their geometries by performing density functional theory (DFT) calculations. The PBE0 (ref. 27) functional with Grimme's D3 dispersion correction²⁸ and the def2-SV(P)²⁹ basis set were deemed suitable for modelling the energetics of BV systems. Using this level of theory, we constructed isomer

interconversion networks and predicted relative isomer populations for Me_2BV (Fig. 3), Me_3BV (Fig. S1†), and Me_4BV (Fig. S2†), which have 15, 42, and 72 unique isomers, respectively.²¹ For each of these isomers, we calculated the three principal moments of inertia (I_1 , I_2 , and I_3 in ascending order) using a KNIME³⁰ Vernalis³¹ chemoinformatic protocol, then used these values to calculate normal PMI ratios, NPR1 (I_1/I_3) and NPR2 (I_2/I_3). The resulting PMI plots (Fig. 4a–g) follow the standard layout (Fig. 4) where the vertices are defined by NPR values associated with rod-like shape [NPR1 = 0, NPR2 = 1], disc-like shape [NPR1 = 0.5, NPR2 = 0.5], and spherical shape [NPR1 = 1, NPR2 = 1]. The diagrams are also labelled with representative structures for each vertex, *i.e.*, butadiyne (rod-like), benzene (disc-like), and adamantane (sphere-like). To guide the eye, parallel lines on the PMI plot correspond to increments of 0.1 in $\sum\text{NPR}$ values ($\sum\text{NPR} = \text{NPR1} + \text{NPR2}$) between the limits of 1.0 and 2.0. The points that lie furthest from the rod–disc axis, *i.e.*, toward the top right of the diagram, are associated with greater sphericity.

The PMI plot of Me_2BV (Fig. 4a) shows the structural diversity in the population of constitutional isomers, which fall in the range $1.44 \leq \sum\text{NPR} \leq 1.57$. For ease of reference, isomers in Fig. 4a are labelled using the α - δ locants, however, the full isomer barcodes are also given in the ESI (Tables S2–S4).† Darker colored points indicate overlap of enantiomers on the diagram as they give identical PMI coordinates, *e.g.*, for $(\pm)\text{-}\beta,\gamma'$ - Me_2BV .

As the unique isomers of a substituted BV are non-degenerate, they are present in varying concentrations at equilibrium. To visualize how this property influences which molecular shapes are most prevalent, we made PMI plots with the data points scaled by the Boltzmann distribution at 298 K. A population-weighted PMI plot (Fig. 4d) shows that the shape-shifting network of Me_2BV consists predominantly ($\sim 93\%$) of the four lowest-energy isomers, which include two achiral isomers, β,β' - Me_2BV and γ,γ' - Me_2BV , and the enantiomeric pair of $(\pm)\text{-}\beta,\gamma'$ - Me_2BV . The β,δ - Me_2BV , $(\pm)\text{-}\beta,\delta'$ - Me_2BV , and $(\pm)\text{-}\gamma,\delta'$ - Me_2BV isomers are also present in ~ 0.5 – 2% each (Table S2†), leaving $\sim 0.3\%$ of the remaining six isomers combined.

The introduction of more substituents to the BV scaffold (*i.e.*, in Me_3BV and Me_4BV) increases structural variety and overall sphericity (Fig. 4b and c). Several structures extend beyond $\sum\text{NPR} > 1.6$, reaching maxima of 1.89 and 1.90 for Me_3BV and Me_4BV , respectively. In both cases, the PMI distributions are markedly broader than that of Me_2BV , spanning $1.3 \leq \sum\text{NPR} \leq 1.9$, which reflects the change in shape that can occur when several substituents are located close to one another around the BV scaffold (giving rod-like shape) or are spread around the BV evenly to maintain sphericity. Like Me_2BV , a subset of the Me_3BV and Me_4BV constitutional isomers are most prevalent in the network at equilibrium. The 14 most stable isomers of Me_3BV are within $\sim 15 \text{ kJ mol}^{-1}$ of one another, so they are each present in amounts ranging from 0.1% to 60% (Table S3†). Similarly, there are 12 isomers of Me_4BV present in proportions of 0.1% to 52%.^{||}³² The energetically preferred isomers are those with most of their substituents at β and γ positions, and in which substituents do not neighbor one another directly.



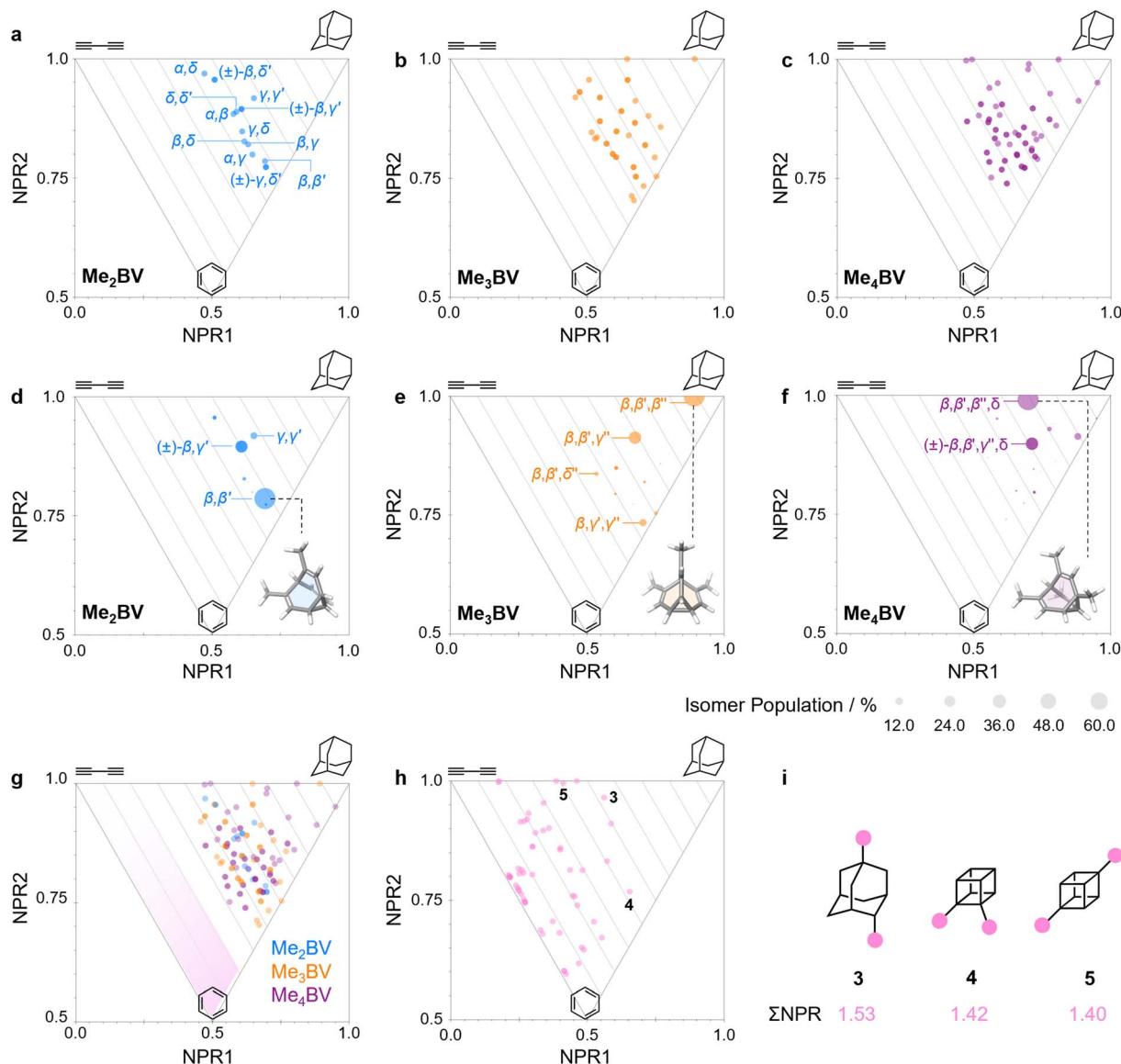


Fig. 4 (a–c) PMI plots for the shapeshifting networks of (a) Me_2BV , (b) Me_3BV , and (c) Me_4BV . Substituent positional labels are given for Me_2BV . For clarity, these labels are not shown on the plots for Me_3BV and Me_4BV . See Tables S6 and S7† for labelled data. (d–e) Population-weighted PMI plots for the shapeshifting networks of (d) Me_2BV , (e) Me_3BV , and (f) Me_4BV where the data points are scaled by calculated Boltzmann distributions at 298 K (PBE0–D3/def2–SV(P)). The modelled structure of the lowest-energy isomer for each BV is shown inset. (g) An overlay of the PMI plots of Me_2BV (blue), Me_3BV (orange), and Me_4BV (purple) showing that none of the isomers have ΣNPR values close to the rod–disc axis. (h) A PMI plot for common ring systems. See Table S8† for compound identities. (i) Structural formulae and ΣNPR values for 1,4-dimethyladamantane (3), 1,2-dimethylcubane (4), and 1,4-dimethylcubane (5).

Therefore, the population-weighted PMI plots (Fig. 4e and f) show that the most spherical isomers are present in higher populations.

To contextualize the PMI data of the functionalized BVs, it is useful to compare the Me_nBV isomers (Fig. 4a) to a small representative library of rings prominent in pharmaceuticals (Fig. S3†).³³ PMI analysis of the dimethyl derivatives of these pharmaceutical building blocks (Fig. 4h) shows the relative lack of 3D diversity in most currently used ring systems.³² The majority of compounds are close to the rod–disc axis ($\Sigma\text{NPR} \leq 1.3$) with the notable exception of a few structures such as substituted adamantanes and cubanes 3–5 (Fig. 4i), which are

increasingly popular as more spherical replacements for flat ring systems.³⁴ The PMI ratios of the Me_2BV isomers ($1.44 \leq \Sigma\text{NPR} \leq 1.57$) and 3–5 ($1.40 \leq \Sigma\text{NPR} \leq 1.53$) are nearly identical, suggesting that BVs could be similarly useful as 3D scaffolds in medicinal chemistry. BVs have the additional, unique property of spontaneously sampling different structures through their Cope rearrangements and are now readily accessible by short synthetic routes.^{1c,2}

Exit vector analysis

EV plots are used to analyze the relationship between two substituents attached to a central scaffold.¹⁶ They are useful in



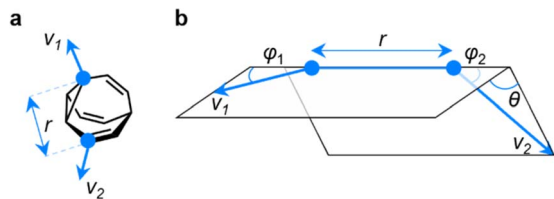


Fig. 5 (a) The vectors v_1 and v_2 for two substituent attachment points on a BV (shown for $(\alpha S)\text{-}\gamma,\delta'\text{-Me}_2\text{BV}$) which are defined by (b) the geometric parameters r , ϕ_1 , ϕ_2 , and θ .

FBDD and bioisostere studies to show the geometries that are available when elaborating a structure outward starting from different ring systems. The relative orientations (Fig. 5a) of

bonds emanating from the scaffold are defined as the exit vectors, v_1 and v_2 . They are described (Fig. 5b) by four geometric parameters; the distance between the functionalized carbon atoms of the scaffold (r), the dihedral angle of the vectors (θ), and the plane angles of each vector (ϕ_1 and ϕ_2).¹⁶ Like PMI plots, the EV plot for a single molecule possesses only one data point if a compound is conformationally rigid, whereas multiple points are plotted to show the effects of conformational flexibility or to compare multiple molecules with different covalent structures on a single diagram.

We constructed EV plots (Fig. 6) for the C–Me bonds of the methyl-substituted BVs to quantify the stereodynamics of the BV scaffold, *i.e.*, the relative orientations and spacings of its substituents. Typically, EV plots span θ values of 0° to 180° ,

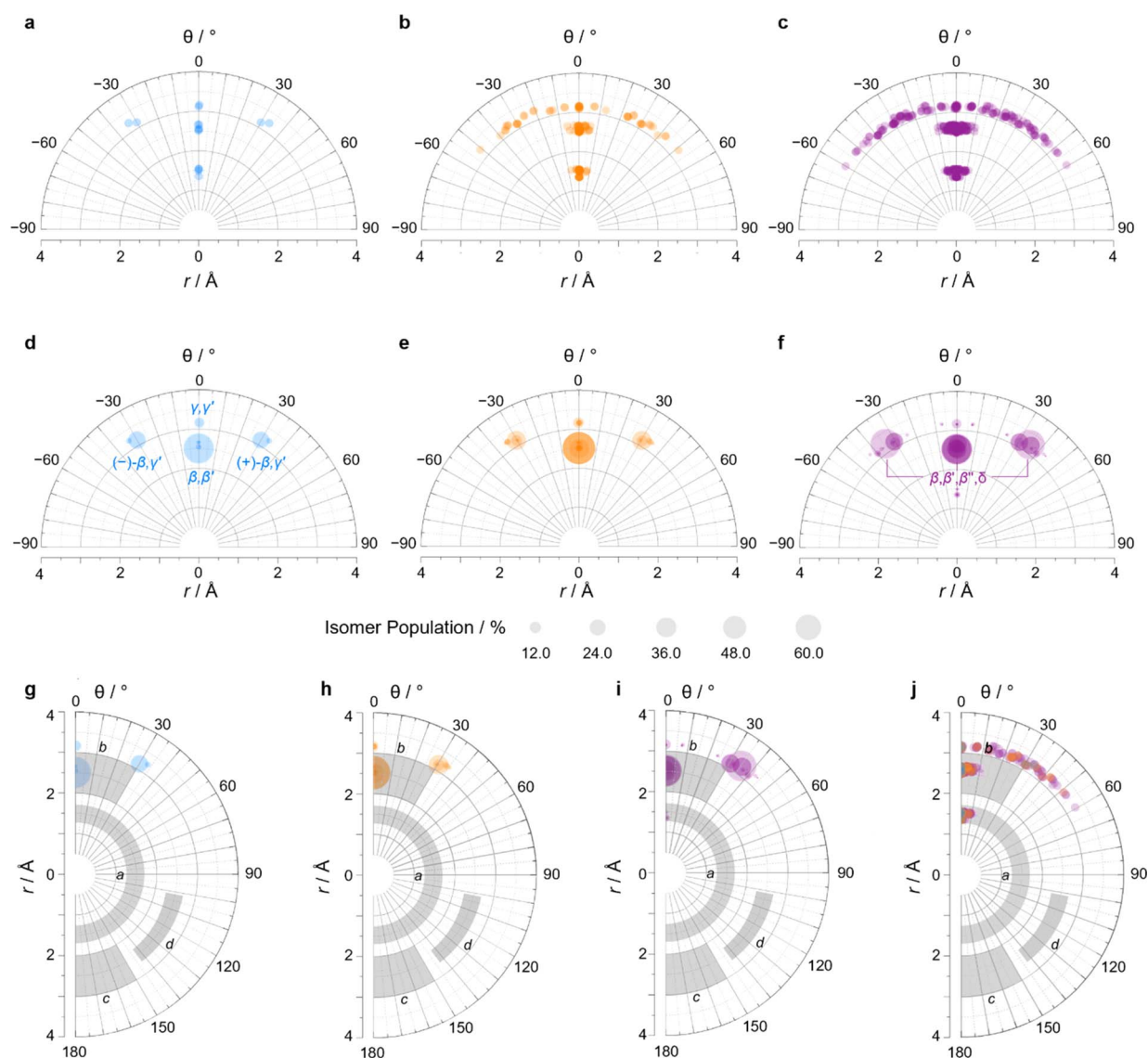


Fig. 6 (a–c) Distance–dihedral angle EV plots and (d–i) Boltzmann population-weighted distance–dihedral angle EV plots (298 K, PBE0-D3/def2-SV(P)) for the isomers of (a, d and g) Me_2BV , (b, e and h) Me_3BV , and (c, f and i) Me_4BV . (j) Overlaid distance versus dihedral angle EV plot of all three methyl-substituted BVs. (g–j) Plots include characteristic areas of EV plots in grey that correspond to those found in disubstituted cycloalkanes,¹⁶ a = *cis*-1,2-disubstituted cyclopropanes, b = *cis*-1,3-disubstituted aliphatic rings and *cis*-1,4-disubstituted cyclohexanes, c = *trans*-1,4-disubstituted cyclohexanes, d = *trans*-1,3-disubstituted cyclopentanes and cyclohexanes.



where all dihedral angles are defined as being positive. For BV isomer networks, however, it is also beneficial to extend the EV plots to include negative values of θ , allowing pairs of enantiomers that are present in the dynamic equilibrium to be shown on the same diagram (e.g., (\pm) - β , γ' - Me_2BV , Fig. 6d), as well as revealing enantiotopic relationships between substituents at equivalent positions on different arms (e.g., the β' and β'' positions of β , β' , β'' , δ - Me_4BV , Fig. 6f). As the cage-like structure of BV imposes dihedral angles between -60° and $+60^\circ$, plotting θ from -90° to $+90^\circ$ (Fig. 6a–f) gives an informative representation of the data.

An EV plot of the Me_2BV isomers reveals C–Me dihedral angles clustered in two regions of either $\theta \sim 0^\circ$ or $\theta \sim \pm 30^\circ$. The points in the former region span distances of $r \sim 1.3$ – 3.1 Å (Fig. 6a). The majority of coplanar EVs ($\theta \sim 0^\circ$) arise from isomers that are either functionalized (i) at two different positions on the same arm of the BV or (ii) at the same type of position on BV on different arms, e.g., β , β' - Me_2BV , accounting for nine of the 15 possible substitution patterns. The four points at $\theta \sim \pm 30^\circ$ correspond to two of the three enantiomer pairs of the isomers with differently substituted arms ((\pm) - β , γ' - Me_2BV and (\pm) - β , δ' - Me_2BV). The final enantiomer pair, (\pm) - γ , δ' - Me_2BV , has coplanar EVs of $\theta = \pm 0.5^\circ$. Boltzmann population-weighted EV analysis of Me_2BV (Fig. 6d) reveals that the most populated isomers β , β' - Me_2BV ($r = 2.5$ Å, $\theta = 0^\circ$, $p = 53\%$) and (\pm) - β , γ' - Me_2BV ($r = 3.1$ Å, $\theta = 30^\circ$, $p = 17\%$ for each enantiomer) exhibit substantial changes in the dihedral angles between the C–Me EVs.

Standard EV plots of Me_2BV , Me_3BV , and Me_4BV spanning θ values of 0° to 180° (Fig. 6g–j) are also shown to aid comparison to the four regions, a–d, determined by Grygorenko *et al.* that are characteristic of EVs found in common disubstituted cycloalkanes.¹⁶ Note that these four regions were determined by plotting the EVs found for ~ 2900 cycloalkanes in the Cambridge Structural Database, so they represent the span of angles that are obtained using a variety of functional groups. The

majority of data points for the BV isomers fall within the a or b region. EVs in region a are characteristic of *cis*-1,2-disubstituted cyclopropyl compounds while region b is associated with *cis*-1,3-disubstituted and *cis*-1,4-disubstituted aliphatic rings.¹⁶ The presence of EVs in these regions for BV is expected, therefore, as the structure of BV contains these motifs.

The diversity of EVs arising from the shapeshifting isomerization becomes even more apparent when considering the plane angles subtended by the C–Me EVs, φ_1 and φ_2 (Fig. 5), of the Me_2BV isomers. There is a spread of possible plane angles spanning from 15° to 60° found in the most energetically accessible isomers (Fig. 7), extending to 67° for higher-energy isomers (Fig. S4†). Therefore, sequential Cope rearrangements between isomers significantly alter the relative orientations of substituents in space, granting BV its unique stereodynamic properties.

As the cage-like geometry of the BV scaffold is well defined and insensitive to the addition of more substituents, Me_3BV and Me_4BV would be expected to exhibit similar EVs between pairs of their C–Me bonds as those found for Me_2BV . This generalization applies when considering the plane angles (Fig. S7–S14†) and distances between substituents. It is also the case for the dihedral angles, but only when considering the most energetically favorable isomers (Fig. 6e and f).

A wider spread of dihedral angles (Fig. 6b and c) that extends beyond region b to angles between -60° and $+60^\circ$ is apparent when the complete isomer networks of Me_3BV and Me_4BV are taken into account (Fig. 6j), including isomers that lie at higher energy. Significant deviations from the 0° and $\pm 30^\circ$ dihedral angles arise to minimize unfavorable steric interactions when substituents are close to one another in space. For the di-, tri-, and tetrasubstituted BVs investigated here, the isomers that bear methyl groups at neighboring positions are relatively high in energy, so are not very prevalent. The most significant (albeit still low) populations of such isomers are present for Me_4BV because of the increased likelihood of substituents being close to one another in the tetrasubstituted system, such as the $\sim 0.1\%$ of β , β' , γ , δ' - Me_4BV , which has $\theta = 51^\circ$ between its β and δ' methyl groups. But more highly substituted BVs, or those bearing bulkier substituents, will likely have larger isomer populations with varied dihedral angles.

Of course, shape differentiation between the orientations of functional groups attached to BV diminishes if conformationally flexible linking groups are used, so elaborating BVs with short rigid groups^{2c,8b,35} or fused rings³⁶ may be advantageous. Each of the molecular shapes described by the PMI (Fig. 4a–c) and EV plots (Fig. 6a–c and S4–S14†) correspond to accessible states for the methyl-substituted BVs, demonstrating that a single BV derivative can cover a significant region of chemical space on its own. Heterosubstituted systems with varied functional groups possess still greater structural diversity. Although realizing the full extent of this shape diversity in some BVs may require higher-energy isomers to be invoked, it is important to note that, in the contexts of drug discovery and materials chemistry, noncovalent bonding interactions with biomolecular targets, confinement effects, or forces imposed by the surrounding medium may compensate for the moderately low

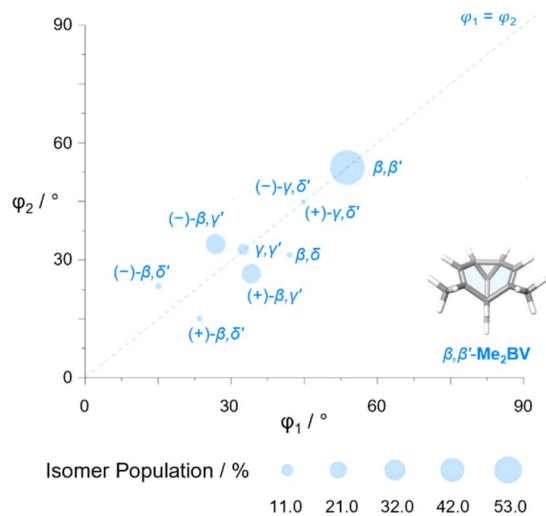


Fig. 7 The plane angles subtended by C–Me EVs in the nine lowest energy isomers of Me_2BV . Data points are scaled according to the Boltzmann population at 298 K (PBE0-D3/def2-SV(P)).



energy differences between isomers to amplify certain BV structures within the shapeshifting network.^{2d,4,6,37} Therefore, isomers that have otherwise low populations should not be dismissed, as they may become more significant components of the network in the right environment.

Conclusions

In summary, the appealing complexity of dynamic BV networks is built upon four types of positional exchange between different sites around the BV scaffold. Isomers in these networks tend to position functional groups with dihedral angles of either 0° or 30° and at a range of plane angles between 15° and 60°, originating from vertices of the BV that are spaced apart by 1.3–3.1 Å. The BV scaffold itself has quasi-spherical overall structure. Functionalized derivatives generally retain this characteristic, particularly because the most energetically favored isomers that emerge are typically those that space substituents out from one another around the scaffold. PMI analysis shows that they have a similar degree of sphericity as adamantane and cubane building blocks, while they orient functional groups at vectors that are characteristic of *cis*-disubstituted cycloalkanes. Greater shape diversity emerges with the inclusion of more substituents, not just because it gives rise to greater numbers of unique isomers, but also because interactions between neighboring groups reduces the predisposition toward them being coplanar with one another, giving more varied EVs. Currently, the most accessible syntheses of BVs^{4c,2a-c} produce structures with two or three functional groups attached, implying it remains advantageous to develop new synthetic approaches that give efficient access to higher-order multifunctional BVs. The analysis described here, which is accelerated using *bullviso*, can be exploited to prescreen computationally the diversity and accessibility of molecular shapes in complex BV libraries.

Data availability

Coordinates of optimized geometries, PMIs and EVs are available in the ESI.† The Python code for *bullviso* is publicly available under the GNU Public License (GPLv3) on GitLab.¹⁹

Author contributions

Conceptualization: RAI, WM, PRM. Methodology: RAI, WM, CR. Software: CR. Writing: RAI, WM, MTG, CR, PRM. Funding acquisition: PRM.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We thank James Firth and Prof Peter O'Brien for useful discussions. R. A. I. and P. R. M. acknowledge a Leverhulme Trust Research Project Grant (RPG-2020-218). W. M., M. G. and

P. R. M. (EP/T518001/1, EP/V047817/2, EP/V040049/2) thank the EPSRC for funding. We acknowledge the University of York High Performance Computing service, Viking.

Notes and references

§ The use of Greek letter locants to label relative positions and prime symbols to label identical rings in multiple ring systems follows IUPAC conventions.¹⁸

¶ A Fortran code to generate input geometries of BV isomers is mentioned in ref. 2a but it has not been made publicly available.

|| A consequence of there being more possible isomers in the networks of highly substituted BVs is that, statistically, any individual isomer is expected to be present in lower concentrations. However, certain BV systems, *e.g.*, hexasubstituted BVs, are biased toward a relatively small number of isomers, which act as energetic sinks that also slow the rate of exchange.³²

- (a) W. von Eggers Doering and W. R. Roth, *Tetrahedron*, 1963, **19**, 715–737; (b) G. Schröder, J. F. M. Oth and R. Merényi, *Angew. Chem. Int. Ed. Engl.*, 1965, **4**, 752–761; (c) S. Ferrer and A. M. Echavarren, *Synthesis*, 2019, **51**, 1037–1048; (d) A. N. Bismillah, B. M. Chapin, B. A. Hussein and P. R. McGonigal, *Chem. Sci.*, 2020, **11**, 324–332; (e) P. K. Saha, T. Tran Ngoc, P. R. McGonigal and J. F. Teichert, *Nat. Synth.*, 2024, **3**, 684–697.
- (a) O. Yahiaoui, L. F. Pašteka, B. Judeel and T. Fallon, *Angew. Chem., Int. Ed.*, 2018, **57**, 2570–2574; (b) O. Yahiaoui, L. F. Pašteka, C. J. Blake, C. G. Newton and T. Fallon, *Org. Lett.*, 2019, **21**, 9574–9578; (c) H. D. Patel, T.-H. Tran, C. J. Sumby, L. F. Pašteka and T. Fallon, *J. Am. Chem. Soc.*, 2020, **142**, 3680–3685; (d) J. F. Teichert, D. Mazunin and J. W. Bode, *J. Am. Chem. Soc.*, 2013, **135**, 11314–11321.
- A. Sanchez, A. Gurajapu, W. Guo, W.-Y. Kong, C. J. Laconsay, N. S. Settineri, D. J. Tantillo and T. J. Maimone, *J. Am. Chem. Soc.*, 2023, **145**, 13452–13461.
- (a) A. R. Lippert, V. L. Keleshian and J. W. Bode, *Org. Biomol. Chem.*, 2009, **7**, 1529–1532; (b) A. R. Lippert, A. Naganawa, V. L. Keleshian and J. W. Bode, *J. Am. Chem. Soc.*, 2010, **132**, 15790–15799.
- C. Dohmen, H. Ihmels and T. Paululat, *Eur. J. Org. Chem.*, 2022, **2022**, e202201172.
- (a) A. P. Bivé, H. D. Patel, J. R. Price, W. M. Bloch and T. Fallon, *Angew. Chem., Int. Ed.*, 2022, **61**, e202115468; (b) C. Dohmen, T. Paululat and H. Ihmels, *Chem.–Eur. J.*, 2024, **30**, e202304311.
- J. R. Reimers, T. Li, A. P. Bivé, L. Yang, A. C. Aragonès, T. Fallon, D. S. Kosov and N. Darwish, *Nat. Commun.*, 2023, **14**, 6089.
- (a) D. J. Tantillo and R. Hoffmann, *Acc. Chem. Res.*, 2006, **39**, 477–486; (b) M. N. Pomfret, P. B. Sun, Z. Huang, A. C. Freund, T. Miyoshi and M. R. Golder, *Angew. Chem., Int. Ed.*, 2023, **62**, e202301695; (c) P. B. Sun, M. N. Pomfret, M. J. Elardo, A. Suresh, Á. Rentería-Gómez, R. F. Lalisce, S. Keating, C. Chen, S. L. Hilburg, P. Chakma, Y. Wu, R. C. Bell, S. J. Rowan, O. Gutierrez and M. R. Golder, *J. Am. Chem. Soc.*, 2024, **146**, 19229–19238.
- A. Ottonello, J. A. Wyllie, O. Yahiaoui, S. Sun, R. A. Koelln, J. A. Homer, R. M. Johnson, E. Murray, P. Williams, J. R. Bolla, C. V. Robinson, T. Fallon, T. P. Soares da Costa



- and J. E. Moses, *Proc. Natl. Acad. Sci. U.S.A.*, 2023, **120**, e2208737120.
- 10 (a) A. E. Baumann, D. A. Burns, B. Liu and V. S. Thoi, *Commun. Chem.*, 2019, **2**, 86; (b) Z. Ji, H. Wang, S. Canossa, S. Wuttke and O. M. Yaghi, *Adv. Funct. Mater.*, 2020, **30**, 2000238.
- 11 (a) R.-B. Lin and B. Chen, *Chem*, 2022, **8**, 2114–2135; (b) K. Geng, T. He, R. Liu, S. Dalapati, K. T. Tan, Z. Li, S. Tao, Y. Gong, Q. Jiang and D. Jiang, *Chem. Rev.*, 2020, **120**, 8814–8933.
- 12 (a) D. A. Erlanson, S. W. Fesik, R. E. Hubbard, W. Jahnke and H. Jhoti, *Nat. Rev. Drug Discovery*, 2016, **15**, 605–619; (b) B. Over, S. Wetzler, C. Grütter, Y. Nakai, S. Renner, D. Rauh and H. Waldmann, *Nat. Chem.*, 2013, **5**, 21–28.
- 13 (a) A. W. Hung, A. Ramek, Y. Wang, T. Kaya, J. A. Wilson, P. A. Clemons and D. W. Young, *Proc. Natl. Acad. Sci. U.S.A.*, 2011, **108**, 6799–6804; (b) W. R. J. D. Galloway, A. Isidro-Llobet and D. R. Spring, *Nat. Commun.*, 2010, **1**, 80; (c) A. D. Morley, A. Pugliese, K. Birchall, J. Bower, P. Brennan, N. Brown, T. Chapman, M. Drysdale, I. H. Gilbert, S. Hoelder, A. Jordan, S. V. Ley, A. Merritt, D. Miller, M. E. Swarbrick and P. G. Wyatt, *Drug Discov. Today*, 2013, **18**, 1221–1227.
- 14 T. D. Downes, S. P. Jones, H. F. Klein, M. C. Wheldon, M. Atobe, P. S. Bond, J. D. Firth, N. S. Chan, L. Waddelove, R. E. Hubbard, D. C. Blakemore, C. De Fusco, S. D. Roughley, L. R. Vidler, M. A. Whatton, A. J. A. Woolford, G. L. Wrigley and P. O'Brien, *Chem.-Eur. J.*, 2020, **26**, 8969–8975.
- 15 W. H. B. Sauer and M. K. Schwarz, *J. Chem. Inf. Comput. Sci.*, 2003, **43**, 987–1003.
- 16 (a) O. O. Grygorenko, D. Demenko, D. M. Volochnyuk and I. V. Komarov, *New J. Chem.*, 2018, **42**, 8355–8365; (b) O. O. Grygorenko, P. Babenko, D. M. Volochnyuk, O. Raievskiy and I. V. Komarov, *RSC Adv.*, 2016, **6**, 17595–17605.
- 17 M. He and J. W. Bode, *Org. Biomol. Chem.*, 2013, **11**, 1306–1317.
- 18 H. A. Favre and W. H. Powell, *Nomenclature of Organic Chemistry: IUPAC Recommendations and Preferred Names 2013, IUPAC Blue book*, RSC Publishing, 2014.
- 19 bullviso, 2024, <https://gitlab.com/conorrankine/bullviso>.
- 20 (a) RDKit: Open-Source Cheminformatics, <https://rdkit.org>; (b) RDKit, 2023, <https://github.com/rdkit/rdkit>.
- 21 B. M. Gimarc and A. R. Brant, *J. Chem. Inf. Comput. Sci.*, 1994, **34**, 1167–1173.
- 22 (a) S. Riniker and G. A. Landrum, *J. Chem. Inf. Model.*, 2015, **55**, 2562–2574; (b) S. Wang, J. Witek, G. A. Landrum and S. Riniker, *J. Chem. Inf. Model.*, 2020, **60**, 2044–2058.
- 23 A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard and W. M. Skiff, *J. Am. Chem. Soc.*, 1992, **114**, 10024–10035.
- 24 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, Gaussian, Inc., Wallingford CT*, 2016.
- 25 (a) F. Neese, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2012, **2**, 73–78; (b) F. Neese, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2018, **8**, e1327; (c) F. Neese, F. Wennmohs, U. Becker and C. Riplinger, *J. Chem. Phys.*, 2020, **152**, 224108.
- 26 (a) S. P. Jones, J. D. Firth, M. C. Wheldon, M. Atobe, R. E. Hubbard, D. C. Blakemore, C. De Fusco, S. C. C. Lucas, S. D. Roughley, L. R. Vidler, M. A. Whatton, A. J. A. Woolford, G. L. Wrigley and P. O'Brien, *RSC Med. Chem.*, 2022, **13**, 1614–1620; (b) P. Garner, P. B. Cox, U. Rathnayake, N. Holloran and P. Erdman, *ACS Med. Chem. Lett.*, 2019, **10**, 811–815.
- 27 (a) C. Adamo and V. Barone, *J. Chem. Phys.*, 1999, **110**, 6158–6170; (b) J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1997, **78**, 1396; (c) J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.
- 28 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
- 29 F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297.
- 30 M. R. Berthold, N. Cebron, F. Dill, T. R. Gabriel, T. Kötter, T. Meinl, P. Ohl, C. Sieb, K. Thiel and B. Wiswedel, *KNIME: The Konstanz Information Miner*, Springer, 2007.
- 31 S. D. Roughley, *Curr. Med. Chem.*, 2018, **27**, 6495–6522.
- 32 K. Rebsamen, H. Röttele and G. Schröder, *Chem. Ber.*, 1993, **126**, 1429–1433.
- 33 (a) R. D. Taylor, M. MacCoss and A. D. G. Lawson, *J. Med. Chem.*, 2014, **57**, 5845–5859; (b) J. Shearer, J. L. Castro, A. D. G. Lawson, M. MacCoss and R. D. Taylor, *J. Med. Chem.*, 2022, **65**, 8699–8712.
- 34 (a) B. A. Chalmers, H. Xing, S. Houston, C. Clark, S. Ghassabian, A. Kuo, B. Cao, A. Reitsma, C. E. P. Murray, J. E. Stok, G. M. Boyle, C. J. Pierce, S. W. Littler, D. A. Winkler, P. V. Bernhardt, C. Pasay, J. J. De Voss, J. McCarthy, P. G. Parsons, G. H. Walter, M. T. Smith, H. M. Cooper, S. K. Nilsson, J. Tsanaktisidis, G. P. Savage and C. M. Williams, *Angew. Chem., Int. Ed.*, 2016, **55**, 3580–3585; (b) L. Wanka, K. Iqbal and P. R. Schreiner, *Chem. Rev.*, 2013, **113**, 3516–3604; (c) T. A. Reekie, C. M. Williams, L. M. Rendina and M. Kassiou, *J. Med. Chem.*, 2018, **62**, 1078–1095; (d) K. C. Nicolaou, J. Yin, D. Mandal, R. D. Erande, P. Klahn, M. Jin, M. Aujay, J. Sandoval, J. Gavrilyuk and D. Vourloumis, *J. Am. Chem. Soc.*, 2016, **138**, 1698–1708; (e) E. G. Tse, S. D. Houston, C. M. Williams, G. Paul Savage, L. M. Rendina,



- I. Hallyburton, M. Anderson, R. Sharma, G. S. Walker, R. Scott Obach and M. H. Todd, *J. Med. Chem.*, 2020, **63**, 11585–11601; (f) P. K. Mykhailiuk, *Org. Biomol. Chem.*, 2019, **17**, 2839–2849; (g) M. A. M. Subbaiah and N. A. Meanwell, *J. Med. Chem.*, 2021, **64**, 14046–14128; (h) E. T. Warda, M. B. El-Ashmawy, E.-S. E. Habib, M. S. M. Abdelbaky, S. Garcia-Granda, S. Thamptharon and A. A. El-Emam, *Sci. Rep.*, 2022, **12**, 21058.
- 35 A. N. Bismillah, J. Sturala, B. M. Chapin, D. S. Yufit, P. Hodgkinson and P. R. McGonigal, *Chem. Sci.*, 2018, **9**, 8631–8636.
- 36 G. Schröder and W. Witt, *Angew. Chem. Int. Ed. Engl.*, 1979, **18**, 311–312.
- 37 (a) A. N. Bismillah, T. G. Johnson, B. A. Hussein, A. T. Turley, P. K. Saha, H. C. Wong, J. A. Aguilar, D. S. Yufit and P. R. McGonigal, *Nat. Chem.*, 2023, **15**, 615–624; (b) B. A. Hussein, W. Maturi, M. K. Rylands, A. N. Bismillah, Y. Wen, J. A. Aguilar, R. Ayub, C. D. Rankine and P. R. McGonigal, *Chem. Sci.*, 2024, DOI: [10.1039/D4SC03699A](https://doi.org/10.1039/D4SC03699A).

