Molecular symmetry deficiency and shape deviation measures

Paul G. Mezey

Résumé **Quasi-symétrie moléculaire et mesures d'écart à la similarité de forme**
Un objet à l'état naturel est rarement d'une symétrie parfaite. Il reste que le concept de symétrie est d'une grande puissance simplificatrice et que même les symétries imparfaites permettent des simplifications utiles. Ainsi l'étude de quasi-symétries mène souvent à de meilleurs modèles et à une meilleure compréhension de certains phénomènes. L'on peut qualifier une quasi-symétrie par un écart de symétrie, celui-ci étant lui-même un cas particulier de mesure d'écart à la similarité de forme. C'est pourquoi les mesures d'écart à la similarité de forme forment un ensemble général qui inclut les mesures d'écart à la symétrie. Dans cet article, les mesures de quasi-symétrie sont formulées en termes de mesures d'écart à la similarité de forme, avec une attention particulière aux implications de ces dernières pour les molécules et pour la structure moléculaire dans la conception de médicaments. La chiralité, ce cas particulier d'écart à la symétrie, est d'une grande importance en biochimie.
Mots-clés **Quasi-symétrie, chiralité, mesures d'écart à la similarité de forme, forme moléculaire, QSAR.**

Abstract Symmetry is seldom perfect for any actual object of nature. Yet, symmetry is a powerful simplifying concept, and even approximate symmetry can provide useful simplifications. Hence, the study of approximate symmetry often leads to better models and better insight. Approximate symmetry can be formulated in terms of symmetry deficiency, and symmetry deficiency is in its turn a special case of shape deviation. For this reason, shape deviation measures form a general framework that also incorporates symmetry deficiency measures. In this contribution, approximate symmetry measures are formulated in terms of shape deviation measures, with particular emphasis on their applications to molecules and to the molecular basis of pharmaceutical drug design. Chirality, as a special case of symmetry deficiency, is of major importance in biochemistry.
Keywords **Symmetry deficiency, chirality, shape deviation measures, molecular shape, QSAR.**

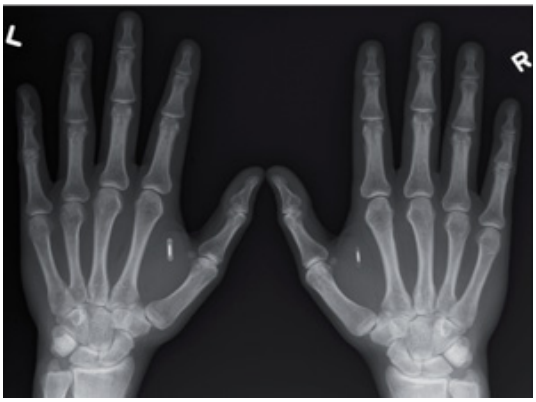
This paper is dedicated to the memory of Professor Jacques-Émile Dubois, in recognition of his pioneering work on the unambiguous coding of chirality, the associated revision of the Cahn-Ingold-Prelog rules, and his work extending chemical informatics over interdisciplinary boundaries.

Handedness, the English equivalent of the more commonly used term of Greek origin, "chirality", refers to the left-right distinction of shapes that are both mirror images of each other, yet different, and non-superimposable. Many objects exhibit various levels of symmetry and chirality; several examples are shown as illustrations in *figure 1*. Chirality is of special importance in chemistry, since two mirror image forms of molecules of otherwise identical composition often have radically different effects. Although constitutionally identical, chiral molecules may possess very different biological properties. In pharmaceuticals, a racemic mixture is one that has equal amounts of left- and right-handed enantiomers of a chiral molecule. In some cases, it is not only more efficient for pharmaceutical manufacturers to market the individual enantiomers but also highly beneficial to the patient because the inactive isomer of a racemic form may have undesired, even toxic side effects or counteract the benefits of the active isomer. One of the two forms may for instance be a harmless substance or even a beneficial medication, whereas the other form may be a potent poison (*figure 2*). It is therefore not surprising that the study of chirality in chemistry has led to important discoveries, and this is reflected in the efforts aimed at concise characterization of chiral molecules. This effort may be regarded as a part of a more general problem: the characterization of molecular shapes.

Symmetries corresponding to various symmetry elements, such as mirror planes or rotation axes, may be regarded as special shape properties. These symmetry elements do not uniquely characterize a shape; nevertheless, two objects with the shame shape must also have the same symmetry elements. Having the same symmetries is a necessary, but not a sufficient condition for two objects to have exactly the same shape. Similarly, shape deviations and shape deficiencies can be regarded as more general cases of symmetry deficiencies. The special case of chirality is in fact characterized by the lack of mirror planes and the lack of S(2n)-type, rotation-reflection symmetry axes, which makes chirality a special case of symmetry deficiency. It is therefore natural to seek relations between chirality measures, more general symmetry deficiency measures, and even more general shape deviation measures. The present contribution focuses on one such approach.

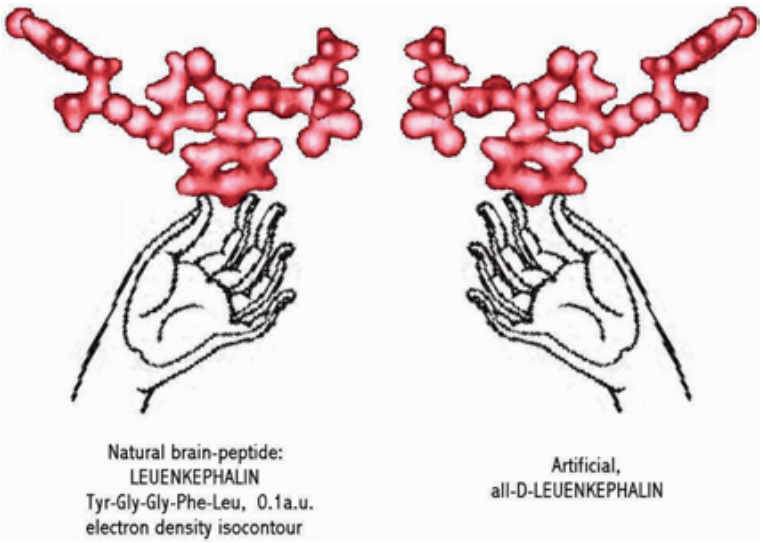
In recent decades, focus on the systematic organization of chemical information, for instance in terms of advanced chemical databases, has increased the need for simple

(a)



(c)

Janus de Chartes (b)



(d)

Figure 1 - Four examples on symmetry and chirality.
(a) Symmetry in architecture: the Taj Mahal (India), considered by many as one of the most perfect examples of symmetry. (b) Chirality in art: the statue of Janus on the Cathedral of Chartres is that of the Roman god Janus who is typically depicted with two, slightly asymmetrical opposing faces, representing beginnings and endings. (c) Chirality in anatomy: the word *chirality* is derived from the Greek χειρ (cheir), the hand, the most familiar chiral object. Although the X-ray image shown is on a planar film thus achiral and the actual bone structure is three-dimensional and thus chiral. (d) Chirality in chemistry: a chiral molecule and its mirror image are said to be enantiomers, from the Greek εναντιοξ (enantios) "opposite" and μεροξ, part. They have identical physical properties except for their ability to rotate plane-polarized light by equal amounts in opposite directions. In proteins and peptides, only the left-handed form of amino acids occur.

numerical codes which allow a systematic organization of shape-related molecular information. The advances in numerical chirality codes studied by Dubois and coworkers had important connections, not only to the development of the DARC chemical informatics approach [1-7], but also to the more general area of molecular shape classification [8-39].

In the present context, we shall distinguish shape and size: two objects which can be transformed into each other by uniform scaling are regarded as having the same shape, shape thus being regarded as a size-independent property, much as symmetry is regarded as a size-independent property. For example, we may consider the unique spherical shape, a common property of all spheres, or we may consider a particular ellipsoidal shape with a fixed ratio in the length of its ellipsoidal axes. In the latter case, the shape of an ellipsoid with axes of respectively 1, 2 and 3 cm long is identical to that of an ellipsoid with axes of respectively 5, 10 and 15 km long. Some simple examples are shown in *figure 3*.

Shape families and sets of objects with common shape properties

Consider a reference shape *S*, and the family *F(S)* of all objects presenting this shape. We may question how much the shape of a given object *A* deviates from this reference shape *S*. For simplicity, we shall regard only objects with well-defined surfaces and volumes, and we shall formulate

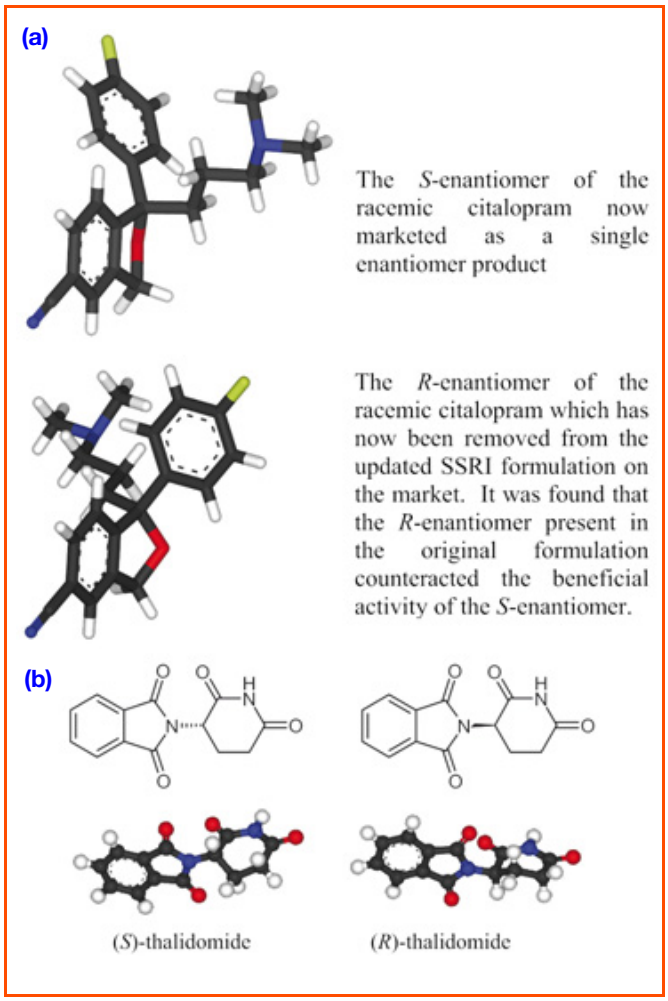


Figure 2 – Chiral molecules.
(a) Citalopram, an antidepressant drug belonging to the class of drugs known as selective serotonin reuptake inhibitors (SSRIs), has a stereocenter, to which 4-fluorophenyl group and an *N*, *N*-dimethyl-3-aminopropyl groups bind. Due to this chirality, the molecule exists in two enantiometric forms (mirror images). They are termed *S*-(+)-citalopram and *R*-(–)-citalopram. Citalopram was initially marketed as Celexa/Cipramil, a racemic mix. Subsequently, an updated formulation called escitalopram, which is the *S*-enantiometer of the racemic citalopram, was marketed as Lexapro/Cipralex, and found to be more efficient. It also caused fewer undesirable side-effects such as nausea and weight gain. (b) Thalidomide: an example of a racemic drug in which one enantiomer (left) is toxic and produces a teratogenic side effect, whereas its chiral counterpart (right) has desirable anti-emetic and sedative effects.

the description of the associated shape deviation measures in terms of sets in the three-dimensional space.

We shall say that a set *B* is an *S*-set if *B* has the shape *S*, that is, if *B* belongs to the family *F(S)*:

$$B \in F(S) \qquad (1)$$

Alternatively, we may consider only some specific shape property, for example *Q*, such a simple shape property being the ratio *q* of the shortest *a* and the longest *b* distances between two parallel planes both touching the object *A* from opposite sides:

$$q = a/b \qquad (2)$$

Consider a shape property *Q*, and the family *F(Q)* of all objects having this shape property. In general, we shall say that a set *B* is a *Q*-set if *B* has the shape property *Q*, that is, if *B* belongs to the family *F(Q)*:

$$B \in F(Q) \qquad (3)$$

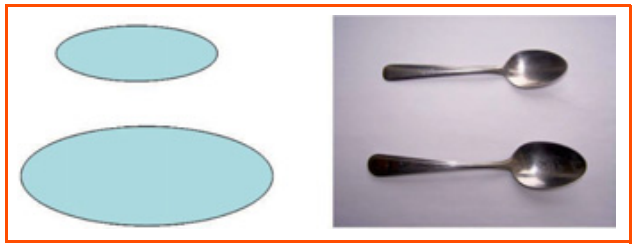


Figure 3 - Shape and size.
Left: two ellipses of different sizes but with the same ratio for the length of their axes, hence these ellipses have the same shape. Right: two spoons of the same shape, of the same symmetry, but of different sizes. ***"Shape thus being regarded as a size-independent property, much as symmetry is regarded as a size-independent property"***. For real objects, such as spoons, symmetry and identity of shape can never be perfect, but these notions are very useful at some imperfect level of resolution.

In fact, *Q* may stand for the entire set of shape properties defining a shape *S*, and consequently, the two cases can be treated on a common basis. In the following, first we shall discuss the case of reference shape *S*, but we should point out that the same derivations are equally valid if *S* is replaced by just a single shape property *Q*.

Shape deviation measure (SDM) and shape property deviation measure (SPDM)

In the following, while in order to visualize the meaning of the statements, we may wish to think of the body enclosed by an isodensity contour of a molecular electron density cloud when referring to a set, the treatment is clearly more general and applicable to other objects as well.

With regard to some reference shape *S*, a set *B* is an *S*-subset of a set *A* if *B* is a subset of *A*, and *B* is a *S*-set.
Set *B* is a maximum volume *S*-subset of a set *A* if *B* is an *S*-subset of *A* and if volume *V*(*B*) is maximum among all *S*-subsets of *A* (note that, while *B* is not necessarily unique, *V*(*B*) is).
Set *C* is an *S*-superset of a set *A* if *C* is a superset of *A* (set *A* is a subset of *C*), and *C* is an *S*-set.
Set *C* is a minimum volume *S*-superset of a set *A* if *C* is an *S*-superset of *A* and if volume *V*(*C*) is minimum among all *S*-supersets of *A* (note that, while *C* is not necessarily unique, *V*(*C*) is).
The internal *S*-shape deficiency measure *IShD*(*A*, *S*) of a set *A* is:

$$IShD(A.\ S) = 1 - V(B)/V(A) \qquad (4)$$

where *B* is a maximum volume *S*-subset of *A*.
The external *S*-shape deficiency measure *EShD*(*A*, *S*) of a set *A* is:

$$EShD(A.\ S) = 1 - V(A)/V(C) \qquad (5)$$

where *C* is a minimum volume *S*-superset of *A*.
A completely analogous treatment is applicable to any individual shape property *Q*. With regard to some shape property *Q*, a set *F* is a *Q*-subset of a set *A* if *F* is a subset of *A*, and *F* is a *Q*-set.
Set *F* is a maximum volume *Q*-subset of a set *A* if *F* is a *Q*-subset of *A* and if volume *V*(*F*) is maximum among all *Q*-subsets of *A* (note that, while *F* is not necessarily unique, *V*(*F*) is).
Set *G* is a *Q*-superset of a set *A* if *G* is a superset of *A* (set *A* is a subset of *G*), and if *G* is a *Q*-set.
Set *G* is a minimum volume *Q*-superset of a set *A* if *G* is an *S*-superset of *A* and if volume *V*(*G*) is minimum among all *Q*-supersets of *A* (note that, while *G* is not necessarily unique, *V*(*G*) is).
The internal *Q*-shape property deficiency measure *IShPD*(*A*, *Q*) of a set A is:

$$IShPD(A, Q) = 1 - V(F)/V(A) \qquad (6)$$

where *F* is a maximum volume *O*-subset of *A*.
The external *Q*-shape property deficiency measure *EShPD*(*A*, *Q*) of a set A is:

$$IShPD(A, Q) = 1 - V(A)/V(G) \qquad (7)$$

where *G* is a minimum volume *Q*-superset of *A*.

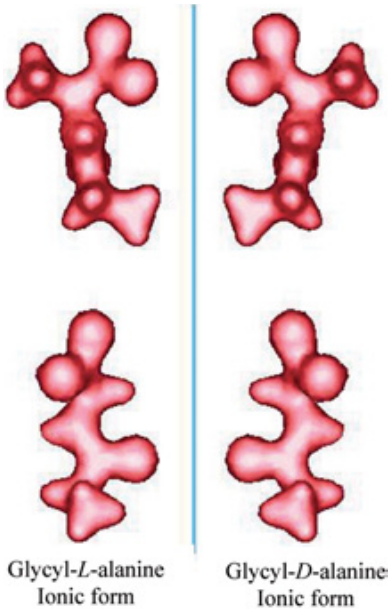


Figure 4 - Two mirror-image forms of the dipeptide of glycyl-alanine, showing the peptide-bond (central region), a structural feature of all proteins.
Proteins do much of the fancy chemistry in our body needed to stay alive. Both forms of this molecule are chiral, having no mirror planes themselves, hence these are typical symmetry-deficient molecules, yet only one of these (left) can contribute to the job in our body, the other one is practically useless.

The connection between shape property deviation measures and symmetry deficiency measures

It is easily seen that symmetry deficiency measures, such as appropriately chosen chirality measures, characterizing the degree of chirality, are a subclass of shape property deficiency measures (*figure 4*).

Since chiral objects cannot possess any mirror plane and any rotation-reflection symmetry axis S(2n), chirality itself can be regarded as deficiency in the above two types of symmetries. The degree of chirality can also be evaluated in terms of the degree of symmetry deficiency with regard to the above two symmetry types.

Each symmetry element of the above two symmetry types can be regarded as a specific shape property. Furthermore, by choosing a new shape property Q' *as the presence of any of the symmetry elements of the above two symmetry types*, the following shape property deficiency measures become actual chirality measures, as given below.

The internal Q'-shape property deficiency measure $IShPD(A, Q')$ of a set A, defined as:

$$IShPD(A, Q') = 1 - V(F')/V(A) \quad (8)$$

where F' is a maximum volume Q'-subset of set A, is in fact an internal chirality measure for set A.

The external Q'-shape property deficiency measure $EShPD(A, Q')$ of a set A, defined as:

$$EShPD(A, Q') = 1 - V(A)/V(G') \quad (9)$$

where G' is a minimum volume Q'-subset of set A, is in fact an external chirality measure for set A.

While chirality is the most widely appreciated symmetry deficiency, especially in chemistry, any other combination of individual symmetry elements can be used to specify some Q shape property, and then the internal and external shape property deficiency measures, $IShPD(A,Q')$ and $EShPD(A,Q)$, can respectively serve as actual measures of symmetry deficiencies.

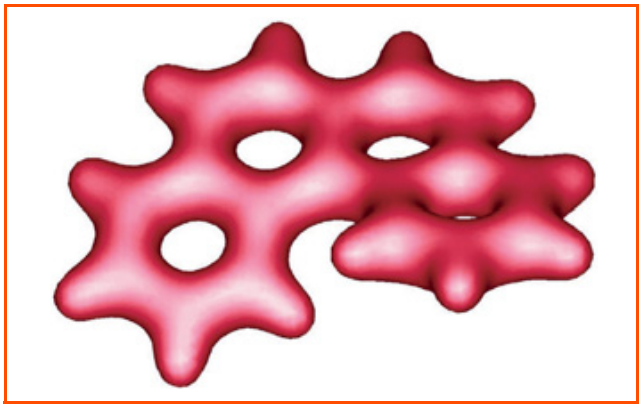


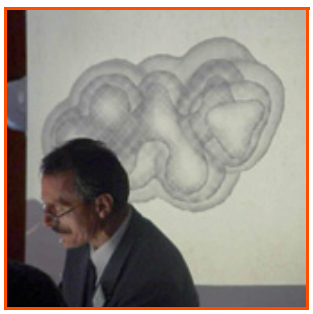
Figure 5 - Image of a benzphenanthrene molecule (electron density contour at density value of 0.1 au) generated in the Scientific Modeling and Simulation Department of Chemistry at the Memorial University of Newfoundland. (Mezey *et al.*, *J. Chem. Inf. Comp. Sci.*, **1996**, *36*, p. 602.
A prototype of the environmentally hazardous, toxic polyaromatic hydrocarbons, has only approximate mirror symmetry, a fact that has an impact on the stability and reactivity, hence, toxicity of this molecule. The degree of symmetry deficiency often correlates with toxicity within molecular families.

Summary

An approach is decribed for the generation of shape deviation measures with the aim of providing numerical measures for molecular shape comparisons, and to interconnect the symmetry deficiency measures, in particular, chirality measures, with the measures of deviations from reference shapes or from individual shape properties.

Such numerical measures are relevant to modern approaches to molecular informatics, and they can provide new tools for toxicological risk assessment, for pharmaceutical drug design (*figure 5*), and more generally for shape-biochemical activity correlations.

Professor Walter Kohn (left), one of the world's leading experts of molecular electron clouds, and the author discuss the role of small molecular fragments in large molecules.



"Molecules on Mezey's mind!"

Acknowledgements

The financial research support of the Natural Sciences and Engineering Research Council (NSERC) of Canada, the Canada Foundation for Innovation, the Canada Research Chairs Program, Memorial University of Newfoundland, Canada, the e-Science RET Project of Hungary, as well as the hospitality of the Institute for Advanced Study, Collegium Budapest, Hungary are gratefully acknowledged.

References

[1] Cayzergues P., Panaye A., Dubois J.-E., *C. R. Acad. Sci. Paris (série C)*, **1980**, *290*, p. 441.
[2] Dubois J.-E., Alliot M.-J., Panaye A., *C. R. Acad. Sci. Paris (série C)*, **1971**, *273*, p. 224.
[3] Dubois J.-E., Cojan-Alliot M.-J., *C. R. Acad. Sci. Paris (série C)*, **1975**, *280*, p. 13.
[4] Dubois J.-E., Panaye A., Cayzergues P., *C. R. Acad. Sci. Paris (série C)*, **1980**, *290*, p. 429.
[5] Dubois J.-E., Panaye A., Cayzergues P., *C. R. Acad. Sci. Paris (série II)*, **1982**, *295*, p. 135.
[6] Dubois J.-E., Panaye A., Cojan-Alliot M.-J., *C. R. Acad. Sci. Paris (série C)*, **1975**, *280*, p. 353.
[7] Panaye A., Dubois J.-E., *The Role of Data in Scientific Progress (Proceedings of the 9th International CODATA Conference, Jerusalem, 24 June 1984)*, Elsevier, North-Holland, **1975**, p. 381.
[8] Kitaigorodski A.I., *Organic Chemical Crystallography*, Chap. 4, Consultants Bureau, New York, **1961**, p. 5.
[9] Mezey P.G., *Symmetries and Properties of Non-Rigid Molecules: A Comprehensive Survey*, J. Maruani, J. Serre (eds), Elsevier, Amsterdam, **1983**, p. 335.
[10] Rassat A., Un critère de classement des systèmes chiraux de points à partir de la distance au sens de Haussdorf, *C. R. Acad. Sci., Paris (Série II)*, **1984**, *299*, p. 53.
[11] Gilat G., *Chem. Phys. Letters*, **1985**, *121*, p. 13.
[12] Mezey P.G., *Structure and Dynamics of Molecular Systems. Vol. I.*, R. Daudel, J-P. Korb, J-P. Lemaistre, J. Maruani (eds), Reidel, Dordrecht, **1985**, p. 41.
[13] Barron L.D., *Chem. Soc. Rev.*, **1986**, *15*, p. 189.
[14] Mezey P.G., *J. Am. Chem. Soc.*, **1986**, *108*, p. 3976.
[15] Mezey P.G., *Theor. Chim. Acta.*, **1986**, *73*, p. 221.
[16] Gilat G., *J. Phys. A. Math. Gen.*, **1989**, *22*, p. L545.
[17] Gilat G., *Found. Phys. Lett.*, **1990**, *3*, p. 189.
[18] Mezey P.G., Maruani J., *Mol. Phys.*, **1990**, *69*, p. 97.
[19] Hel-Or Y., Peleg S., Avnir D., *Langmuir*, **1990**, *6*, p. 1691.
[20] Meyer A.Y., Avnir D., *Struct. Chem.*, **1991**, *2*, p. 475.
[21] Auf der Heyde T.P.E., Buda A.B., Mislow K., *J. Math. Chem.*, **1991**, *6*, p. 255.
[22] Buda A.B., Mislow K., *J. Mol. Struct. (Theochem)*, **1991**, *232*, p. 1.
[23] *New Developments in Molecular Chirality*, P.G. Mezey (ed), Kluwer, Dordrecht, **1991**.
[24] Mezey P.G., *J. Math. Chem.*, **1991**, *7*, p. 39.
[25] Mezey P.G., *Canad. J. Chem.*, **1992**, *70*, p. 343.
[26] Mezey P.G., Maruani J., *Int. J. Quantum Chem.*, **1993**, *45*, p. 177.
[27] Mezey P.G., *Shape in Chemistry: An Introduction to Molecular Shape and Topology*, VCH Publishers, New York, **1993**.
[28] Mezey P.G., *Internat. Rev. Phys. Chem.*, **1997**, *16*, p. 361.
[29] Mezey P.G., *Chirality*, **1998**, *10*, p. 173.
[30] Mezey P.G., *J. Mol. Struct. Theochem*, **1998**, *455*, p. 183.
[31] Mezey P.G., *J. Math. Chem.*, **1998**, *23*, p. 65.
[32]Mezey P.G., *Encyclopedia of Computational Chemistry*, P.v.R. Schleyer, N.L. Allinger, T. Clark, J. Gasteiger, P.A. Kollman, H.F. Schaefer III, P.R. Schreiner (eds), John Wiley & Sons, Chichester, **1999**, *4*, p. 2582
[33] Mezey P.G., *Advances in BioChirality*, G. Palyi, C. Zucchi, L. Caglioti (eds), Elsevier, Amsterdam, **1999**, p. 35.
[34] Mezey P.G., *Progress in Biological Chirality*, Chap. 17, G. Palyi, C. Zucchi, L. Caglioti (eds), Elsevier, Oxford, **2004**, p. 209.
[35] Mezey P.G., *Computational Medicinal Chemistry and Drug Discovery*, Bultinck P., Tollenaere J.P., De Winter H., Langenaeker W. (eds), Marcel Decker, New York, **2004**, p. 345.
[36] Mezey P.G., *J. Internat. Soc. Interdisc. Study Symmetry*, **2004**, p. 166.
[37] Mezey P.G., *Journal of Computational Methods in Sciences and Engineering (JCMSE)*, **2005**, *5*, p. 109.
[38] Mezey P.G., *Molec. Phys.*, **2006**, *104*, p. 723; Corrigendum, *Molec. Phys.*, **2006**, *104*, p. 2575 (special volume dedicated to Prof. Michael A. Robb).
[39] Wang L., Wang L., Arimoto S., Mezey P.G., *J. Math. Chem.*, **2006**, *40*, p. 145.

Paul G. Mezey
is Professor of Chemistry, Canada Research Chair in Scientific Modeling and Simulation, Memorial University of Newfoundland*.

* Department of Chemistry and Department of Physics and Physical Oceanography, Memorial University of Newfoundland, St. John's, NL, Canada A1B 3X7.
E-mail: paul.mezey@gmail.com