PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

How does symmetry impact the flexibility of proteins?

Bernd Schulze, Adnan Sljoka and Walter Whiteley

Phil. Trans. R. Soc. A 2014 **372**, 20120041, published 30 December 2013

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A

rsta.royalsocietypublishing.org

Research

Cite this article: Schulze B, Sljoka A, Whiteley W. 2014 How does symmetry impact the flexibility of proteins? *Phil. Trans. R. Soc. A* **372**: 20120041.
http://dx.doi.org/10.1098/rsta.2012.0041

One contribution of 14 to a Theo Murphy Meeting Issue 'Rigidity of periodic and symmetric structures in nature and engineering'.

Subject Areas:
biomathematics, computational mathematics, graph theory, group theory, geometry

Keywords:
rigidity of frameworks, flexibility, symmetry, proteins, allostery, pebble game algorithms

Author for correspondence:
Bernd Schulze
e-mail: bschulze@mathstat.yorku.ca

How does symmetry impact the flexibility of proteins?

Bernd Schulze, Adnan Sljoka and Walter Whiteley

Department of Mathematics and Statistics, York University, Toronto, Ontario, Canada M3J 1P3

It is well known that (i) the flexibility and rigidity of proteins are central to their function, (ii) a number of oligomers with several copies of individual protein chains assemble with symmetry in the native state and (iii) added symmetry sometimes leads to added flexibility in structures. We observe that the most common symmetry classes of protein oligomers are also the symmetry classes that lead to increased flexibility in certain three-dimensional structures—and investigate the possible significance of this coincidence. This builds on the well-developed theory of generic rigidity of body–bar frameworks, which permits an analysis of the rigidity and flexibility of molecular structures such as proteins via fast combinatorial algorithms. In particular, we outline some very simple counting rules and possible algorithmic extensions that allow us to predict continuous symmetry-preserving motions in body–bar frameworks that possess non-trivial point-group symmetry. For simplicity, we focus on dimers, which typically assemble with twofold rotational axes, and often have allosteric function that requires motions to link distant sites on the two protein chains.

1. Introduction

A review of a table of symmetric protein quaternary structures formed by oligomeric complexes (two or more copies of the same protein bound to a larger molecule) [1] highlights that the most common symmetry groups found in such structures are the group $\mathcal{C}_2$, which describes half-turn symmetry, and then the groups $\mathcal{D}_2$ and $\mathcal{D}_3$, where $\mathcal{D}_2$ is the dihedral symmetry group of order four generated by two half-turns about perpendicular axes and $\mathcal{D}_3$ is the dihedral symmetry group of order six generated by a threefold and a twofold rotation about perpendicular axes (see table 1 and

Royal Society Publishing

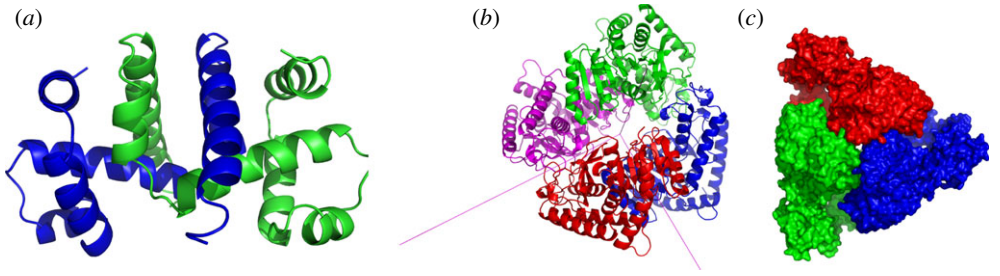

Figure 1. Proteins with $\mathcal{C}_2$, $\mathcal{D}_2$ and $\mathcal{D}_3$ symmetry. (*a*) The dimer tryptophan repressor—shown in cartoon representation—has $\mathcal{C}_2$ symmetry, where the two colours represent the two chains. (*b*) Bacterial L-lactate dehydrogenase (PDB ID: 1lth) has $\mathcal{D}_2$ symmetry. (*c*) *N*-phosphonacetyl-L-aspartate (PDB ID: 8atc)—shown in surface representation—has $\mathcal{D}_3$ symmetry, where colours distinguish separate chains. This complex is formed with two types of chains, six copies of each, which corresponds to the 6 + 6 notation for this family of proteins in table 1. Each colour contains two dimers with two different chains in each dimer (i.e. in each colour, one dimer is built of chains A and C, and the other dimer is built of chains B and D). These structures were generated with Pymol (http://pymol.sourceforge.net). (Online version in colour.)

Table 1. A selection of proteins and their point groups. This table is an excerpt from the more complete table given in [1].

protein	no. subunits	point group
alcohol dehydrogenase	2	$\mathcal{C}_2$
immunoglobulin	4	$\mathcal{C}_2$
s-malate dehydrogenase	2	$\mathcal{C}_2$
superoxide dismutase	2	$\mathcal{C}_2$
triose phosphate isomerase	2	$\mathcal{C}_2$
phosphorylase	2	$\mathcal{C}_2$
alkaline phosphatase	2	$\mathcal{C}_2$
concanavalin A	4	$\mathcal{D}_2$
lactate dehydrogenase	4	$\mathcal{D}_2$
prealbumin	4	$\mathcal{D}_2$
pyruvate kinase	4	$\mathcal{D}_2$
insulin	6	$\mathcal{D}_3$
asparate transcarbamoylase	6 + 6	$\mathcal{D}_3$

figure 1). Encountering such a pattern, it is natural to ask why these groups are more abundant than other (perhaps simpler) symmetry groups such as $\mathcal{C}_3$ (threefold rotational symmetry).

If we also review the table of which simple symmetry groups in 3-space convert a framework that is minimally rigid when realized in a generic (random) geometry into a flexible framework when realized with the corresponding symmetry, we find only the groups $\mathcal{C}_2$, $\mathcal{D}_2$ and $\mathcal{D}_3$ (table 2). This striking coincidence invites an investigation of possible connections.

As observed in the review article of Goodsell & Olson, the evolutionary selection of symmetrical oligomeric complexes is not a coincidence—it 'is driven by functional, genetic, and physico-chemical needs' [2, p. 105]. We propose that there is a further specific functional connection related to the role of symmetry, flexibility and rigidity in the assembly and functioning of these oligomeric proteins. This paper will probe the background for the two observations, as well as some proposals for the functional role of symmetry in the proteins and how their corresponding flexibility supports the functions. Symmetry is a common feature; however, the native state reader should be aware that some reported symmetries in the protein structure

Table 2. Orbit counts for symmetric body–bar frameworks satisfying the count $|E| = 6|B| - 6$.

S	triv_S	$\lvert E\rvert$	$\lvert E_0\rvert$	$6\lvert B_0\rvert - \text{triv}_S$	f_S
$\mathcal{C}_1$	6	$6\lvert B\rvert - 6$	$6\lvert B_0\rvert - 6$	$6\lvert B_0\rvert - 6$	0
$\mathcal{C}_2$	2	$6\lvert B\rvert - 6$	$6\lvert B_0\rvert - 3$	$6\lvert B_0\rvert - 2$	1
$\mathcal{C}_3$	2	$6\lvert B\rvert - 6$	$6\lvert B_0\rvert - 2$	$6\lvert B_0\rvert - 2$	0
$\mathcal{C}_6$	2	$6\lvert B\rvert - 6$	$6\lvert B_0\rvert - 1$	$6\lvert B_0\rvert - 2$	−1
$\mathcal{D}_3$	0	$6\lvert B\rvert - 6$	$6\lvert B_0\rvert - 1$	$6\lvert B_0\rvert$	1

databases can be induced by the crystallization/crystal packing conditions and by artefacts of the modelling software.

This work builds on the last two decades of progress in predicting the flexibility and rigidity of three-dimensional frameworks with fast combinatorial counts and associated pebble game algorithms (see §3) [3–6]. Recently, this work culminated in the proof that these methods developed for general frameworks and implemented in software packages for predicting the rigidity and flexibility of proteins, such as FIRST [7] and ProFlex [8], also correctly predict the flexibility or rigidity of molecular frameworks—which mathematically model protein structures [9].

Some recent theoretical work has investigated the impact of symmetry on the flexibility properties of various types of frameworks [10–19]. In §5, we build on the results in [19] to present some very simple symmetry-adapted combinatorial counts that can be used to detect hidden motions (or at least modest fluctuation) in certain symmetric molecular structures that count as minimally rigid without symmetry. In particular, these counts provide the mathematical background for our observation that a symmetric molecular structure that counts as minimally rigid without symmetry becomes continuously flexible if and only if it is realized 'generically' with $\mathcal{C}_2$, $\mathcal{D}_2$ or $\mathcal{D}_3$ symmetry.

While motions that are predicted by our symmetry-adapted counts typically preserve the full symmetry group of the original structure throughout the path, we demonstrate in §6 that these counts can also be used to predict motions that break some, but preserve other, symmetries of the structure. In particular, we can gain information about what structural symmetries a motion preserves which is predicted by the standard non-symmetric counts and algorithms.

Finally, since our symmetry-adapted counts can predict hidden flexibility, which cannot be detected with the standard non-symmetric counts, and since these symmetry-adapted counts essentially only require a very simple counting of the number of orbits (i.e. sets of symmetric copies) of structural components under the given group action (see §§4 and 5 for details), it is natural to ask how these counts can be incorporated into software packages such as FIRST and ProFlex. In §7, we present some initial symmetry-adapted pebble game algorithms that provide improved predictions for the rigidity and flexibility of proteins that exhibit non-trivial symmetries.

We close with some proposals on how the symmetry-based motions could support observed protein behaviour and protein assembly (§8). Some of the connections are speculative—but based on general mechanical considerations that follow from the predicted flexibility and some general comments about biochemical properties associated with 'symmetry' in the biochemical literature. We are proposing plausible symmetry-based mechanical behaviour, drawing on analogies with other mechanical systems. Finally, in §9 we offer some areas for further work that are highlighted by questions raised in this analysis.

2. Symmetry in proteins

Many functioning soluble and membrane proteins assemble into symmetric multi-subunit oligomeric complexes composed of two or more identical proteins (monomers), which are held

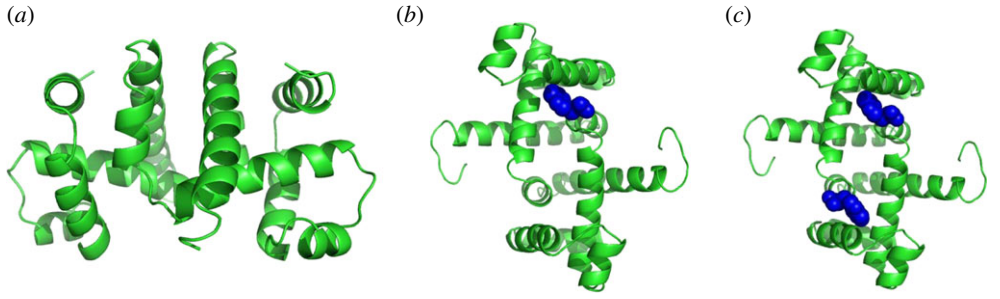

Figure 2. Tryptophan repressor with (*a*) no tryptophan bound (PDB ID: 3wrp), (*b*) one tryptophan bound and (*c*) both tryptophans bound (PDB ID: 1wrp). Both (*a*) and (*c*) have $\mathcal{C}_2$ symmetry, and it is speculated that $\mathcal{C}_2$ symmetry is preserved in (*b*), as has been observed for symmetric allosteric proteins such as arginine repressor [24]. (Online version in colour.)

together mostly by non-covalent interactions. Understanding symmetry and how it could impact protein flexibility and protein motions can provide direct insights and better understanding of protein function.

The preference of symmetries in proteins and their impact on protein function have been described [2,20–23]. In comparison to single non-symmetric proteins, symmetry often gives rise to large oligomeric complexes, which are generally more stable against denaturation and are better able to protect their surface area from the solvent. In other words, symmetry leads to lower solvent accessibility. Symmetric proteins are also more likely to avoid the unwanted aggregations that are often associated with 'amyloid and prion-related diseases' such as Alzheimer's, sickle-cell disease, mad cow disease and its human variant Creutzfeldt–Jakob disease, as oligomers with non-trivial point-group symmetries usually only have a small number of components [2]. It has also been speculated that symmetric proteins have greater folding efficiency (faster collapse) [2,20]. Furthermore, the energy landscapes of folding of symmetrical complexes are generally smoother than those of asymmetric structures [2,21].

Symmetry also plays an important role in allosteric interactions in proteins, where the binding of a small ligand at one ('allosteric') site of the protein causes a conformational change at a distant ('active') site (or possibly multiple sites) of the protein. Many large allosteric proteins exhibit non-trivial symmetries, because symmetric allosteric proteins are more likely to be selected by evolutionary forces, as they can have several identical active sites (as opposed to monomeric proteins, which have only a single active site) [2].

Molecular dynamics simulations indicate that symmetric proteins (such as tryptophan repressor) have a preference to have either no ligand bound at all or an entire 'symmetric orbit' of ligands bound (figure 2*a*,*c*). In other words, for symmetric proteins, the binding of one ligand at some site of the protein very quickly leads to the binding of further ligands at the corresponding symmetric sites via allosteric interactions, so that the resulting structure exhibits the same symmetry as the original protein.

Recent work by Strawn *et al*. [24] on arginine repressor reveals strong evidence for the famous symmetric model of allostery, also known as the MWC (Monod–Wyman–Changeux) or the concerted model, first asserted in 1965 by Monod *et al* [23]. Their results indicate that, as binding sites of symmetric allosteric proteins are filled sequentially and lead to subsequent conformational transitions, even during the partial ligation process (i.e. one, two, etc., or all ligands bound), the proteins maintain their symmetries. In other words, at every stage of the ligation process (prior to, during and after the ligation process), the symmetry of the protein (without the ligands) is conserved. This is rather remarkable considering that heterotropic allosteric proteins (unlike tryptophan or haemoglobin) can have interactions between different ligands (i.e. allosteric and effector ligands are not the same), which may need different binding pocket geometries. We will see in §8 that our results in §5 actually provide further support for the MWC model.

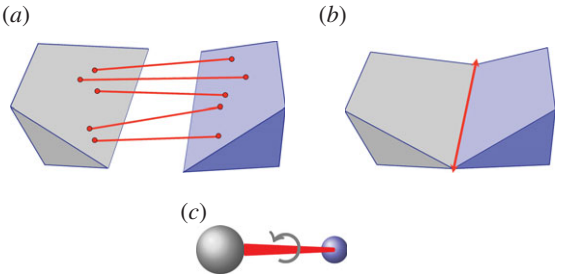

Figure 3. (*a*) A body–bar framework, (*b*) a body–hinge framework and (*c*) a molecular framework. (Online version in colour.)

It is important to observe that, while there exist examples of proteins from all rotational point groups in 3-space [2], the groups $\mathcal{C}_2$ (half-turn symmetry), $\mathcal{D}_2$ (dihedral symmetry of order four) and $\mathcal{D}_3$ (dihedral symmetry of order six) are the most common symmetry groups found in proteins. This is well reflected by table 5.4 in [1], which is partially shown in table 1.

In a survey of all naturally occurring proteins in *Escherichia coli*, it has been estimated that roughly 20% of the proteins are monomers and the majority of the proteins are dimers (40%) and tetramers (20%). Dimers (proteins consisting of two copies of a monomer) typically form with $\mathcal{C}_2$ symmetry and tetramers (proteins consisting of four monomers) mostly form with D_2 symmetry [2]. Note that it is also possible that monomeric proteins in the Protein Data Bank (PDB) are overstated due to the difficulties involved in protein crystallization.

It turns out that, while some symmetries have no impact, others have a significant impact on the flexibility of molecular structures. The boat conformation of cyclohexane, for instance, has $\mathcal{C}_2$ symmetry and is flexible, whereas the chair conformation of cyclohexane has $\mathcal{C}_3$ symmetry and is rigid (see example 5.3 and figure 9). In §5, we will show that the groups $\mathcal{C}_2$, $\mathcal{D}_2$ and $\mathcal{D}_3$ (that is, the most common symmetry groups found in proteins) are exactly the three groups that induce (symmetry-preserving) flexibility in molecular frameworks that are realized generically with these groups and count as minimally rigid without symmetry. We will investigate the possible conclusions from this for the assembly process and functioning of proteins in §8.

As a closing remark of this section, we would like to alert the reader that some claims about symmetries in proteins found in the biochemical data and literature are overrepresented. This is because symmetry is often presumed in the tools and methods used to refine and deposit molecular structures in biological databases (such as the PDB), as it can simplify and accelerate the required computations. (If a dimer is assumed to have half-turn symmetry, for example, then one may apply NMR spectroscopy to only one-half of the structure.) In other words, the symmetry of a protein found in a PDB file can sometimes arise as an artefact of the way the data for this structure have been collected, processed or refined.

3. Generic flexibility of molecular structures

It is well known that the rigidity of molecules and proteins can be analysed using the generic rigidity of body–bar frameworks [5,25,26]. A three-dimensional *body-bar framework* consists of a set of rigid bodies in $\mathbb{R}^3$ (each of dimension at least two) connected by rigid bars (figure 3*a*). Each of the bodies is free to move continuously in $\mathbb{R}^3$ subject to the constraints that the distance between any pair of points that are connected by a bar remains fixed. If *every* such motion also preserves the distance between *all* pairs of points belonging to different bodies (i.e. the motion extends to a congruent motion of Euclidean 3-space), then the framework is called *rigid*. Otherwise the framework is called *flexible*.

The underlying combinatorial structure for a three-dimensional body–bar framework is a multigraph $G = (B, E)$ (i.e. a graph with one or more edges running between connected vertices), which allows up to *six* edges (bars) between any pair of vertices (bodies). This upper bound for the number of bars between any pair of bodies is motivated by the fact that a three-dimensional

rigid body has six degrees of freedom (spanned by three translations and three rotations). So, in order to join two rigid bodies in $\mathbb{R}^3$ in such a way that the resulting structure is again rigid, one needs six properly placed bars.

For a body–bar framework with underlying multigraph $G=(B,E)$, we denote $\bar{B}$ as the set of endpoints of all bars of the framework. The positions of these endpoints in $\mathbb{R}^3$ (i.e. the positions of the attachment points of the bars on the bodies) are defined by the *configuration* $q:\bar{B}\to\mathbb{R}^3$ of the framework. Throughout this paper, we assume for simplicity that the map q is injective, i.e. all the attachment points on the bodies are pairwise distinct. A body–bar framework with underlying multigraph G and configuration q is denoted by (G,q).

To determine whether a given body–bar framework is rigid is, in general, a very difficult problem, as it requires solving a system of quadratic equations. It is therefore common to linearize this problem by differentiating the length constraints given by the rigid bars. This gives rise to the infinitesimal theory of rigidity, which we briefly outline next. Further details can be found in [4], for example.

Let (G,q) be a three-dimensional body–bar framework with underlying multigraph $G=(B,E)$. For every body b of (G,q), we choose a minimal set of points P_b that contains all the attachment points q_i of bars on b and spans an affine subspace of $\mathbb{R}^3$ of dimension at least two. We define P to be the set $P=\bigcup_{b\in B}P_b$. An *infinitesimal motion* of (G,q) is a map $u:P\to\mathbb{R}^3$ that satisfies

— $(p_i-p_j)\cdot(u(p_i)-u(p_j))=0$ for all points p_i and p_j in P that belong to the same body of (G,q);
— $(q_i-q_j)\cdot(u(q_i)-u(q_j))=0$ for all attachment points q_i and q_j in P that are connected by a bar of (G,q).

An infinitesimal motion u of (G,q) is called a *trivial infinitesimal motion* if $(p_i-p_j)\cdot(u(p_i)-u(p_j))=0$ for all pairs of points p_i and p_j in P (i.e. if u corresponds to a congruent motion of $\mathbb{R}^3$). Otherwise u is called an *infinitesimal flex* of (G,q). Framework (G,q) is said to be *infinitesimally rigid* if every infinitesimal motion of (G,q) is a trivial infinitesimal motion. Otherwise (G,q) is said to be *infinitesimally flexible*.

The infinitesimal rigidity properties of a body–bar framework (G,q) are completely described by its *rigidity matrix* $\mathbf{R}(G,q)$. This matrix (which is often written in the language of the projective Grassmann–Cayley algebra [4,6]) has six columns for each body of (G,q) (i.e. one column for each of the six degrees of freedom of the body) and one row for each bar of (G,q) (i.e. one row for each constraint on the body motions). So, $\mathbf{R}(G,q)$ is an $|E|\times 6|B|$ matrix, which has the following basic structure (see [4–6], for example, for details):

$$\begin{array}{c} \\ e_{i,j} \end{array}\begin{array}{c} \begin{array}{ccccc} & b_i & & b_j & \\ \end{array} \\ \begin{pmatrix} \cdots & \cdots & \cdots & \cdots & \cdots \\ 0\cdots 0 & (******) & 0\cdots 0 & -(******) & 0\cdots 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix} \end{array}.$$

The kernel of $\mathbf{R}(G,q)$ is the space of infinitesimal motions of (G,q). It is well known that the space of trivial infinitesimal motions of (G,q) is of dimension six (it is spanned by three infinitesimal translations and three infinitesimal rotations). So, if the dimension of the kernel of $\mathbf{R}(G,q)$ is strictly larger than six (or, equivalently, if $\operatorname{rank}(\mathbf{R}(G,q))<6|B|-6$), then (G,q) is infinitesimally flexible. Note that it follows immediately from the size of the rigidity matrix that if $|E|<6|B|-6$, then (G,q) has an infinitesimal flex.

It is a classic result that infinitesimal rigidity is a sufficient condition for rigidity.

A three-dimensional body–bar framework (G,q) is said to be *regular* if its rigidity matrix $\mathbf{R}(G,q)$ has maximal rank, i.e. if $\operatorname{rank}(\mathbf{R}(G,q))\geq\operatorname{rank}(\mathbf{R}(G,q'))$ for all possible realizations (G,q') of G as a body–bar framework. It is a simple consequence of this definition that the set of regular realizations of a multigraph G forms a dense open subset of all possible realizations of G as a body–bar framework. Thus, a regular body–bar framework is sometimes also called a 'generic' body–bar framework.

Asimov & Roth showed [27] that, for regular frameworks, infinitesimal rigidity is equivalent to rigidity. Moreover, it was shown by Tay [3] that the rigidity of a regular three-dimensional body–bar framework is completely determined by the combinatorial properties of its underlying multigraph [3] (see also [6]). Thus, we say that a multigraph G is *rigid* if it has a rigid regular realization as a body–bar framework.

Theorem 3.1 ([3]). *A regular three-dimensional body–bar framework (G, q) with underlying multigraph $G = (B, E)$ is infinitesimally rigid (and rigid) if and only if there exists a subset E^* of E that satisfies the conditions*:

1. $|E^*| = 6|B| - 6$;
2. $|E'| \leq 6|B'| - 6$ *for all subgraphs induced by subsets of* E^*.

While algorithmically this condition looks like we need to check all possible subsets of bars of (G, q) (an exponential process), these counts on a multigraph define independent sets in a matroid, and the counts lead to a greedy algorithm called the pebble game, which has a fast running time of $O(|B||E|)$ (e.g. [25,28,29]).

It is shown in [26,30] that these same counts (and the corresponding pebble game algorithms) also characterize regular rigid body–hinge frameworks in which bodies are connected by revolute hinges along assigned lines. These hinges function as implicit packages of five bars meeting the assigned hinge line (see also figure 3*b*). Moreover, very recently it has been confirmed that Tay's counts (and the corresponding pebble game algorithms) also characterize regular rigid molecular structures in which the atoms become the bodies, and each hinge is assigned the hinge line that joins the two atoms that it bonds (see also figure 3*c*). This result is called the molecular theorem (confirming the more than 20-year-old molecular conjecture) [9].

These algorithmic methods are implemented for basic predictions of the flexibility and rigidity of a protein, from a single geometric snapshot of the protein (e.g. a PDB file), in the web-based server FIRST developed and supported by Mike Thorpe [7].

4. Preliminaries about symmetry in body–bar frameworks

To formalize the notion of a symmetric body–bar framework (G, q) with underlying multigraph $G = (B, E)$, we consider the simple graph $\bar{G}$ (i.e. $\bar{G}$ has no multiple edges or loops) whose vertex set is the set $\bar{B}$ of endpoints of all bars of (G, q) (recall §3) and whose edge set $\bar{E}$ consists of all pairs of endpoints that are either connected by a bar in (G, q) or belong to the same body of (G, q). We define a *symmetry operation* of a body–bar framework (G, q) in $\mathbb{R}^3$ as an isometry R of $\mathbb{R}^3$ such that, for some graph automorphism $\alpha \in \mathrm{Aut}(\bar{G})$, we have

$$R(q(v)) = q(\alpha(v)) \text{ for all vertices } v \text{ of } \bar{G}.$$

The *symmetry element* corresponding to R is the affine subspace of points in $\mathbb{R}^3$ that are fixed by R. So, for instance, if R is a rotation, then the symmetry element of R is the rotational axis corresponding to R. The set of all symmetry operations of a body–bar framework (G, q) forms a group under composition, called the *point group* of (G, q).

We use the Schoenflies notation for the point groups and symmetry operations considered in this paper, as this is one of the standard notations in the literature about symmetric structures (e.g. [10,11,13–15,17,18,31]). Note that, while a molecule may have a reflection or improper rotation (i.e. a rotation followed by a reflection in the mirror perpendicular to the rotational axis) as a symmetry operation, the only possible symmetry operations of a protein are pure rotations (including the identity, which we may think of as a trivial rotation), in part because all α-helices of a protein have the same handedness. Since our main goal in this paper is to study the rigidity and flexibility of proteins, we will focus on the purely rotational point groups, which in the Schoenflies notation are denoted by $\mathcal{C}_n$, $\mathcal{D}_n$, $\mathcal{T}$, $\mathcal{O}$ and $\mathcal{I}$.

$\mathcal{C}_1$ is the trivial group, which consists only of the identity. $\mathcal{C}_n$, $n \geq 2$, is a group generated by an n-fold rotation (i.e. a rotation by an angle of $2\pi/n$); and $\mathcal{D}_n$, $n \geq 2$, is a group generated by an n-fold rotation and a twofold rotation whose rotational axis is perpendicular to the n-fold rotational axis. Finally, $\mathcal{T}$, $\mathcal{O}$ and $\mathcal{I}$ denote the groups that consist of all the rotational symmetry operations of a regular tetrahedron, a regular octahedron and a regular icosahedron, respectively. Throughout this paper, we will denote the identity operation by Id and a rotation by an angle of $2\pi/n$, $n \geq 2$, by C_n.

Note that the symmetry operations in the point group S of a body–bar framework (G, q) induce permutations of both the bodies and the bars of (G, q). In other words, the point group S of (G, q) acts on the set of bodies of (G, q) and on the set of bars of (G, q). The *body orbit* containing a body b of (G, q) consists of all the bodies of (G, q) we can arrive at by applying symmetry operations of S to b. Similarly, the *edge orbit* containing a bar e of (G, q) consists of all the bars of (G, q) we can arrive at by applying symmetry operations of S to e.

A body is said to be fixed by a symmetry operation R in S if it is fixed by the permutation of the bodies induced by R, i.e. if each attachment point on the body is mapped to a (possibly different) attachment point on the same body. Similarly, a bar is fixed by R if either $R(q(v)) = q(v)$ and $R(q(w)) = q(w)$ or $R(q(v)) = q(w)$ and $R(q(w)) = q(v)$, where $q(v)$ and $q(w)$ are the endpoints of the bar. So, loosely speaking, a structural component (body, bar) is fixed by R if it is not moved (but may be reoriented) by R.

Throughout this paper, we will only consider body–bar frameworks (G, q) whose point group acts freely on the bodies and bars of (G, q). In other words, (G, q) has neither body nor bar that is fixed by any non-trivial symmetry operation of S. This assumption is made because, in large structures such as proteins, few if any structural components occupy positions of non-trivial site symmetry [22]. Moreover, these restrictions, in particular the requirement that the point group acts freely on the bodies of the framework, facilitate the construction of orbit rigidity matrices (see [19]) with rows and columns indexed by the orbits of bars and bodies, respectively, leading to practical and theoretical advantages for the prediction of motions (and self-stresses) belonging to given irreducible representations of S [11,19] (see also §5).

Note that, if the point group S of a body–bar framework (G, q) with underlying multigraph $G = (B, E)$ acts freely on the bodies and bars of (G, q), then the size of each orbit is equal to the order of the group S. Thus, we have

$$|B| = |S||B_0| \tag{4.1}$$

and

$$|E| = |S||E_0|, \tag{4.2}$$

where $|B_0|$ and $|E_0|$ are the respective numbers of orbits of bodies and bars of (G, q) under the action of S, respectively.

While our symmetric counts are simplified by the assumption that the group S acts freely on structural components, they can immediately be transferred to frameworks that have joints or bars that are fixed by a non-trivial symmetry operation as well. In this case, we need to adjust the comparison of the counts, since there are now orbits of bars or bodies of different sizes (the size of an orbit corresponding to a structural component that is fixed by a non-trivial symmetry operation will no longer be equal to $|S|$). Furthermore, if bodies are fixed by non-trivial symmetry operations in S, then not only may we have body orbits of different sizes, but also the sets of columns corresponding to the bodies in the orbit matrix may now be of varying size. For example, if a body is fixed by a rotation, then the number of columns will be reduced to two [19].

5. When does symmetry lead to added flexibility?

In the following, we describe some very simple counts for detecting symmetry-preserving motions in symmetric body–bar frameworks.

Let (G, q) be a body–bar framework with symmetry group S. For every body b of (G, q), we choose a minimal set of points P_b that contains all the attachment points q_i of bars on b and

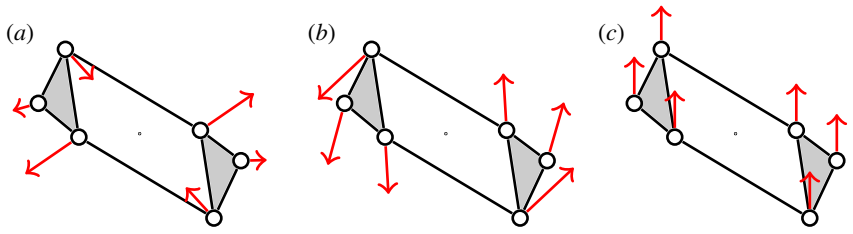

Figure 4. An example of a body–bar framework (G, q) with point group $\mathcal{C}_2$, shown looking down the axis. (*a*) The infinitesimal flex shown is $\mathcal{C}_2$-symmetric, since all the velocity vectors remain unchanged by the half-turn C_2. (*b*) The infinitesimal motion shown is a $\mathcal{C}_2$-symmetric trivial infinitesimal motion (corresponding to a rotation about the C_2 axis). (*c*) The trivial infinitesimal motion shown is not $\mathcal{C}_2$-symmetric, since the initial velocity vectors of the motion are reversed by C_2. (Online version in colour.)

spans an affine subspace of $\mathbb{R}^3$ of dimension at least two. We define P to be the set $P = \bigcup_{b\in B} P_b$. Moreover, if the set P contains points that are not endpoints of bars of (G, q), then we assume that these points were chosen so that they also exhibit the symmetries of S (figure 4). An infinitesimal motion $u : P \to \mathbb{R}^3$ of (G, q) is called S-symmetric if all the initial velocity vectors of u remain unchanged under all symmetry operations in S (see also figure 4) [19].

A simple way of testing whether a body–bar framework (G, q) with symmetry group S has an S-symmetric infinitesimal flex is to set up its *orbit rigidity matrix* $\mathbf{O}(G, q, S)$ (see e.g. [19] for details). In the language of the Grassmann–Cayley algebra [4,6], this matrix has one row for each edge orbit of G, and six columns for each vertex orbit of G under the action of the group S. Thus, $\mathbf{O}(G, q, S)$ is an $|E_0| \times 6|B_0|$ matrix with the following basic structure:

$$\begin{array}{c} \\ e_{i,R(j)} \\ \\ e_{i,R(i)} \\ \\ \end{array} \begin{array}{c} \begin{array}{ccccccc} & & b_i & & & b_j & \\ \end{array} \\ \begin{pmatrix} \cdots & \cdots & \cdots & \cdots & \cdots \\ 0\cdots0 & (******) & 0\cdots0 & (******) & 0\cdots0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0\cdots0 & (******) & 0\cdots0 & (000000) & 0\cdots0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}. \end{array}$$

For our purposes in this paper, the key property of the orbit rigidity matrix $\mathbf{O}(G, q, S)$ is that the kernel of $\mathbf{O}(G, q, S)$ is isomorphic to the space of S-symmetric infinitesimal motions of (G, q) [19]. So, if triv_S denotes the dimension of the space of S-symmetric trivial infinitesimal motions, then the count $|E_0| < 6|B_0| - \mathrm{triv}_S$ guarantees that (G, q) has an S-symmetric infinitesimal flex.

The dimension triv_S of the space of S-symmetric trivial infinitesimal motions can be read off from the character table of S. Comprehensive sets of character tables can be found in [32] (see also [10,11,13], for example).

A three-dimensional body–bar framework (G, q) with symmetry group S is said to be *S-regular* if its orbit rigidity matrix $\mathbf{O}(G, q, S)$ has the maximal rank, i.e. if $\mathrm{rank}(\mathbf{O}(G, q, S)) \geq \mathrm{rank}(\mathbf{O}(G, q', S))$ for all possible realizations (G, q') of G as a body–bar framework with symmetry group S. It follows from this definition that the set of S-regular realizations of a multigraph G forms a dense open subset of all possible S-symmetric realizations of G as a body–bar framework. Thus, we may think of an S-regular body–bar framework as a framework that is 'generic' modulo the symmetry constraints imposed by S.

It follows from the results in [18] that if an S-regular body–bar framework (G, q) has an S-symmetric infinitesimal flex, then it also has a non-trivial continuous motion that preserves the symmetry throughout the path. Together with the results in [19], this leads to the following necessary condition for an S-regular body–bar (or molecular) framework to be rigid.

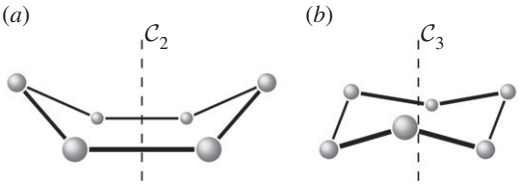

Figure 5. Two basic conformations of cyclohexane: (*a*) the 'boat' has half-turn symmetry and is flexible; (*b*) the 'chair' has threefold rotational symmetry and is rigid.

Theorem 5.1 ([19]). *If an S-regular body–bar* (*or molecular*) *framework* (G,q) *in* $\mathbb{R}^3$ (*with S acting freely on the bodies and bars of* (G,q)) *has only trivial symmetry-preserving continuous motions, then there exists a subset of representatives for the edge orbits,* E_0^*, *such that*:

(1) $|E_0^*| = 6|B_0| - \mathrm{triv}_S$;
(2) $|E_0'| \leq 6|B_0'| - \mathrm{triv}_S$ *for all subgraphs induced by subsets* E_0' *of* E_0^*.

Note that theorem 5.1 applies to both body–bar and molecular frameworks, because conditions (1) and (2) are necessary but not sufficient conditions for an *S*-regular body–bar framework to be rigid, and the rank of the orbit rigidity matrix of an *S*-regular realization of a multigraph *G* as a body–bar framework is at least as big as the rank of the orbit rigidity matrix of an *S*-regular realization of *G* as a molecular framework (see also §7 and conjecture 9.1 in §9).

In particular, we immediately obtain the following corollary from theorem 5.1.

Corollary 5.2. *If an S-regular body–bar* (*or molecular*) *framework* (G,q) *satisfies*

$$|E_0| < 6|B_0| - \mathrm{triv}_S,$$

then (G,q) *has a non-trivial symmetry-preserving continuous motion.*

Example 5.3. Consider a ring of six carbon atoms—carbon cyclohexane (figure 5). This starts with six atoms and six bonds—which count as 6×5 constraints. The overall Tay count is $|E| = 6 \times 5 = 30 = 6 \times 6 - 6 = 6|B| - 6$. Moreover, it is easy to check that the sparsity condition $|E'| \leq 6|B'| - 6$ of Tay's theorem is also satisfied. Thus, at regular configurations, the structure is rigid.

Suppose now that the structure is realized with $\mathcal{C}_2$ symmetry—the boat conformation of cyclohexane. Then, by (4.1) and (4.2), we have $|E_0| = 3 \times 5 = 15 = 6 \times 3 - 3 = 6|B_0| - 3$. However, since there is only a two-dimensional space of $\mathcal{C}_2$-symmetric trivial infinitesimal motions (spanned by a translation along the $\mathcal{C}_2$ axis and a rotation about the $\mathcal{C}_2$ axis [11,32]), we have $|E_0| = 15 < 16 = 6 \times 3 - 2 = 6|B_0| - \mathrm{triv}_S$. Thus, it follows from theorem 5.1 that, at $\mathcal{C}_2$-regular configurations, the boat has a non-trivial continuous motion that preserves the half-turn symmetry throughout the path.

Note that for the chair conformation of cyclohexane (cyclohexane realized with $\mathcal{C}_3$ symmetry), which is known to be rigid, the orbit counts give $|E_0| = 2 \times 5 = 10 = 6 \times 2 - 2 = 6|B_0| - \mathrm{triv}_S$.

Example 5.3 illustrates that a body–bar (or molecular) framework that counts as minimally rigid without symmetry becomes flexible (with a symmetry-preserving motion) if it is realized with half-turn symmetry.

We have seen in §2 that proteins (in particular dimers) often form with $\mathcal{C}_2$ symmetry. Given that the other most common groups found in proteins are $\mathcal{D}_2$ and $\mathcal{D}_3$ (recall table 1), it is natural to ask whether these groups also lead to flexibility in frameworks that count as minimally rigid without symmetry. In the following, we show that this is in fact the case.

It is a consequence of equations (4.1) and (4.2) that, if a body–bar framework (G,q) with point group *S*, where *S* acts freely on the bodies and bars of (G,q), satisfies the overall count $|E| = 6|B| - 6$, then the order of *S* must be equal to one, two, three or six. The orbit counts for all rotational point groups with this property are summarized in table 2. In this table (as well as

Table 3. Orbit counts for symmetric body–bar frameworks that count as minimally rigid without symmetry (i.e. for each group S, $|E|$ is the smallest number so that $|E| \geq 6|B| - 6$ and $|E|$ is divisible by $|S|$), but do not satisfy the count $|E| = 6|B| - 6$.

S	triv_S	$\lvert E\rvert$	$\lvert E_0\rvert$	$6\lvert B_0\rvert - \text{triv}_S$	f_S
$\mathcal{D}_2$	0	$6\lvert B\rvert - 4$	$6\lvert B_0\rvert - 1$	$6\lvert B_0\rvert$	1
$\mathcal{C}_4$	2	$6\lvert B\rvert - 4$	$6\lvert B_0\rvert - 1$	$6\lvert B_0\rvert - 2$	−1
$\mathcal{C}_5$	2	$6\lvert B\rvert - 5$	$6\lvert B_0\rvert - 1$	$6\lvert B_0\rvert - 2$	−1
$\mathcal{C}_{k\geq 7}$	2	$6\lvert B\rvert$	$6\lvert B_0\rvert$	$6\lvert B_0\rvert - 2$	−2
$\mathcal{D}_{k\geq 4}$	0	$6\lvert B\rvert$	$6\lvert B_0\rvert$	$6\lvert B_0\rvert$	0
$\mathcal{T}$	0	$6\lvert B\rvert$	$6\lvert B_0\rvert$	$6\lvert B_0\rvert$	0
$\mathcal{O}$	0	$6\lvert B\rvert$	$6\lvert B_0\rvert$	$6\lvert B_0\rvert$	0
$\mathcal{I}$	0	$6\lvert B\rvert$	$6\lvert B_0\rvert$	$6\lvert B_0\rvert$	0

in table 3), the integer $f_S = (6|B_0| - \text{triv}_S) - |E_0|$ in the final column indicates an f_S-dimensional space of S-symmetric infinitesimal flexes if $f_S > 0$, and an f_S-dimensional space of S-symmetric self-stresses (row dependences of the orbit rigidity matrix) if $f_S < 0$.

By theorem 5.1, the final column of table 2 implies that, at symmetry-regular configurations, the body–bar frameworks with $\mathcal{C}_2$ or $\mathcal{D}_3$ symmetry always have a non-trivial symmetry-preserving continuous motion.

The orbit counts for the remaining rotational groups S that do not allow the count $|E| = 6|B| - 6$ for body–bar frameworks with point group S are summarized in table 3. For each group S in table 3, $|E|$ is chosen to be the smallest number that satisfies $|E| \geq 6|B| - 6$ and is divisible by $|S|$; that is, $|E|$ is chosen to be the least number of edges for the framework to be rigid without symmetry and to be compatible with the symmetry constraints given by S.

By theorem 5.1, the final column of table 3 implies that, at symmetry-regular configurations, a body–bar framework with $\mathcal{D}_2$ symmetry that counts as over-braced by two without symmetry ($|E| = 6|B| - 4$) becomes flexible with a symmetry-preserving motion.

In summary, the groups $\mathcal{C}_2$, $\mathcal{D}_2$ and $\mathcal{D}_3$ are the most common symmetry groups found in proteins (see table 1) and they are also exactly the groups for which our symmetry-adapted counts detect (symmetry-preserving) flexibility in structures that count as minimally rigid without symmetry. We will investigate the significance of this for the behaviour of dimers and other proteins in §8.

We have focused on the impact of symmetry on flexes. There is a dual concept in rigidity theory of a stress in a structure—a dependence among the lengths and angles, which is sometimes viewed by biochemists as 'stabilizing the molecule'. Such a dependence will require that any fluctuation in lengths be fully coordinated—given some of the changes in length and angle, others will be constrained to specific connected values.

At a local level with rigidity matrix $\mathbf{R}(G, q)$, we have the linear equations $\mathbf{R}(G, q)d = s$ where s is the strain and d is the displacement. If rows are independent (no stress), then $\mathbf{R}(G, q)d = s$ has a solution for all $s \in \mathbb{R}^{|E|}$, and the dimension of the tangent space matches the dimension of the algebraic variety of displacements. Locally, any small strain is achievable with a small displacement.

If the rows of the rigidity matrix are dependent with a purely antisymmetric stress λ, then $\lambda\mathbf{R}(G, q)d = 0 = \lambda s$. Only a subspace of strains (changes in bond lengths, angles) of codimension one satisfying $\lambda s = 0$ are achievable without loading energy. At symmetry-generic configurations, all symmetric strains will satisfy this condition. Specifically, if $\mathbf{O}(G, q, S)$ has independent rows, then $\mathbf{O}(G, q, S)d_O = s_O$ has a solution for all symmetric strains. The reduction in the space of strains from the antisymmetric self-stress is in the subspace of antisymmetric strains, not symmetric strains.

If there is a symmetric stress, then the low-energy strains must satisfy the symmetric condition: $\lambda_O \mathbf{O}(G,q,S)d_O = \lambda_O s_O = 0$. This will cut the possible symmetric strains down to a codimension-one subspace. This reduces thermal fluctuations and gives more stability.

Recent work has shown that stresses in generic body–bar frameworks (redundant rigidity) are equivalent to global rigidity of the stressed substructure [33]—a strong form of stability since no other configuration of the bodies will respect the given constraints. There is a conjecture that this global rigidity also follows for generically redundantly rigid molecules.

The discussion above (and related extensions of results on global rigidity) suggests that a symmetric stress may be needed to ensure global rigidity within the space of symmetric molecular structures. We conjecture that the antisymmetric self-stress can trap the molecule into low-energy motions within the class of symmetric realizations, as happens with the boat configuration.

We conclude this section by drawing some further conclusions from the counts in tables 2 and 3 about how symmetry may impact the rigidity properties of a molecular structure when the counts do not predict flexibility but predict either independence or stresses.

(1) For the groups $\mathcal{D}_3$, $\mathcal{T}$, $\mathcal{O}$ and $\mathcal{I}$, a molecular structure that counts as infinitesimally rigid ($|E| = 6|B|$) also satisfies the 'symmetry-isostatic' count $f_S = (6|B_0| - \mathrm{triv}_S) - |E_0| = 0$. The initial count gives generic redundant rigidity, which was conjectured above to give global rigidity at generic molecular configurations. The count $f_S = 0$ suggests that, for symmetric molecular structures (proteins), the symmetry groups $\mathcal{C}_2$, $\mathcal{D}_2$, $\mathcal{T}$, $\mathcal{O}$ and $\mathcal{I}$ do not give rise to continuous flexibility and resist most antisymmetric strains, which would break symmetry. However, the standard arguments from Saard's theorem for independent structures suggest that symmetric configurations will have second symmetric realizations. Combined, these counts suggest that the structures support symmetry-preserving fluctuations of the structures, which only encounter modest energy barriers (see also proposition 1 in [12]).
(2) For $\mathcal{C}_2$, even the initial over-braced count $|E| = 6|B| - 4$ also leads to the symmetry-isostatic count $f_{\mathcal{C}_2} = (6|B_0| - \mathrm{triv}_{\mathcal{C}_2}) - |E_0| = 0$. This will behave in ways similar to (1) above.
(3) A molecular structure that counts as minimally rigid for any rotational group of the form $\mathcal{C}_{k\geq 4}$ (i.e. $|E| = 6|B| - k$ for $\mathcal{C}_k$, $k = 4,5,6$, and $|E| = 6|B|$ for $\mathcal{C}_{k\geq 7}$) has at least one fully symmetric self-stress if realized with such a point-group symmetry. Except for $k = 5,6$, these also have antisymmetric stresses. These added stresses further stabilize the structure against even symmetric fluctuations.

6. Symmetry-preserving and symmetry-breaking motions

The symmetry-adapted counts of theorem 5.1 may be used to detect hidden motions in symmetric structures. Moreover, with this result we can investigate which symmetries are preserved by motions predicted by Tay's standard non-symmetric counts. We illustrate this by means of the following example.

Example 6.1. Consider an eightfold molecular ring. The non-symmetric overall count for the underlying multigraph of the corresponding body–bar framework is $(6|B| - 6) - |E| = (6 \times 8 - 6) - 8 \times 5 = 2$. Thus, at regular configurations, this structure has two continuous degrees of freedom.

Suppose now that the eightfold ring is realized with $\mathcal{C}_2$ symmetry as shown in figure 6. Then the orbit counts for $\mathcal{C}_2$ are $(6|B_0| - \mathrm{triv}_{\mathcal{C}_2}) - |E_0| = (6 \times 4 - 2) - 20 = 2$. Thus, we conclude that for $\mathcal{C}_2$-regular configurations, the eightfold ring has two continuous $\mathcal{C}_2$-preserving degrees of freedom. This suggests that $\mathcal{C}_2$ symmetry does not lead to any added flexibility for the eightfold ring. In other words, the two degrees of freedom detected by the $\mathcal{C}_2$ orbit counts are essentially the 'same' degrees of freedom as the ones predicted by Tay's non-symmetric counts.

Let us now realize the eightfold ring with $\mathcal{C}_4$ symmetry. In this case, the orbit counts become $(6|B_0| - \mathrm{triv}_{\mathcal{C}_4}) - |E_0| = (6 \times 2 - 2) - 10 = 0$, and hence we do not detect any $\mathcal{C}_4$-symmetric

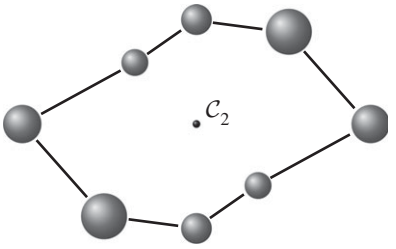

Figure 6. An eightfold molecular ring realized with $\mathcal{C}_2$ symmetry. Larger and smaller spheres indicate atoms that lie respectively in front of and behind the median plane of the structure.

motion. Together with the previous counts, this suggests that $\mathcal{C}_4$-regular realizations of the eightfold ring have two degrees of freedom, each of which preserves the $\mathcal{C}_2$ symmetry, but breaks the $\mathcal{C}_4$ symmetry.

Finally, if we realize the eightfold ring with $\mathcal{C}_8$ symmetry, then the structure is forced to lie in a hyperplane of $\mathbb{R}^3$, and the orbit counts become $(6|B_0| - \mathrm{triv}_{\mathcal{C}_8}) - |E_0| = (6 \times 1 - 2) - 5 = -1$. This says that there is a $\mathcal{C}_8$-symmetric self-stress, and hence an additional infinitesimal flex (which is not $\mathcal{C}_8$-symmetric). However, owing to the presence of the $\mathcal{C}_8$-symmetric self-stress, we cannot make any predictions for the (continuous) flexibility of the structure.

We note that more comprehensive counts for analysing the symmetry properties of motions of symmetric structures are described [10,11,13,31]. These counts are derived using techniques from group representation theory and character theory. In particular, in [11], conditions are given for a symmetric structure whose symmetry group acts freely on its structural components to have at least one infinitesimal flex within each irreducible representation of the group. A detailed analysis of when a (not necessarily fully symmetric) infinitesimal flex of a symmetric structure extends to a continuous motion is given in [12,18].

7. Symmetric pebble game algorithms

In §5, we have seen that with the symmetry-adapted counts in theorem 5.1 we can detect hidden motions in symmetric structures that are predicted to be rigid by Tay's standard non-symmetric counts. The following example shows that it is also possible for a symmetric body–bar framework to satisfy the symmetric counts in theorem 5.1, but to fail Tay's non-symmetric counts for rigidity. Such a framework will have a non-trivial symmetry-breaking motion.

Example 7.1. Consider the structure shown in figure 7, which consists of two fourfold molecular rings, sharing a $\mathcal{C}_2$ axis, connected by four additional bars. Without symmetry, each of the fourfold rings counts as $|E| = 4 \times 5 = 20 = 6 \times 4 - 4 = 6|B| - 4$, which is over-braced by two. With the $\mathcal{C}_2$ symmetry, each of the rings counts as $|E_0| = 10 = 6|B_0| - 2$, which still predicts no symmetric motions. With the four attaching connections between the rings (with twofold symmetry), we have $|E_0| = 4 \times 5 + 2 = 22 = 6 \times 4 - 2 = 6|B_0| - 2$. This still predicts no symmetric motions. However, counted without symmetry, each ring of four has two redundant edges, so we have a maximum of $2 \times 18 + 4 = 40$ independent edges without symmetry. Since $6|B| - 6 = 42$, it follows that there is at least a two-dimensional space of (non-symmetric) non-trivial infinitesimal motions.

We conclude that to test a symmetric structure for rigidity we must use both Tay's original criteria and the symmetry-adapted criteria, as they give distinct information. We need to run the two algorithms in a careful sequence.

While this is still an area of ongoing research, we can describe efficient algorithms that test necessary conditions for having rigidity, and therefore can predict significant flexibility (see also [34]). In the following, we describe a symmetry-adapted pebble game algorithm for the group $\mathcal{C}_2$ in terms of a dimer, in which each atom is a body and each molecular bond becomes five bars.

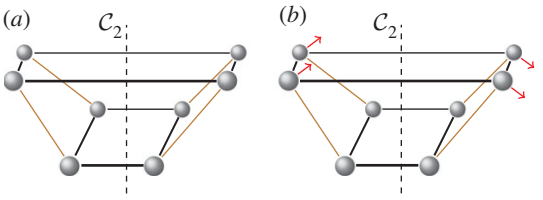

Figure 7. (*a*) A flexible structure that satisfies the orbit counts of theorem 5.1, but violates Tay's non-symmetric counts. The structure consists of two fourfold molecular rings (placed horizontally), which are connected by four single bars. (*b*) The vectors indicate an antisymmetric infinitesimal flex of the structure. (Online version in colour.)

Algorithm 7.2. *Given a body–bar multigraph $G = (B, E)$ with C_2 symmetry (e.g. a dimer), let $\hat{G}$ be the graph that is generated by one representative of each orbit of vertices under the action of C_2, and the induced edges for these vertices (see also figure 8). Apply the following sequence of steps*:

(1) (i) *Apply the standard $6|B| - 6$ pebble game to $\hat{G}$ [28,29]. Any edges that fail (i.e. are not pebbled) are placed in the set R_1 of essentially redundant edges, while edges that pass (i.e. are pebbled) are placed in E_1 (figure 8a).*
   (ii) *Copy this to the induced subgraph of the other vertex orbit representatives $\hat{G}^*$, again placing edges in R or in E_1^*.*
   (iii) *Continue to apply the $6|B| - 6$ pebble game to the remaining (bridging) edges, working on the two edges of an orbit, consecutively. If an edge passes this test, put it in E_2. If it fails this test, then do the additional test in steps (iv) and (v).*
   (iv) *If the failed search region [28,29] contains two copies of some vertex orbit, then put the edge in Q. Otherwise, put the edge in R_1 (figure 8b, c).*
   (v) *Repeat step (iv) immediately on the second copy of the same orbit.*
   (vi) *If these combined steps return a maximal set of edges $E_1 \cup E_1^* \cup E_2$ of size $< 6|B| - 6$, then the framework is flexible.*

(2) *The pebble game $6|B_0| - 2$ for the orbit graph G_0 (which is obtained from $\hat{G}$ by adding a set of representatives for the bridging edges)*:

   (i) *Start from the $6|B| - 6$ output of step (1)(i) on the vertices of $\hat{G}$. Run the $6|B_0| - 2$ pebble game on the orbit of each edge in $E_2 \cup Q$. If an orbit passes, put it in E_3. If it fails, put it in R_2.*
   (ii) *If this produces a maximal set $E_1 \cup E_3$ with $|E_1 \cup E_3| < 6|B_0| - 2$, then the dimer is C_2-symmetric flexible.*

Figure 8*a*, *b* illustrates why we need the extra checks in part (1)(iv),(v) of the algorithm to ensure that the output is independent of how we selected the representatives of the vertex orbits. The full invariance of the size of the pebbled set in step (2) of the algorithm in terms of such selections is a detail that needs to be confirmed.

Example 7.3. We illustrate algorithm 7.2 by applying it to the C_2-symmetric boat conformation of cyclohexane, whose underlying multigraph is shown in figure 9*a*. See [35] for a more detailed illustration.

In steps (1)(i),(ii) of the algorithm, we first apply the standard $6|B| - 6$ pebble game to $\hat{G}$ shown in figure 9*c* and then copy the result to the C_2-symmetric copy $\hat{G}^*$ of $\hat{G}$ shown in figure 9*d*. This yields $|E_1| = 2 \times 5 = 10$ and $|E_1^*| = 10$, and leaves $3 \times 6 - 10 = 8$ free pebbles for $\hat{G}$ (and for $\hat{G}^*$).

Then we proceed with step (1)(iii) of the algorithm, i.e. with the application of the $6|B| - 6$ pebble game to the remaining $2 \times 5 = 10$ bridging edges. This gives $|E_2| = 10$ (each edge in E_2 passes) and $|E_1 \cup E_1^* \cup E_2| = 10 + 10 + 10 = 30 = 6|B| - 6$, so that no non-trivial motion is detected (as expected).

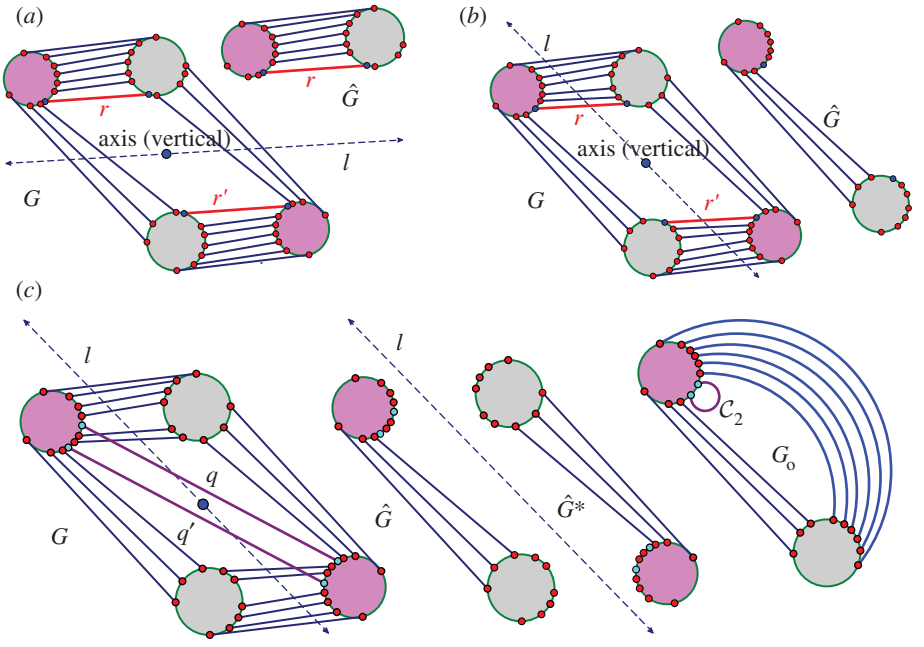

Figure 8. Three samples for the runs of algorithm 7.2. See also [35] for a detailed demonstration of the symmetric pebble game. (*a*) The algorithm declares the edges r and r' to be in R_1 in the first step, if tested last in the run on $\hat{G}$. It accepts all other edges not crossing ℓ in $E_1 \cup E_1^*$ and all bridging edges (crossing ℓ) in E_2. So, in step (1), the algorithm predicts no non-trivial motion since $|E_1 \cup E_1^* \cup E_2| = 6 + 6 + 6 = 18 = 6|B| - 6$. In step (2), the algorithm predicts a non-trivial symmetric motion since $|E_1 \cup E_3| = 6 + 3 = 9 < 10 = 6|B_0| - 2$. (*b*) The algorithm (applied to the same graph as in (*a*), but with a different choice of representatives for the vertex orbits) accepts all edges not crossing the new line ℓ in $E_1 \cup E_1^*$. All bridging edges except r and r' are placed in E_2, while r and r' (tested last among bridging pairs) are placed in R_1. So the counts are $|E_1 \cup E_1^* \cup E_2| = 3 + 3 + 12 = 18 = 6|B| - 6$ (again predicting no non-trivial motion in step (1)) and $|E_1 \cup E_3| = 3 + 6 = 9 < 10 = 6|B_0| - 2$ (again predicting symmetric flexibility in step (2)). (*c*) All edges not crossing ℓ are placed in $E_1 \cup E_1^*$, all bridging edges except q and q' are placed in E_2, and q and q' are placed in Q. Therefore, in step (1) of the algorithm we obtain $|E_1 \cup E_1^* \cup E_2| = 3 + 3 + 12 = 18 = 6|B| - 6$, and in step (2) we obtain $|E_1 \cup E_3| = 3 + 7 = 10 = 6|B_0| - 2$, and hence no non-trivial motion is detected. (Online version in colour.)

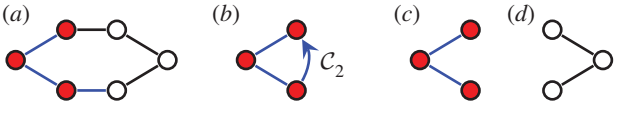

Figure 9. (*a*) The underlying multigraph G of cyclohexane, where each edge represents an implicit package of five edges. (*b*) The C_2-symmetric 'orbit graph' G_0 of cyclohexane. (*c*,*d*) The graph $\hat{G}$ and its symmetric copy $\hat{G}^*$. (Online version in colour.)

In step (2) we apply the symmetric $6|B_0| - 2$ pebble game to the orbit graph G_0 of G shown in figure 9*b*, starting from the $6|B| - 6$ output of step (1)(i). Since $Q = \emptyset$, we test orbits of bridging edges in E_2 only. Since there are eight free pebbles left after step (1)(i) and since there are five edge orbits in E_2, we obtain $|E_3| = 5$ (each edge orbit in E_2 passes). Thus, we have $|E_1 \cup E_3| = 10 + 5 = 15 < 16 = 6|B_0| - 2$, confirming that the boat has a C_2-symmetric non-trivial motion.

We caution the reader that, while algorithm 7.2 clearly provides necessary conditions for rigidity, and therefore sufficient conditions for flexibility, there are a number of details to be studied to confirm whether this provides both necessary and sufficient conditions for rigidity at symmetry-generic configurations. Note that, even if this is confirmed for general body–bar

frameworks, this result still needs to be extended to necessary and sufficient conditions for the rigidity of symmetry-regular molecular frameworks (see also conjecture 9.1 in §9). These problems are important areas of ongoing research.

Regarding step (1) of algorithm 7.2, we conjecture that, if a structure is predicted to have n degrees of freedom with the standard $6|B| - 6$ pebble game, then each of these n degrees of freedom is still present as a finite motion in a $\mathcal{C}_2$-regular realization of the structure.

Analogous symmetry-adapted pebble game algorithms for the groups $\mathcal{D}_2$ and $\mathcal{D}_3$ can be generated to provide sufficient conditions for flexibility.

8. Conclusions about the behaviour of proteins

Recall that our key observations in §5 were that

— the groups $\mathcal{C}_2$, $\mathcal{D}_2$ and $\mathcal{D}_3$ are the most common symmetry groups found in proteins, and
— the groups $\mathcal{C}_2$, $\mathcal{D}_2$ and $\mathcal{D}_3$ are also exactly the groups for which our orbit counts detect flexibility (with a symmetry-preserving motion) in frameworks that count as minimally rigid without symmetry.

In the following, we will explore some possible explanations for this connection.

First, we claim that symmetry-preserving motions can facilitate the self-assembly of stressed but not strained proteins. In normal circumstances, when assembling a stressed structure, there should be some resistance in placing the last set of bars (bond). If the structure has $\mathcal{C}_2$, $\mathcal{D}_2$ or $\mathcal{D}_3$ as a symmetry group, in the assembly process just before stress occurs, the simultaneous placement of the last symmetric components (bars or bonds) will only introduce nominal stress, whereas in the asymmetric assembly process the stress is more likely also to cause strain (non-trivial deformation) in the structure. As an example, consider a structure that is realized with $\mathcal{D}_2$ symmetry and is redundantly rigid (stressed) and over-counted by two in the regular non-symmetric count. We know that this structure will still have a symmetry-preserving motion (table 3), and with this available additional motion the amount of resistance in the placement of the last set of bars (bonds) will be decreased. In this context, we can envision that in the stressed protein the formation of the last set of symmetric hydrogen bonds and/or hydrophobic interactions across the symmetric copies, with their predetermined geometrical alignment, may be more easily achieved with the presence of symmetry. In general, we expect that the presence of these common symmetry groups facilitates the assembly of stressed structures, without incurring too much strain in the system.

The preference and evolutionary selection of symmetry groups $\mathcal{C}_2$, $\mathcal{D}_2$ and $\mathcal{D}_3$ in biological complexes should not be accidental. In fact, the symmetric flexibility induced by these groups plays an important role in carrying specific biological functions of the oligomeric complexes that have these symmetries. As an example, we consider the allosteric behaviour of dimeric proteins, for which we observe a preference for $\mathcal{C}_2$-symmetric states.

We recall that allostery describes a phenomenon in which binding/releasing of a ligand at one site of a protein with a subsequent change in shape/conformation at its binding pocket also induces a change in shape at a second site of the protein, and in the process increases the likelihood of the binding/releasing of a ligand at the second site. For instance, in the tryptophan repressor dimer, when there is a lot of tryptophan present, tryptophan binds to one of the two dimers, which induces a change in shape and a subsequent increase in affinity at the other binding site. When there is little tryptophan present, one of the tryptophans is released, which leads to a decrease in affinity of the other binding site and a release of its tryptophan (figure 1). It has been observed that, for $\mathcal{C}_2$-symmetric oligomeric complexes, there is a drive to resume the $\mathcal{C}_2$ symmetry, where the dimer is functional, via allosteric actions (both or none of the ligands are bound) (see §2), and symmetry is often maintained during the entire ligation process. This would suggest that, even with only one tryptophan bound, the tryptophan repressor should maintain $\mathcal{C}_2$ symmetry. It is reasonable to assume that, as one ligand docks and changes the shape of its binding pocket, a

corresponding $\mathcal{C}_2$ symmetry-preserving change of shape at the other binding pocket is preferred, as it now has the ideal geometry for binding the second ligand. This would indicate that the entire allosteric transition occurs along the pathway of a C_2 symmetry-preserving motion. Thus, we claim that the pathway of a $\mathcal{C}_2$ preserving motion may better support the allosteric communication in $\mathcal{C}_2$ allosteric proteins, such as tryptophan repressor, so that when one tryptophan binds (or is released) the entire protein is pushed along to make the same corresponding symmetric change at the second binding site.

Consider the docking of a ligand of a fixed shape into a binding pocket that is fluctuating over a range of shapes. What fraction of the time is the pocket shaped correctly for the ligand to bind—and how can symmetry of the overall protein (not of the pocket), restricting the fluctuation to a symmetric path, improve the binding? As a simple thought experiment, assume there are k parameters for the variation in the shape of the pocket. Assume the variation of the ith parameter is over a range m_i, and d_i is the fraction of that parameter space that contributes 'shape' to support docking. Then the overall fraction of the variation space that supports docking is $\prod_i (d_i/m_i)$. Now, if the symmetry traps the fluctuation into a one-parameter path, then the odds of a fit with the ligand is increased to (d_1/m_1)—greatly improving the chances of a fit. Of even greater impact, if the two binding pockets are coordinated by the symmetry, then they always have the same shape. Therefore, as soon as a ligand docks at one site, the other symmetric site is frozen into the position that will always afford a docking of a second copy of the ligand!

In a dimer pair docking with $\mathcal{C}_2$ symmetry, the contacts, which would generically lock the pair in a rigid form without symmetry, leave an added degree of freedom. Either this flexibility is preserved for functional reasons, or the dimer moves along the $\mathcal{C}_2$-symmetric path to a point where an additional bonding occurs to give strong stability. We note that a structure that is infinitesimally rigid without any redundancy (stress) will be able to fluctuate with any thermodynamic variation in even one bond length or angle with no resistance from required deformations of other bonds or bond angles. However, the presence of stress means that changing one bond length or angle will require a coordinated response from other lengths and angles in the stress—so added bonds will add stability.

Recall that if a molecular structure that counts as minimally rigid is realized with $\mathcal{C}_2$ symmetry, then it picks up not only a $\mathcal{C}_2$-symmetric motion, but also an (antisymmetric) self-stress. One may anticipate that this self-stress will add stability holding the structure to preserve C_2 symmetry as it moves along the path. While one fluctuating length or angle will require a compensating fluctuation in the stress (perhaps the opposite change at the $\mathcal{C}_2$-symmetric bond or angle due to the antisymmetric self-stress), the motion along the path requires no changes in length—just thermodynamic changes in the configuration of the vertices. This supports our anticipation that the only local realizations of the conformation are also $\mathcal{C}_2$-symmetric. In other words, the dimer could be stably trapped in $\mathcal{C}_2$-symmetric configurations, with one symmetric degree of freedom (and a self-stress without strain) to function. This behaviour is consistent with the observed data for protein dimers (recall §2), and we conjecture that this plays an important role in their allosteric behaviour.

Finally, we note that our symmetry-adapted counts and methods can also be used to analyse the motions of virus capsids. It is well known that virus capsids typically have symmetry $\mathcal{I}$, i.e. the rotational symmetries of the icosahedron (see also figure 10). One model of virus functioning is that the capsid swells (flexes) in some way to permit material to exchange from the capsid to infect a cell. The other model is that it disassembles to release the material. It is a common observance in virus capsids that they swell and expand their size and still retain the icosahedral symmetry. A study by Kovacs *et al.* [36,37] using physical model building and symmetry-based calculations reproduces some of these properties. More specifically, they describe a model that shows flexing of the shape along the twofold axes, with the fivefold axes fixed, and the threefold axes opening up. This reinforces our claim that symmetry-preserving flexibility (induced by half-turn symmetry, for example) plays a crucial role in the behaviour and functioning of many proteins, viruses and other biological assemblies.

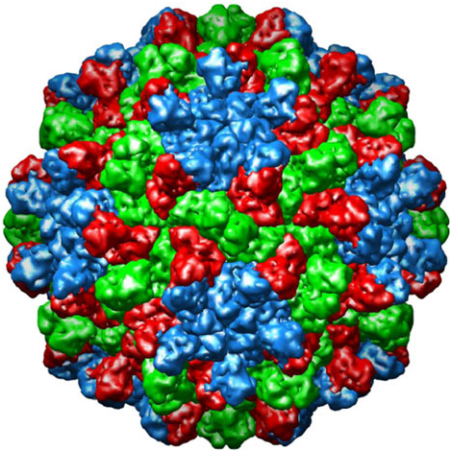

Figure 10. A tomato bushy stunt virus (PDB ID: 2tbv) with icosahedral symmetry. (Online version in colour.)

9. Future work

The results, tools and observations presented in this paper clearly indicate that any rigidity exploration of dimers and other symmetric oligomeric proteins should take into account their altered rigidity properties, which occur due to twofold rotational symmetries, as well as $\mathcal{D}_2$ and $\mathcal{D}_3$ symmetry.

These connections invite further exploration of the symmetry within allosteric pathways of other proteins with $\mathcal{C}_2$, $\mathcal{D}_2$ and $\mathcal{D}_3$ symmetry. As noted in §2, sometimes the tools used to create such structures assume the symmetry as a simplifying assumption, so it may be necessary to collect additional biochemical data that specifically look for symmetry and symmetry breaking in allosteric behaviour of such oligomers.

We have seen in §5 that symmetry-preserving flexibility (or fluctuation) induced by certain symmetries in proteins can be predicted with some very simple combinatorial counts. Moreover, in §7, we have presented an initial symmetry-adapted pebble game algorithm that tracks both the standard non-symmetric and the symmetric counts for the group $\mathcal{C}_2$. This algorithm, as well as the analogous symmetry-adapted pebble game algorithms for the groups $\mathcal{D}_2$ and $\mathcal{D}_3$, should be implemented into software packages such as FIRST.

An important problem that still remains open is to find both necessary and sufficient conditions for the rigidity of a symmetric body–bar framework. Moreover, if such conditions are found, they need to be extended to both necessary and sufficient conditions for the rigidity of symmetric molecular structures. We propose the following symmetric version of the molecular theorem [9].

Conjecture 9.1. An S-regular realization of a multigraph G as a body–bar framework is rigid if and only if an S-regular realization of G as a molecular framework is rigid.

Note that, while there has recently been some significant progress in finding both necessary and sufficient conditions for symmetric orbit matrices of various types of frameworks to have full rank, i.e. for symmetry-regular frameworks to have no non-trivial motion that maintains the symmetry of the structure throughout the path (see [19,38,39] for example), fairly little is known about sufficient conditions for symmetry-regular frameworks to have no non-trivial motion at all [13,40]. In fact, for finite body–bar frameworks in 3-space, the only known results are summarized in [13].

Another interesting area of future research is to study the rigidity and flexibility of finite structures (such as body–bar or molecular frameworks) that have no non-trivial point-group symmetry overall, but consist of substructures that exhibit non-trivial symmetries. From this

we could gain further insight into the rigidity properties of proteins that consist of multiple symmetric components, for example. However, owing to the possibly complex interactions between the various symmetric components, it is still unclear what methods to use for analysing the rigidity properties of such structures.

Another natural question concerning the behaviour of proteins, crystals and other biological structures (such as zeolites, for example) is how to analyse the rigidity of finite repetitive frameworks, that is, finite sub-frameworks of infinite periodic frameworks. In particular, such an analysis could be applied to study the rigidity of some key secondary structures of proteins, such as β-sheets for example. While there has been a rapid and extensive development in the rigidity theory of infinite periodic frameworks (with and without added crystallographic symmetry) in recent years [19,38,39], the behaviour of finite substructures of such infinite periodic structures is still quite poorly understood.

Funding statement. B.S. is partially supported by the Fields Institute and under a grant from NSERC (Canada). A.S. and W.W. are supported under a grant from NSERC (Canada).

References

1. Schulz GE, Schirmer RH. 1979 *Principles of protein structures*. New York, NY: Springer.
2. Goodsell D, Olson A. 2000 Structural symmetry and protein function. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 105–153. (doi:10.1146/annurev.biophys.29.1.105)
3. Tay T-S. 1984 Rigidity of multi-graphs, linking rigid bodies in *n*-space. *J. Comb. Theory B* **36**, 95–112. (doi:10.1016/0095-8956(84)90016-9)
4. Whiteley W. 1996 *Some matroids from discrete applied geometry.* Contemporary Mathematics, vol. 197, pp. 171–311. Providence, RI: American Mathematical Society.
5. Whiteley W. 2005 Counting out to the flexibility of molecules. *Phys. Biol.* **2**, 1–11. (doi:10.1088/1478-3975/2/4/S06)
6. White N, Whiteley W. 1987 The algebraic geometry of motions of bar-and-body frameworks. *SIAM J. Alg. Discr. Methods* **8**, 1–32. (doi:10.1137/0608001)
7. FIRST, protein flexibility web server. See http://flexweb.asu.edu.
8. ProFlex. See http://www.bmb.msu.edu/~kuhn/software/proflex/index.html.
9. Katoh N, Tanigawa S. 2011 A proof of the molecular conjecture. *Discr. Comput. Geom.* **45**, 647–700. (doi:10.1007/s00454-011-9348-6)
10. Connelly R, Fowler PW, Guest SD, Schulze B, Whiteley W. 2009 When is a symmetric pin-jointed framework isostatic? *Int. J. Solids Struct.* **46**, 762–773. (doi:10.1016/j.ijsolstr.2008.09.023)
11. Fowler PW, Guest SD, Schulze B. 2014 Mobility of symmetry-regular bar-and-joint frameworks. In *Fields Institute Workshop on Rigidity and Symmetry, Toronto, Canada, 17–21 October 2011*.
12. Guest SD, Fowler PW. 2007 Symmetry conditions and finite mechanisms. *Mech. Mater. Struct.* **2**, 293–301. (doi:10.2140/jomms.2007.2.293)
13. Guest SD, Schulze B, Whiteley W. 2010 When is a symmetric body–bar structure isostatic? *Int. J. Solids Struct.* **47**, 2745–2754. (doi:10.1016/j.ijsolstr.2010.06.001)
14. Kangwai RD, Guest SD. 1999 Detection of finite mechanisms in symmetric structures. *Int. J. Solids Struct.* **36**, 5507–5527. (doi:10.1016/S0020-7683(98)00234-0)
15. Kangwai RD, Guest SD. 2000 Symmetry-adapted equilibrium matrices. *Int. J. Solids Struct.* **37**, 1525–1548. (doi:10.1016/S0020-7683(98)00318-7)
16. Owen JC, Power SC. 2010 Frameworks, symmetry and rigidity. *Int. J. Comput. Geom. Appl.* **20**, 723–750. (doi:10.1142/S0218195910003505)
17. Schulze B. 2010 Block-diagonalized rigidity matrices of symmetric frameworks and applications. *Contrib. Alg. Geom.* **51**, 427–466.
18. Schulze B. 2010 Symmetry as a sufficient condition for a finite flex. *SIAM J. Discr. Math.* **24**, 1291–1312. (doi:10.1137/090776238)
19. Schulze B, Whiteley W. 2011 The orbit rigidity matrix of a symmetric framework. *Discr. Comput. Geom.* **46**, 561–598. (doi:10.1007/s00454-010-9317-5)
20. Andre I, Strauss CEM, Kaplan DB, Bradley P, Baker D. 2008 Emergence of symmetry in homooligomeric biological assemblies. *Proc. Natl Acad. Sci. USA* **105**, 16 148–16 152. (doi:10.1073/pnas.0807576105)

21. Wolynes PG. 1996 Symmetry and the energy landscapes of biomolecules. *Proc. Natl Acad. Sci. USA* **93**, 14 249–14 255. (doi:10.1073/pnas.93.25.14249)
22. Stenkamp RE. 2009 Protein quaternary structure: symmetry patterns. In *Encyclopedia of Life Sciences* (*ELS*). Wiley Online Library. (doi:10.1002/9780470015902.a0003121.pub2)
23. Monod J, Wyman J, Changeux JP. 1965 On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.* **12**, 88–118. (doi:10.1016/S0022-2836(65)80285-6)
24. Strawn R, Melichercik M, Green M, Stockner T, Carey J, Ettrich R. 2010 Symmetric allosteric mechanism of hexameric *Escherichia coli* arginine repressor exploits competition between L-arginine ligands and resident arginine residues. *PLoS Comput. Biol.* **6**, e1000801. (doi:10.1371/journal.pcbi.1000801)
25. Jacobs DJ, Rader AJ, Kuhn LA, Thorpe MF. 2001 Protein flexibility predictions using graph theory. *Proteins Struct. Funct. Genet.* **44**, 150–165. (doi:10.1002/prot.1081)
26. Tay T-S, Whiteley W. 1985 Recent advances in generic rigidity of structures. *Struct. Topol.* **9**, 31–38.
27. Asimov L, Roth B. 1978 The rigidity of graphs. *Trans. Am. Math. Soc.* **245**, 279–289. (http://www.jstor.org/stable/1998867)
28. Lee A, Streinu I. 2005 Pebble game algorithms and sparse graphs. Paper presented at *EuroComb'05, European Conf. on Combinatorics, Graph Theory and Applications, Berlin, 5–9 September 2005*. See http://arxiv.org/abs/math/0702129v1.
29. Sljoka A. 2012 Algorithms in rigidity theory with applications to protein flexibility and mechanical linkages. PhD thesis, York University, Toronto. See http://www.math.yorku.ca/∼adnanslj/adnanthesis.pdf.
30. Whiteley W. 1988 Matroid unions and rigidity. *SIAM J. Discr. Math.* **1**, 237–255. (doi:10.1137/0401025)
31. Fowler PW, Guest SD. 2000 A symmetry extension of Maxwell's rule for rigidity of frames. *Int. J. Solids Struct.* **37**, 1793–1804. (doi:10.1016/S0020-7683(98)00326-6)
32. Bishop DM. 1973 *Group theory and chemistry*. Oxford, UK: Clarendon Press.
33. Connelly R, Jordán T, Whiteley W. Generic global rigidity of body–bar frameworks. *J. Comb. Theory B* **103**, 689–705. (doi:10.1016/j.jctb.2013.09.002)
34. Schulze B, Sljoka A, Whiteley W. 2011 Protein flexibility of dimers: do symmetric motions play a role in allosteric interactions? In *Advances in Mathematical and Computational Methods: Addressing Modern Challenges of Science, Technology, and Society, Waterloo, Canada, 25–29 July 2011* (eds I Kotsireas, R Melnik, B West). AIP Conf. Proc., vol. 1368, pp. 135–138. Melville, NY: American Institute of Physics. (doi:10.1063/1.3663478)
35. Schulze B, Sljoka A, Whiteley W. 2013 Symmetric pebble game algorithm demonstration. See http://www.math.yorku.ca/∼ adnanslj/adnanthesis.pdf.
36. Kovacs F, Tarnai T, Guest SD, Fowler P. 2004 Double-link expandohedra: a mechanical model for expansion of a virus. *Proc. R. Soc. A* **460**, 3191–3202. (doi:10.1098/rspa.2004.1344)
37. Kovacs F, Tarnai T, Guest SD, Fowler P. 2004 Construction of a mechanical model for the expansion of a virus. In *Proc. IASS-2004 Symp., Shell and Spatial Structures from Models to Realization, Montpellier, France, 20–24 September 2004*. International Association for Shell and Spatial Structures.
38. Malestein J, Theran L. 2011 Generic rigidity of frameworks with orientation-preserving crystallographic symmetry. See http://arxiv.org/abs/1108.2518v2.
39. Ross E, Schulze B, Whiteley W. 2011 Finite motions from periodic frameworks with added symmetry. *Int. J. Solids Struct.* **48**, 1711–1729. (doi:10.1016/j.ijsolstr.2011.02.018)
40. Schulze B. 2010 Symmetric versions of Laman's theorem. *Discr. Comput. Geom.* **44**, 946–972. (doi:10.1007/s00454-009-9231-x)