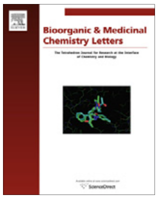
Bioorganic & Medicinal Chemistry Letters 20 (2010) 5887–5890

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry Letters

journal homepage: www.elsevier.com/locate/bmcl



Molecular alignment using multipole moments

Loris Moretti *, W. Graham Richards

Physical and Theoretical Chemistry Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QZ, UK

ARTICLE INFO

Article history:
Received 10 June 2010
Revised 23 July 2010
Accepted 25 July 2010
Available online 30 July 2010

Keywords:
Multipole moments
Structural alignment
Electrostatic similarity
Virtual screening

ABSTRACT

Molecules can be aligned on the basis of their computed electrical multipole moments. This description of molecular electrostatics is proposed and evaluated for similarity applications. It accurately models the charge distribution in compounds with medicinal chemistry interest where electrostatics is known to play an important role in their interaction with the target.

Within the field of drug discovery the in silico determination of similar and dissimilar compounds is a known and significant issue.[1–4] The key factor to define the similarity of molecules is the characteristic aspect or property taken into consideration. Many features can be used, alone or in combination, and as a result divergent solutions are possible which will be valuable in different ways depending on the nature of the studied case.

Here, we focus on the use of electrostatics, which reflect electron distributions, as the basis upon which to explore molecular similarity. Several theories and methodologies exploiting this property have been developed in molecular modelling[5] for alignment procedures, similarity evaluations and virtual screening (VS).

The most widely applied description of electrostatics is the Molecular Electrostatic Potential (MEP) computed either from the electron density,[6–9] from its distributed multipole analysis,[10–12] or from atomic point charges.[13–16] A common alternative to this representation is the so-called Molecular Interaction Potential (MIP). This is a variation of MEP where the interaction between a charged probe and the electrostatics of a molecule is calculated.[17–20]

In general, some of these methodologies exploit quantum chemistry, known to be time consuming, while others use classical models which are faster but still require the computation of the explicit potentials. Although this calculation can be relatively quick in certain implementations, multiplying it millions of times, the typical size of a current database for VS, is computationally demanding. Thus, in our view the need for an electrostatic representation fast enough and relatively accurate has not yet been met.

In this Letter, an approach to model molecular electrostatics for similarity determination is proposed. Each compound can be described by the Cartesian electrical multipole moments, which quantify the symmetry of a charge distribution. Multipoles are the coefficients of the MEP expansion and are computed before solving the potential. It is a fast way to access the electrostatics, suitable for molecular similarity studies on large databases within the field of drug discovery. The reliability and accuracy of this model has been addressed by its implementation for structural alignment of molecules. In the past, several similarity indices based on electrostatics have been used to guide the superposition of small molecules. The majority of the methods evaluates the potential at specific Cartesian references. Initially, the alignment was obtained with the maximisation of sets of MEPs calculated on the nodes of a three-dimensional grid around the molecules[6] or a subset of these.[16] But the breakthrough in speed and accuracy was achieved by the substitution of the grid-based positions with a series of Gaussian functions.[21] MIP has also been used for molecules superposition. As with MEP, one of the first applications exploited the nodes of a rectilinear grid around molecules as reference locations.[17] Then again only a subset of the grid was employed, those points with negative energy values which represent the attractive interactions.[18–20] The last model used is the direct comparison of partial charges to select the best overlay of two compounds.[25–27]

Here, a procedure for molecule superposition based on multipole moments is described. In this protocol, dipole and quadrupole moments are used to drive the geometrical transformations to overlay one molecule onto another. The method is also tested with known cases from the area of medicinal chemistry.

Multipoles are the expression of certain symmetries and therefore some geometrical information is embedded in these entities. In addition, contrary to the spherical multipole moments, which

---

* Corresponding author. Tel.: +44 1865262028; fax: +44 1865243072.
E-mail addresses: loris.moretti.dc@gmail.com (L. Moretti), graham.richards@chem.ox.ac.uk (W. Graham Richards).

0960-894X/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2010.07.107

use internal coordinates, Cartesian multipoles are based on global reference that permits geometrical transformations of molecules on a common basis. These observations led to the idea of employing the multipoles within a protocol for molecular alignment in order to ascertain the use of this electrostatic representation for drug discovery. Thus, a superposition protocol was designed and implemented based just on multipole moments. The same been suggested by other authors, but without any details of implementation.[22,23]

The alignment procedure here proposed consists of three main steps: (i) the molecule is centred based on its charge-weighted centroid, which is reproducible and consistent for electrostatically equivalent molecules, (ii) its dipole vector is aligned onto the *Z*-axis and (iii) the minimum of the quadrupole difference between the reference and the molecule to be aligned is selected during the rotation of the molecule around the *Z*-axis. Details about the theoretical aspects of multipoles and protocol implementation can be found in Supplementary data.

To note that multipoles are the deviation from specific symmetries, two electrostatically similar molecules have similar components but similar components do not necessary imply similar electrostatics. Therefore, it is probable that with very different molecules such alignment will be less meaningful. Another drawback is that in certain cases there will be twofold degeneracy by using quadrupoles to finalise the alignment. This behaviour will arise when the molecule under investigation has specific symmetry that is present, in our view, in a small proportion of the chemical space used for pharmaceutical purposes.

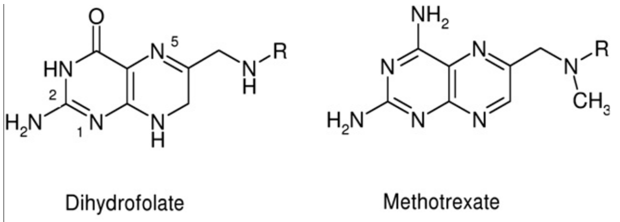
The proposed alignment falls into the class of rigid-body optimisation protocols, where the molecular flexibility is not taken into account. Therefore, the execution of the molecular overlay is fast. In addition, the final pose is obtained in one cycle in contrast with other methods of the same kind which need to iterate over a number of starting positions. In terms of speed, this procedure superposes approximately 16 drug-size molecules per second compared to 8–10 compounds a second of the well known ROCS, or Rapid Overlay of Chemical Structures.[28]

To validate this method known case studies of medicinal chemistry were selected. The first is a typical test for structural alignment methods:[6,19,23–25,28] the dihydrofolate, the natural substrate of dihydrofolate reductase (DHFR) and the methotrexate, an inhibitor. While the second case is the endogenous ligand of ionotropic glutamate receptors (iGluRs), the glutamate, and two agonists: AMPA and kainate.

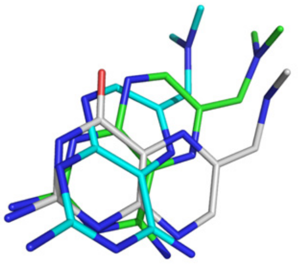
The binding of all these molecules with their respective target is experimentally known through solved crystallographic structures. Therefore, the alignments performed by the multipole method can be directly compared.

The molecules of dihydrofolate (DHF) and methotrexate (MTX) are binders of the DHFR enzyme and share a common scaffold, the pteridine ring (see Fig. 1). This case is popular in medicinal chemistry because the pteridine moieties of the two ligands, although interacting in the same protein region, present a relative rotation of 180° along the axis passing by the carbon in position 2 and the nitrogen in position 5.[6,29] Thus this moiety interacts with the protein in 2 different orientations depending on the ligand. The different substitutes and protonation states[29,30] of the pteridine ring are likely to vary the hydrogen-bonding pattern. Therefore, the electrostatics is supposed to play a central role for the different binding mode of the two ligands.

The crystallographic structures of the complexes ligand–protein were found at the Protein Data Bank (PDB)[31] and prepared.[32] Only the pteridine rings with substituents, as represented in Figure 1, were used and the ligand protonation states were selected according to publications.[29,30] The main difference is the protonation of the nitrogen at position 1 of the MTX. The multipole alignment of methotrexate onto dihydrofolate was performed, the result is depicted in Figure 1b. The final pose is compared to the structural alignment of the receptor–ligand complexes.[35]



(a) The DHFR ligands



(b) Ligands alignments

Figure 1. The dihydrofolate reductase ligands. (a) The two-dimensional structures of the ligands under investigation. 'R', in both compounds, represents the *p*-aminobenzoyl-L-glutamate moiety. (b) Superpositions of pteridine core of methotrexate to dihydrofolate based on the crystallographic structures and on electrostatic alignment. Molecules are represented with oxygen atoms in red and nitrogen in blue. Carbons are grey for dihydrofolate, light-green for methotrexate from the crystallographic alignment and cyan for the same molecule from the electrostatic alignment.

The outcomes of the two overlays are visually quite close with same directionality of the methotrexate structures. The difference between the resulting position of the inhibitor was quantify to be 1.44 Å in terms of root mean square deviation (RMSD).

Multipole components values of the final positions of the two molecules are given in Table 1. The monopole is different with a total charge of 1 for the MTX and 0 for the DHF due to the chosen protonation states. Interestingly, the other multipoles are close in value beside few components of the quadrupole, where one has a different sign (the XZ-component).

Certain conditions of the alignment experiment were varied to test further the protocol.[37] Results were affected by different charge assignments but not from different starting positions.

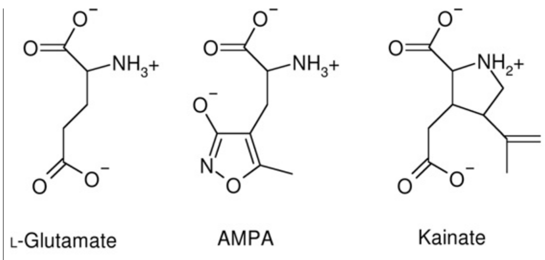
The molecules glutamate, AMPA and kainate (represented in Fig. 2a) are ligands of iGluRs interacting with the target at the level of the ligand-binding domain.[41] This protein portion is composed by two domains, S1 and S2, and it fluctuates dynamically between two limiting states: open and closed. The main difference between these two structures is the relative degree of closure of the two lobes.

The conformations adopted by the ligands within the receptors are revealed by the crystallographic complexes.[40] These compounds act as agonists for this receptor family stabilising the

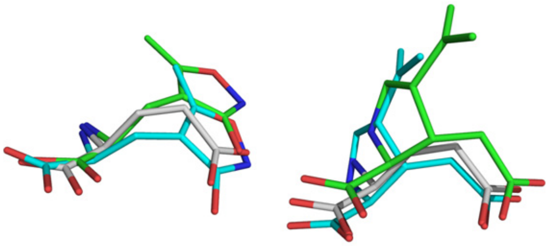
Table 1
Multipoles values of DHFR ligands depicted in Figure 1 are reported

	M	D	Q					
		z	*xx*	*xy*	*xz*	*yy*	*yz*	*zz*
Dihydrofolate	0	1.07	1.01	0.83	−0.16	5.36	2.86	2.56
Methotrexate	1	0.71	1.66	0.84	2.60	8.39	1.78	12.76

The monopole (M), the Z-component of the dipole (D) and the quadrupole (Q) components, are present.



(a) The iGluRs ligands



(b) AMPA (c) Kainate

Figure 2. The ionotropic glutamate receptor–ligands. (a) The two-dimensional structures of the ligands under investigation. (b and c) Superpositions of kainate and AMPA to glutamate, based on the crystallographic structures and on electrostatic alignment. Molecules are represented with oxygen atoms in red and nitrogen in blue. Carbons are grey for glutamate, light-green for AMPA and kainate from the crystallographic alignment and cyan for the same molecules from the electrostatic alignment.

closed state with similar closures of the S1–S2 domain. In more detail, the anchor points of the ligands to the target protein are mainly three polar interactions: a salt bridge, a three-hydrogen-bonding centre and an interaction between a negative charge of the ligand and the positive end of a macro dipole formed by a helix. Thus, the similarity of these electrostatic features makes these compounds another good test case for the alignment procedure based on the multipoles.

Ligands were prepared[32] with protonation states depicted in Figure 2a, in agreement with the original publication[41] and crystallographic information. AMPA and kainate were electrostatically aligned using glutamate as the reference structure. The relative positions and orientations of the compounds obtained are depicted in Figure 2 panel b and c. As comparison the structural alignments of the three receptor–ligand complexes[35] is also reported.

The overlays of these ligands from the alignment of the crystallographic structures and the one based on the multipoles are again visually quite close. Same directionality of the structures with the close proximity of correspondent polar groups. The difference in RMSD between the structures coming from the two alignments is 1.64 Å for AMPA and 1.77 Å for kainate.

Confirmation of the electrostatic similarity of these molecules is given by the multipole values of the aligned structures reported in Table 2. As expected the total charge is −1 in all three cases. The dipoles (the Z-component) and the 6 quadrupole components are of the same sign and of similar magnitude. Thus, the use of multipole moments for aligning these structures produced a good performance highlighting the major role of the electrostatics in this ligand-binding.

Table 2
Multipoles values of iGluR ligands depicted in Figure 2 are reported

	M	D	Q					
		z	xx	xy	xz	yy	yz	zz
Glutamate	−1	3.99	−6.18	6.63	−3.45	−5.19	3.23	−3.08
AMPA	−1	3.62	−9.24	8.73	−4.60	−6.34	3.27	−4.46
Kainate	−1	4.27	−5.63	7.07	−3.62	−5.14	2.71	−3.45

The monopole (M), the Z-component of the dipole (D) and the quadrupole (Q) components, are present.

Certain conditions of the alignment experiment were varied.[37] In all the cases the alignments seen in Figure 2 were reproduced.

Molecular similarity was proved to be a very useful tool in computational approaches to drug discovery, but almost invariably restricted to similarity of shape. Here we have introduced a simple way of incorporating the electrostatic information. Cartesian electrical multipole moments are an effective representation of the charge distribution in molecules. They are easy to implement and fast to calculate, therefore suitable for dealing with large numbers of molecules as in virtual screening.

To explore the accuracy of this model a protocol for molecular superposition was designed and successfully tested on known cases from medicinal chemistry. By means of this procedure, ligands of dihydrofolate reductase and of ionotropic glutamate receptors were aligned in a manner consistent with the experimental data.

Although the methodology has to be further validated with real applications to drug discovery, these results are considered encouraging for future developments. Being aware of the limitations, as for example with very different ligands, this model can be applied to study those cases where the electrostatics of the binding is crucial. On the other hand, the approach can also be used as complement to other methods/theories, whereas it might be a convenient way of including electrostatics, together with other molecular features, in fast in silico investigations. Therefore, several applications to drug discovery are possible by the exploitation of this electrostatic model.

Acknowledgements

This study was supported by a grant of the 6th Framework Program of the European Union (DeZnIT, Project Number 037303). The authors would like to thank Paul Finn, Garrett Morris, Richard Cooper, Stuart Armstrong and Marco Nardecchia for the fruitful discussions.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmcl.2010.07.107.

References and notes

1. Willett, P.; Barnard, J. M.; Downs, G. M. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 983.
2. Bender, A.; Glen, R. C. *Org. Biomol. Chem.* **2004**, *2*, 3204.
3. Maldonado, A. G.; Doucet, J. P.; Petitjean, M.; Fan, B. T. *Mol. Divers.* **2006**, *10*, 39.
4. Eckert, H.; Bajorath, J. *Drug Discovery Today* **2007**, *12*, 225.
5. Naray-Szabo, G.; Ferenczy, G. G. *Chem. Rev.* **1995**, *95*, 829.
6. Manaut, F.; Sanz, F.; José, J.; Milesi, M. *J. Comput. Aided Mol. Des.* **1991**, *5*, 371.
7. Sanz, F.; Manaut, F.; Rodríguez, J.; Lozoya, E.; López-de Briñas, E. *J. Comput. Aided Mol. Des.* **1993**, *7*, 337.
8. Petke, J. D. *J. Comput. Chem.* **1993**, *14*, 928.
9. Marín, R. M.; Aguirre, N. F.; Daza, E. E. *J. Chem. Inf. Model.* **2008**, *48*, 109.
10. Wenden, E. M.; Price, S. L.; Apaya, R. P.; IJzerman, A. P.; Soudijn, W. *J. Comput. Aided Mol. Des.* **1995**, *9*, 44.
11. Apaya, R. P.; Lucchese, B.; Price, S. L.; Vinter, J. G. *J. Comput. Aided Mol. Des.* **1995**, *9*, 33.
12. Burgess, E. M.; Ruell, J. A.; Zalkow, L. H.; Haugwitz, R. D. *J. Med. Chem.* **1995**, *38*, 1635.
13. Hodgkin, E. E.; Richards, W. G. *Int. J. Quantum Chem.* **1987**, *32*, 105.
14. Wild, D. J.; Willett, P. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 159.
15. Xian, B.; Li, T.; Sun, G.; Cao, T. *J. Mol. Struct.: Theochem.* **2004**, *674*, 87.
16. Vainio, M. J.; Puranen, J. S.; Johnson, M. S. *J. Chem. Inf. Model.* **2009**, *49*, 492.
17. Cocchi, M.; Benedetti, P. G. D. *J. Mol. Model.* **1998**, *4*, 113.
18. de Caceres, M.; Villa, J.; Lozano, J. J.; Sanz, F. *Bioinformatics* **2000**, *16*, 568.
19. Melani, F.; Gratteri, P.; Adamo, M.; Bonaccini, C. *J. Med. Chem.* **2003**, *46*, 1359.
20. Cheeseright, T.; Mackey, M.; Rose, S.; Vinter, A. *J. Chem. Inf. Model.* **2006**, *46*, 665.

21. Good, A. C.; Hodgkin, E. E.; Richards, W. G. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 188.
22. McMahon, A. J.; King, P. M. *J. Comput. Chem.* **1997**, *18*, 151.
23. Parretti, M. F.; Kroemer, R. T.; Rothman, J. H.; Richards, W. G. *J. Comput. Chem.* **1997**, *18*, 1344.
24. Lemmen, C.; Lengauer, T. *J. Comput. Aided Mol. Des.* **1997**, *11*, 357.
25. Kearsley, S. K.; Smith, G. M. *Tetrahedron Comput. Methodol.* **1990**, *3*, 615.
26. Lemmen, C.; Hiller, C.; Lengauer, T. *J. Comput. Aided Mol. Des.* **1998**, *12*, 491.
27. Klebe, G.; Mietzner, T.; Weber, F. *J. Comput. Aided Mol. Des.* **1999**, *13*, 35.
28. Rocs, http://www.eyesopen.com/products/applications/rocs.html.
29. Bolin, J. T.; Filman, D. J.; Matthews, D. A.; Hamlin, R. C.; Kraut, J. *J. Biol. Chem.* **1982**, *257*, 13650.
30. London, R. E.; Howell, E. E.; Warren, M. S.; Kraut, J.; Blakley, R. L. *Biochemistry* **1986**, *25*, 7229.
31. Crystallographic structures with PDB-code 7DFR (dihydrofolate) and 3DFR (methotrexate).
32. The atomic coordinates of the ligands were extracted from the crystallographic complexes and partial charges assigned within MOE software package[33] based on the semi-empirical method AM1.[34]
33. Moe, Chemical Computing Group, Inc.: Montreal, Quebec, Canada. http://www.chemcomp.com.
34. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.
35. The structural alignment of the complexes was carried out with BODIL software package.[36]
36. Lehtonen, J. V.; Still, D. J.; Rantanen, V. V.; Ekholm, J.; Björklund, D.; Iftikhar, Z.; Huhtala, M.; Repo, S.; Jussila, A.; Jaakkola, J.; Pentikäinen, O.; Nyrönen, T.; Salminen, T.; Gyllenberg, M.; Johnson, M. S. *J. Comput. Aided Mol. Des.* **2004**, *18*, 401.
37. Several starting positions for molecules were randomly selected as well as other partial charge schemes were adopted. Beside the semi-empirical AM1 method, the MMFF94[38] and Gasteiger[39] atomic charges were also used and assigned within MOE.
38. Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 490.
39. Gasteiger, J.; Marsili, M. *Tetrahedron* **1980**, *36*, 3219.
40. PDB-codes: 1FTL (glutamate), 1FTM (AMPA) and 1FTK (kainate).[41]
41. Armstrong, N.; Gouaux, E. *Neuron* **2000**, *28*, 165.