On the Symmetric Molecular Conjectures

Josep M. Porta, Lluis Ros, Bernd Schulze, Adnan Sljoka, and Walter Whiteley

Abstract A molecular linkage consists of a set of rigid bodies pairwise connected by revolute hinges where all hinge lines of each body are concurrent. It is an important problem in biochemistry, as well as in robotics, to efficiently analyze the motions of such linkages. The theory of generic rigidity of body-bar frameworks addresses this problem via fast combinatorial algorithms. However, recent work has shown that symmetry (a common feature of many molecular and mechanical structures) can lead to additional motions. These motions typically maintain the original symmetry of the structure throughout the path, and they can often be detected via simple combinatorial counts. In this paper, we outline how these symmetry-based mathematical counts and methods can be used to efficiently predict the motions of symmetric molecular linkages, and we numerically analyze configuration spaces supporting the symmetric Molecular Conjectures formulated herein.

Key words: rigidity, flexibility, symmetry, molecule, linkage, configuration space.

1 Introduction

Accurate measurements of flexibility and dynamics of proteins and other molecules can help us interpret the relationship between structure and function, which has significant implications in medicine and drug design [12, 7]. This is an important area of research in computational biology, material science and bioinformatics, and has lately attracted a lot of interest in the robotics community [11].

The authors are listed in alphabetical order.

Josep M. Porta and Lluis Ros

Institut de Robòtica i Informàtica Industrial, UPC-CSIC e-mail: {porta,ros}@iri.upc.edu

Bernd Schulze

Dept. of Mathematics and Statistics, Univ. of Lancaster e-mail: b.schulze@lancaster.ac.uk

Adnan Sljoka and Walter Whiteley

Dept. of Mathematics and Statistics, York Univ. e-mail: {adnanslj,whiteley}@mathstat.yorku.ca

Over the last 15 years, rigidity-based methods like FIRST, ProFlex, or Kinari [3, 7] have been used to give fast computational predictions of the flexibility of molecular structures, such as DNA, RNA, and proteins. These methods detect rigidity or flexibility in generic molecular linkages using combinatorial (counting) characterizations of rigidity, which can be verified via fast pebble game algorithms [9, 16]. They are based on the mathematical theory of generic rigidity of molecular linkages (Section 2), which are built of rigid bodies (atoms) linked in pairs via revolute hinges (rotatable bonds between the atoms) [7, 21].
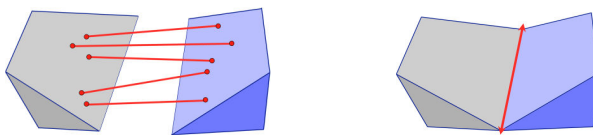
However, recent work by Schulze et al. [13, 14, 15] shows that symmetry can lead to additional flexibility that is not detected by the more general characterizations. The point is crucial because many molecular rings, proteins, or viral capsids are symmetric and symmetry plays a central role in macromolecular stability, assembly, ligand recognition and drug docking, and in phenomena such as allostery [4, 16].

In this paper, we demonstrate how this increased flexibility can still be detected through simple symmetry-adapted counts, and then we formulate the symmetric molecular conjectures of rigidity theory (Section 3). To test the power of the counts, and to provide evidence supporting the conjectures, we use the higher-dimensional continuation tools of the CUIK suite [10, 5] on illustrative examples (Section 4).

2 Detecting flexibility in body-bar and molecular linkages

A 3-dimensional *body-bar framework* consists of a set of rigid bodies in $\mathbb{R}^3$ which are connected by rigid bars (see Fig. 1, left). Each body is free to move continuously in $\mathbb{R}^3$, subject to the constraints that the distance between any pair of points which are connected by a bar remains fixed. A body-bar framework is called *rigid* if *every* such motion also preserves the distance between *all* pairs of points belonging to different bodies. Otherwise the framework is called *flexible* [20, 21].

The underlying combinatorial structure for a body-bar framework is a multigraph $G = (B, E)$, where the vertices in B represent the bodies of the framework and the edges in E represent the rigid bars of the framework [20, 21]. To determine whether a given body-bar framework with underlying multigraph $G = (B, E)$ 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 leads to a homogeneous system of $|E|$ linear equations in $6|B|$ unknowns, and the corresponding coefficient matrix of this

Fig. 1 A body-bar framework and its corresponding body-hinge framework.

linear system is called the *rigidity matrix* of the framework. The elements in the kernel of this matrix are called the *infinitesimal motions* of the framework, and it is well known that the space of trivial infinitesimal motions of a 3-dimensional body-bar framework (i.e. the space of infinitesimal motions of the framework which correspond to congruent motions of Euclidean 3-space) is of dimension 6. Thus, a body-bar framework possesses a non-trivial infinitesimal motion if and only if the dimension of the kernel of the rigidity matrix is strictly larger than 6.

It is a well known fact that if a body-bar framework is infinitesimally rigid (i.e. it has only has trivial infinitesimal motions), then it is also rigid. The converse is not true in general. However, if the framework is *regular*, that is, if the rigidity matrix of the framework with underlying multigraph G has maximal rank among all realizations of G as a body-bar framework, then the existence of a non-trivial infinitesimal motion also guarantees the existence of a non-trivial continuous motion of the framework [1]. 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, and hence 'almost all' realizations of G as a body-bar framework are regular.

Tay showed that for a regular body-bar framework, the rigidity is captured by the following simple combinatorial counts [18]:

Theorem 1. *(Tay, 1984) A regular 3-dimensional body-bar framework with underlying multigraph $G = (B, E)$ is rigid if and only if there exists a subset E^* of E which satisfies the conditions:*

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

Tay's counts are related to general Grübler-Kutzbach counting rules, however they are more precise as they can detect topological redundancies. While algorithmically Tay's condition looks like we need to check all possible subsets of bars of the framework (an exponential process!), these counts on a multigraph define independent sets in a matroid, leading to a fast greedy algorithm called the pebble game, with running time of $O(|B||E|)$ [9, 16].

A 3-dimensional *body-hinge framework* consists of a set of rigid bodies which are connected, in pairs, along revolute hinges (i.e., lines in 3-space), and each of the bodies is free to move continuously, subject to the constraints that the contacts along the hinges are preserved (see Fig. 1, right). Since a hinge removes 5 of the 6 relative degrees of freedom between a pair of rigid bodies in 3-space, a body-hinge framework can be modeled as a special case of a body-bar framework by replacing each hinge with 5 independent bars, each intersecting the hinge line.

The counts in Theorem 1 (and the corresponding pebble game algorithms) also characterize regular rigid body-hinge frameworks in 3-space [19, 20], where a body-hinge framework with multigraph G is called regular if its rigidity matrix has maximal rank among all body-hinge realizations of G. Moreover, the recent Molecular Theorem [8] confirmed that Tay's counts also characterize regular rigid molecular linkages, where a *molecular linkage* is a body-hinge framework with the special geometry that all hinges of each body are concurrent in a single point. This result solved the more than 20 year old Molecular Conjecture [19].

3 Detecting symmetry-preserving motions in symmetric linkages

We say that a structure is symmetric if it possesses a non-trivial point group, i.e., a non-trivial group of symmetry operations (isometries in Euclidean 3-space) which leave the structure unchanged. Many natural structures such us proteins, as well as many human-built structures such as linkages and other mechanical machines are symmetric [4]. As a consequence, there has been a growing interest in the impact of symmetry on the rigidity and flexibility of these structures. In particular, a symmetric analog of the rigidity matrix, called the *orbit rigidity matrix*, has been constructed in [15] to detect hidden symmetry-preserving motions in symmetric frameworks.

For simplicity, we will only consider body-bar frameworks whose symmetry group acts freely on the bodies of the framework (that is, no body of a symmetric framework is unshifted by any non-trivial symmetry operation of the framework). Note that for a framework with a purely rotational point group, this means that no body is centered on a rotational axis. For such frameworks, the symmetry-adapted rigidity analysis leads to some very simple Tay-type counts (see Theorem 2).

If we focus only on motions which preserve the full symmetry group S of the original framework, the motion of a single body b uniquely determines the motions of all bodies that lie in the same orbit as b under the group action of S (i.e., all bodies $s(b)$, $s \in S$) as well. So we have 6 degrees of freedom for each orbit of bodies. Similarly, each orbit of bars (i.e., each bar together with all of its symmetric copies) restricts the motion of the combined structure by one. Therefore, the orbit rigidity matrix has $|E_o|$ rows and $6|B_o|$ columns, where $|E_o|$ and $|B_o|$ denote the number of bar orbits and body orbits of the framework, respectively.

The key property of the orbit rigidity matrix is that its kernel is the space of S-symmetric infinitesimal motions of the framework (see [15] for details). In the following, we denote the dimension of the space of trivial S-symmetric infinitesimal motions of the framework by $triv_S$. Note that for any symmetry group S in dimension 3, $triv_S$ can be read off from the character tables in [2], for example.

Analogously to the non-symmetric situation, a body-bar (or molecular) framework with symmetry group S and underlying multigraph G is said to be S-regular, if its orbit rigidity matrix has maximal rank among all S-symmetric realizations of G as a body-bar (or molecular) framework. From the results in [13] it follows that if an S-regular body-bar framework has an S-symmetric non-trivial infinitesimal motion, then it also has a non-trivial continuous motion which preserves the symmetry group S throughout the path. This fundamental result leads to the following necessary conditions for an S-regular body-bar (or molecular) framework to be rigid:

Theorem 2. *(Schulze, Whiteley, 2010 [15, 13]) If an S-regular body-bar (or molecular) framework in $\mathbb{R}^3$ (with S acting freely on the bodies) has only trivial symmetry-preserving continuous motions, then there exists a subset of representatives for the bar orbits, E_0^*, such that:*

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

It is shown in [17] that there are some further necessary conditions for rigidity, and that this larger set of conditions is also sufficient for an *S*-regular body-bar framework to have no symmetry-preserving motion.

Very recently, some initial symmetry-extended versions of the pebble game algorithm - based on the orbit counts in Theorem 2 and on Tay's original counts - have been developed for the group $\mathscr{C}_2$ in order to test frameworks for (possibly symmetry breaking) flexibility [14]. While these algorithms clearly provide sufficient conditions for the flexibility of a $\mathscr{C}_2$-symmetric structure, an area of ongoing research is whether they also provide necessary conditions for flexibility.

Note that Theorem 2 applies to both body-bar and molecular frameworks, because (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. We offer the following two conjectures

Conjecture 1. (**Symmetric Molecular Conjecture I**) The orbit rigidity matrix of an *S*-regular body-bar realization of a multigraph *G* has the *same* rank as the orbit rigidity matrix of an *S*-regular molecular realization of *G*.

In other words, we conjecture that under *S*-regular conditions, body-bar realizations of *G* and molecular realizations of *G* have the same number of *S*-preserving degrees of freedom.
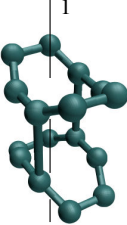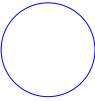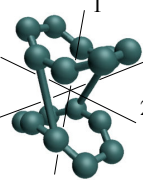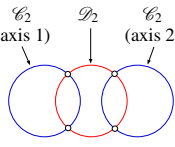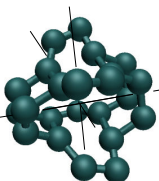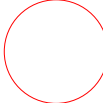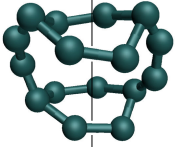
Conjecture 2. (**Symmetric Molecular Conjecture II**). The rank of the entire rigidity matrix - and not just the orbit matrix - of an *S*-generic body-bar realization of a multigraph *G* is the same as the rank of the rigidity matrix of an *S*-generic molecular realization of *G*, where a structure is called *S*-generic if the rank of the entire rigidity matrix is maximal, among all *S*-symmetric realizations.

It would follow from this conjecture that under *S*-generic conditions, body-bar realizations and molecular realizations of *G* have exactly the same number of (not necessarily *S*-preserving) infinitesimal motions.

4 Evidence for the symmetric Molecular Conjectures

To collect evidence supporting the previous conjectures, we have applied the higher-dimensional continuation methods implemented in the CUIK suite [10, 5] to a variety of symmetric molecular linkages. These methods allow an exhaustive tracing of the C-space component of the linkage that is path-connected to a given configuration. The exploration marches from such configuration in all directions systematically by constructing local charts on the tangent bundle. Neighboring charts are mutually coordinated to keep track of the C-space region explored up to a given point, and branch-switching methods are applied at bifurcation points in order not

Table 1 Tay and orbit counts for selected molecules, and the results of the numerical analysis.

Molecular structure	Symmetry group	Tay's counts	Orbit counts	C-space analysis
	$\mathscr{C}_2$ $(triv_{\mathscr{C}_2} = 2)$	$\|B\| = 16$ $\|E\| = 18 \cdot 5 = 90$ $\|E\| = 6\|B\| - 6$ Predicts rigid molecule	$\|B_0\| = 8$ $\|E_0\| = 45$ $\|E_0\| < 6\|B_0\| - 2$ Predicts 1-dim. $\mathscr{C}_2$ symmetric motion	$\mathscr{C}_2$ (axis 1)
	$\mathscr{D}_2$ $(triv_{\mathscr{D}_2} = 0)$	$\|B\| = 16$ $\|E\| = 18 \cdot 5 = 90$ $\|E\| = 6\|B\| - 6$ Predicts rigid molecule	$\mathscr{D}_2 \; \|B_0\| = 4$ $\|E_0\| = 23$ $\|E_0\| < 6\|B_0\| - 0$ Predicts 1-dim. $\mathscr{D}_2$ symm. motion $\mathscr{C}_2$: $\|B_{0'}\| = 8$ $\|E_{0'}\| = 45$ $\|E_{0'}\| < 6\|B_{0'}\| - 2$ Predicts 1-dim. $\mathscr{C}_2$ symmetric motion	$\mathscr{C}_2$ (axis 1) $\quad \mathscr{D}_2 \quad$ $\mathscr{C}_2$ (axis 2)
	$\mathscr{D}_2$ $(triv_{\mathscr{D}_2} = 0)$	$\|B\| = 24$ $\|E\| = 28 \cdot 5 = 140$ $\|E\| > 6\|B\| - 6$ Predicts rigid molecule	$\|B_0\| = 6$ $\|E_0\| = 35$ $\|E_0\| < 6\|B_0\| - 0$ Predicts 1-dim. $\mathscr{D}_2$ symmetric motion	$\mathscr{D}_2$
	$\mathscr{C}_2$ $(triv_{\mathscr{C}_2} = 2)$	$\|B\| = 16$ $\|E\| = 18 \cdot 5 = 90$ $\|E\| = 6\|B\| - 6$ Predicts rigid molecule	$\|B_0\| = 8$ $\|E_0\| = 45$ $\|E_0\| < 6\|B_0\| - 2$ Predicts 1-dim. $\mathscr{C}_2$ symmetric motion	2-dimensional C-space with $\mathscr{C}_2$ sub loop. Only the 6-revolute rings always maintain $\mathscr{C}_2$ symmetry.

to leave areas unexplored [6]. In the end, the component topology can be inferred from the neighborhood relationships between charts.

Table 1 provides a selection of the molecular linkages analyzed so far. For each linkage, we show the initial configuration used for continuation, its symmetry group, the number of predicted motions according to the Tay and orbit counts, and the C-space topology inferred from the numerical analysis. We focus on linkages exhibiting the half-turn symmetry group $\mathscr{C}_2$ and the dihedral group $\mathscr{D}_2$ of order 4 (which is generated by two perpendicular half-turn axes), as these are the most common groups for molecular structures such as proteins [4, 14].

Note that each of the linkages in Table 1 is predicted to be rigid with Tay's non-symmetric counts. (In fact, the third linkage even counts to be over-constrained by two.) However, in each case we have $6|B_0| - triv_S - |E_0| = 1$, so that we may conclude from Theorem 2 that for S-regular realizations, each of these molecular linkages has a symmetry-preserving non-trivial motion (Note that $triv_{\mathscr{C}_2} = 2$ in Table 1 since the space of $\mathscr{C}_2$-symmetric trivial motions is generated by a translation along the $\mathscr{C}_2$ axis and a rotation about the $\mathscr{C}_2$ axis.)

From the topology of the configuration space of the first and the third linkage it follows that there are no additional (symmetry-preserving or symmetry-breaking) non-trivial motions for these linkages, which provides evidence for both symmetric Molecular Conjectures.

Similarly, the configuration space of the second linkage also supports the first symmetric Molecular Conjecture, since it shows that the linkage has only one $\mathscr{D}_2$-preserving degree of freedom at every $\mathscr{D}_2$-regular point, and only one $\mathscr{C}_2$-preserving degree of freedom at every $\mathscr{C}_2$-regular point. (At the bifurcation points, the linkage is still $\mathscr{D}_2$-regular, and the additional motion only preserves one of the half-turn axes of $\mathscr{D}_2$ so it is not $\mathscr{C}_2$-regular at those points.) This example is also consistent with the Symmetric Molecular Conjecture II, since the bifurcation points are neither $\mathscr{C}_2$- nor $\mathscr{D}_2$-generic, as the rank of the entire rigidity matrix at the bifurcation points is clearly not maximal among the $\mathscr{C}_2$- or $\mathscr{D}_2$-symmetric realizations of the structure.

Finally, while the orbit counts predict the fourth linkage in Table 1 to have one non-trivial $\mathscr{C}_2$-preserving motion, a more detailed symmetry-based analysis of the subgraphs of the two 6-revolute rings shows that there is in fact an additional symmetry-breaking motion as well. Our computations of the configuration space again confirm that there is indeed a two-dimensional C-space for this molecular linkage. In addition, our computations show that each of the two 6-revolute rings of the structure maintains its half-turn symmetry at all times, which supports our expectation that stressed (over-constrained under symmetry) components of a structure (such as the symmetric 6-rings in the linkage) are more likely to maintain their symmetry than unstressed ones (such as the connecting links between the 6-rings) [14].

5 Conclusions

This paper has described the symmetric molecular conjectures of rigidity theory, giving their context, formulation, and a selection of examples providing supporting evidence for them. The associated combinatorial methods allow to predict hidden motions in symmetric molecules that otherwise seem apparently over-constrained and rigid. The ultimate aim of this work is to move from the theoretical analysis and to apply the symmetry-based methods and pebble game algorithms to actual proteins, to better understand their functions and possible drug targets.

References

[1] Asimov, L., Roth, B.: The Rigidity Of Graphs. AMS **245**, 279–289 (1978)

[2] Atkins, P., Child, M., Phillips, C.: Tables for group theory. Oxford University Press (1970)

[3] F., J., Hardy, J., Streinu, I.: Using rigidity analysis to probe mutation-induced structural changes in proteins. J. of Bioinf. and Comp. Bio. **10**(10) (2012)

[4] Goodsell, D., Olson, A.: Structural symmetry and protein function. Annu. Rev. Biophys. Biomol. Struct. **29**, 105–153 (2000)

[5] Henderson, M.E.: Multiple parameter continuation: Computing implicitly defined k-manifolds. Int. Journal of Bif. and Chaos **12**(3), 451–476 (2002)

[6] Henderson, M.E.: Multiparameter parallel search branch switching. Int. Journal of Bif. and Chaos in App. Sci. and Eng. **15**(3), 967–974 (2005)

[7] Jacobs, D., Rader, A., Kuhn, L., Thorpe, M.: Protein flexibility predictions using graph theory. Proteins: Struc., Func., and Gen. **44**, 150–165 (2001)

[8] Katoh, N., Tanigawa, S.: A proof of the molecular conjecture. Discrete & Computational Geometry **45**(4), 647–700 (2011)

[9] Lee, A., Streinu, I.: Pebble game algorihms and (k, l)-sparse graphs. European Conference on Combinatorics pp. 181–186 (2005)

[10] Porta, J.M., Ros, L., Bohigas, O., Manubens, M., Rosales, C., Jaillet, L.: The CUIK suite: Motion analysis of closed-chain multibody systems (Submitted)

[11] Porta, J.M., Ros, L., Thomas, F., Corcho, F., Cantó, J., Pérez, J.J.: Complete maps of molecular-loop conformational spaces. J. Comput. Chem. **28**, 2170–2189 (2007)

[12] R.M. Daniel R.V. Dunn, J.F., Smith, J.: The role of dynamics in enzyme activity. Annu. Rev. Biophys. Biomol. Struct. **32**, 69–92 (2003)

[13] Schulze, B.: Symmetry as a sufficient condition for a finite flex. SIAM Journal on Discrete Mathematics **24**(4), 1291–1312 (2010)

[14] Schulze, B., Sljoka, A., Whiteley, W.: How does symmetry impact the rigidity of proteins? To appear in Phil. Trans. of the Royal Society A (2013)

[15] Schulze, B., Whiteley, W.: The orbit rigidity matrix of a symmetric framework. Discrete & Computational Geometry **46**(3), 561–598 (2011)

[16] Sljoka, A.: Algorithms in rigidity theory with applications to protein flexibility and mechanical linkages www.math.yorku.ca/~adnanslj/adnanthesis.pdf. Ph.D. Thesis, York University (2012)

[17] Tanigawa, S.: Matroids of Gain Graphs in Applied Discrete Geometry. arXiv:1207.3601 (2012)

[18] Tay, T.S.: Rigidity of multi-graphs, linking rigid bodies in n-space. J. Comb. Theory, B **36**, 95–112 (1984)

[19] Tay, T.S., Whiteley, W.: Recent advances in generic rigidity of structures. Structural Topology **9**, 31–38 (1985)

[20] Whiteley, W.: Some Matroids from Discrete Applied Geometry. Contemporary Mathematics **197**, 171–311 (1996)

[21] Whiteley, W.: Counting out to the flexibility of molecules. Physical Biology **2**, 1–11 (2005)