SCIENTIA gerundensis, 21: 165-173 (1995)

FIRST-ORDER MOLECULAR DESCRIPTORS FOR MOLECULAR STERIC SIMILARITY

Jordi Mestres, Miquel Solà and Ramon Carbó

Institut de Química Computacional, Universitat de Girona, 17071 Girona.

RESUM

En aquest treball s'analitza la contribució estèrica de les molècules a les seves propietats químiques i físiques, mitjançant l'avaluació del seu volum i de la seva mesura de semblança, a partir d'ara definits com a descriptors moleculars de primer ordre. La diferència entre aquests dos conceptes ha estat aclarida: mentre que el volum és la magnitud de l'espai que ocupa la molècula com a entitat global, la mesura de semblança ens dóna una idea de com està distribuïda la densitat electrònica al llarg d'aquest volum, i reflecteix més les diferències locals existents. L'ús de diverses aproximacions per a l'obtenció d'ambdós valors ha estat analitzat sobre diferents classes d'isòmers.

RESUMEN

En este trabajo se analiza la contribución estérica de las moléculas a sus propiedades químicas y físicas, mediante la evaluación de su volumen y de su medida de semejanza, a partir de ahora definidos como descriptores moleculares de primer orden. La diferencia entre estos dos conceptos ha sido clarificada: mientras que el volumen es la magnitud del espacio que ocupa la molécula como entidad global, la medida de semejanza nos da una idea de cómo está distribuida la densidad electrónica a lo largo de este volumen, reflejando más las diferencias locales existentes. El uso de distintas aproximaciones para la obtención de estos dos valores ha sido analizado sobre diferentes clases de isómeros.

ABSTRACT

In this work the steric contribution of molecules to their chemical and physical properties is analized in terms of their volume and their similarity measure, hereafter called first-order molecular descriptors. The difference between these two concepts has been clarified: while the volume is the magnitude of space occupied by the molecule as a global entity, the similarity measure give us an idea of how the electronic density is distributed along this volume, reflecting more the existent local differences. The use of several approximations to the obtention of these values has been analized on different types of isomers.

Keywords: Molecular Volume, Molecular Similarity Measure, Atomic Similarity Measure, Molecular Descriptors, Molecular Steric Similarity.

INTRODUCTION

Application of molecular similarity strategies as one more step for drug discovery purposes is becoming, nowadays, a common procedure to take into account in any pharmaceutical laboratory (1-3). However, the complexity of the molecular systems

makes similarity studies very difficult from a theoretical point of view. Is for that reason that molecular similarity studies have been often limited to a topological level, giving rise to some widespread QSAR methods (4,5). In a general way, these methods are mostly based on molecular topological descriptors, which means that they are only dealing with 2D parameters that describe the molecule by itself.

Another way to work within a molecular similarity framework is to describe molecules with molecular fields or surfaces (namely, electronic density, electrostatic potential, hidrophobic surface,...) obtained from either an empirical or a quantum mechanical calculation. This change of work philosophy needs of a 3D molecular structure to deal with. As explained above, due to the complexity of the molecules under study, the molecular structure optimization is often reduced to a molecular mechanics level of calculation. After that, the quantum mechanical description of molecules (that is, the obtention of a wave function from which we will extract the molecular surfaces necessary to perform our analysis) is usually done at a semiempirical level of calculation, the use of the more accurate *ab initio* methods being practically non-viable due to computational limits.

Once the wave function of the complete set of molecules under study has been obtained, it is easy to define some molecular descriptors in order to have a first insight into the differences between molecules. One can call first-order molecular descriptors those which depend only on the molecule itself. One step further implies the obtention of *nth*-order molecular descriptors as a result of a molecular similarity matching between n molecules. For example, a second-order molecular descriptor will be a vector that quantifies the similarity of one molecule with respect to each one of the other molecules under study; a third-order molecular descriptor will be a matrix that quantifies similarities between one molecule and two other molecules of the molecular set, and so on. As can be seen, the bottleneck of this way of work will be, precisely, the molecular similarity matching process in an exact form. And this is the reason why several approximations for the matching process has begun to appear. For instance, a fitting of the electronic density has been recently proposed (6) to reduce the enormous computational cost of the maximization process of the molecular matching, and applied then to the study of several current chemical problems (7,8).

As a first stage, this work presents a comparative study of some approximations that can be used to compute the first-order molecular descriptors and which can be more adequate to estimate the steric contribution of molecules to their chemical and physical properties, namely, the molecular volume and the self-similarity measure.

METHODOLOGY

All molecular geometries have been fully optimized at the RHF/3-21G level of theory, using the GAUSSIAN 92 program (9). The molecular volume has been calculated through a Monte-Carlo integration with the electronic analysis program ELECTRA (10). Each volume has been calculated 10 times per molecule, an average value having been taken as the final value. The integration box has been defined by adding 4 au to the limiting positive and negative atomic coordinates. We have used a density of 100 points per au^3 and two different surface approximations: an

atomic van der Waals radii cutoff (VOL_{vdw}) and an electronic density cutoff of 0.005 e/au^3 (VOL_{den}). On the other hand, the molecular similarity measure (MSM) give us an idea of how the electronic density is distributed along the molecule (11). The exact value has been obtained both at the AM1 semiempirical level (MSM_{am1}) and at the RHF/3-21G level (MSM_{scf}). Three possible approximations to the exact MSM_{scf} value have been suggested: i) the use of a fitted density to compute the similarity measure integrals (MSM_{fit}); ii) the use of a unique gaussian function to describe the atoms in the molecule (MSM_{gun}); and iii) the use of a sum of atomic similarity measures obtained from a single-zeta Slater type functions (SASM) (11). The exact and density fitted measures (MSM_{am1}, MSM_{scf} and MSM_{fit}) have been calculated by means of the MESSEM program (12).

RESULTS AND DISCUSSION

In order to perform a comparative study of the different approximations proposed, we have taken several problematic cases where these approximations may fail. If the set of molecules were structurally very different, maybe we could not be able to see where the differences come from. This has been the reason we have chosen couples of molecules that reflect the different possible isomerisms, namely, conformational, configurational and constitutional (although the last one could not be considered a type of isomerism strictly speaking). Figure 1 depicts the two sets of molecules representative of a conformational isomerism: staggered and eclipsed ethane and cis and trans 2-butene; in Figure 2 we present the set of molecules representative of a configurational isomerism: 2,3-butanediol in its ss and sr forms. Their mirror images (rr and rs) have not been taken into account as they have the same volume and similarity measure than ss and sr; finally, in Figure 3 we have depicted two examples of couple of molecules that the only thing that they have in common is that they own exactly the same constituent atoms. These two sets of molecules are, from one side, acetamide and acetaldoxime and, from the other side, furan and 2-butinal.

The overall results are shown in Table I, where the values obtained for the different approximations to the volume and molecular similarity measures have been collected, together with the corresponding electronic energy at the RHF/3-21G level of calculation.

One can emphasize here the clear difference between the concepts of **molecular volume** and **molecular self-similarity measure**: while the former can be related to the idea of *how many* space is being occupied, the latter is referred to *how* the electronic charge density occupies this given space. In other words, while molecular volume describes the molecule as a whole, regardless of its particular atomic constitution, the molecular self-similarity measure is able to distinguish local molecular differences from a charge density concentration point of view. These are the reasons why we have chosen these two concepts as two different first-order steric molecular descriptors.

Looking at the volume results obtained (Table I), it can be easily seen that there is a good correspondence between VOL_{vdw} and VOL_{den}. However, it can be noticed that VOL_{vdw} is always larger than VOL_{den}: this only reflects the fact that the density

cutoff used for the computation of VOL_{den} (0.005 e/au3) lies always inside the atomic van der Waals envelope. If we take a density cutoff of 0.001 e/au3 then VOL_{den} will become always larger than VOL_{vdw}, as the VOL_{den} surface will extend beyond the van der Waals boundaries. All in all, although the volume values presented in Table I can not be taken as the exact ones (they have been obtained from a Monte-Carlo integration) it can be thought that are a good enough approximation to show the correct trends. As expected, volume values for the conformational and configurational isomers are very close, the larger being always those isomers sterically not favored (eclipsed-ethane, cis-2-butene and (rr,ss)-2,3-butanediol), this fact being also reflected by the corresponding electronic energy.

Taking now a look to the results obtained from the different molecular similarity approximations, one can see in some cases a good correspondence between the volume value and the MSM value. Looking first at the MSM_{scf} column, it is shown that for the conformational and configurational isomers, the smaller the volume, the larger the MSM_{scf} value. This is because the MSM value give us an idea of how the electronic density is distributed along the volume. In the conformational and configurational cases under study, one has always the same atoms with the same kind of bonds, thus: if the volume diminishes the MSM_{scf} value increases, reflecting the fact that the same electronic density is being located in a more reduced volume. An explanation for the diferences in the MSM_{scf} values of the called constitutional isomers comes from looking at the structural nature of the molecules themselves (Figure 3). Differences between acetamide and aldoxime can be explained by comparing the more important bonding changes occured from one structure to the other: we have to compare the overlapping of the C=O and C-N bonds in acetamide with that of the N-O and C=N bonds in acetaldoxime. As a result, the valence electronic overlapping in acetaldoxime will be more important than in acetamide and this will be reflected in a large MSM_{scf} value. On the other hand, the fact that the MSM_{scf} value for furan is smaller than that of 2-butinal can be easily explained by the electronic density spreading over the ring in furan.

It is also very interesting to compare the MSM_{scf} value with the electronic energy value. Comparing the different couples of molecules, it is shown that MSM_{scf} and energy have the same trend: the larger the electronic energy (in absolute values), the larger the MSM_{scf} value is. This is a logical consequence of increasing the electronic charge density. However, when comparing two different isomers of the same electronic charge density, the MSM_{scf} does not always correlate well with the relative stability of the molecules. Two cases can be clearly distinguished: i) if the nature of the bonds remains the same (as it is the case of conformational and configurational isomers), the larger the MSM_{scf} value, the more stable the isomer is; ii) if the nature of the bonds is completely different (as it is the case of constitutional isomers) no relationship can be predicted between the MSM_{scf} value and the electronic energy. From the overall results of Table I it can be concluded that the relative stability of two isomers will also depend on the corresponding molecular volume: in the cases studied, the smaller the volume that contains the same electronic density, the more stable the isomer is.

The exact MSM results obtained using the AM1 semiempirical level (MSM_{am1}) must be taken with some caution. First of all, we have to keep in mind that the values obtained from this way will only reflect the valence electronic overlapping,

as this is an electronic coreless method. For the conformational and configurational isomers, the MSM_{aml} results seem not bad. They show almost the correct trend between isomers of the same couple of molecules and between those corresponding to different couples of molecules. However, a suspicious result is obtained when comparing values between the couples acetamide/acetaldoxime and furan/2-butinal: while from the exact MSM_{scf} value we obtained that the MSM should increase (ca. 194 to ca. 205), it diminishes in MSM_{aml}.

Following, we present the results obtained from the three different approximations used to the MSM value:

i) The strategy of fitting the electronic density to compute the MSM (MSM_{fit}) in order to speed up its calculation during the matching procedure, give rise to excellent results as shown in earlier works (6-8) . Both MSM value and trend are correctly obtained from this approximation.

ii) The use of the simple gaussian function description of the atoms to obtain the MSM value (MSM_{gau}) give also satisfactory results. The fact that a unique gaussian function is being used is the reason why its results reflect more a valence electronic overlapping than a total electronic overlapping. This also explain the fact that the MSM_{gau} value for acetamide is larger than that of acetaldoxime and that the MSM_{gau} value for furan is larger than that of 2-butinal. As can be seen, comparing a couple of molecules belonging to the same isomeric form, the larger the volume, the smaller the MSM_{gau} value, but a good trend is shown between different couples of molecules corresponding to different isomeric forms. Even more, it can be noticed that multiplying the obtained value by a factor of 10 one obtains a good approximation of the exact MSM value using such a simple approximation, at a high gain in computational cost.

iii) The idea of using a sum of atomic similarity measures (SASM) obtained from an atomic calculation using Slater type functions to obtain the MSM seems to be suitable from our results. Because of the use of Slater type functions for the correct atomic description, the value obtained will always be an upper bound of the exact MSM value for all molecules with the same constituent atoms. However, although this approximation can be of great help to obtain MSM values for molecules with clear different structures (which is generally the case) one of its main problems is that it can not distinguish between any of the possible isomeric forms. To solve this problem it is proposed to construct a data base of atomic interactions between atoms at different distances depending on the atomic nature (C_{sp3}-H, C_{sp3}-C_{sp3}, C_{sp2}=O, ...) as done in the molecular mechanics framework.

As a final remark to clearly distinguish again between the two concepts of molecular volume and similarity measure, we can compare the results obtained for trans-2-butene and 2-butinal. As shown in Table I, both molecules have a close value of their molecular volume (477.28 and 460.22 au^3, respectively), trans-2-butene even being slightly larger than 2-butinal. However, regarding their MSM_{scf} values it can be seen that their electronic charge density distributions are completely different (125.3169 and 205.2052, respectively), reflecting the fact that 2-butinal concentrates much more electronic density than trans-2-butene in practically the same molecular volume, due to the presence of the oxygen atom. Thus, it is interesting to emphasize the fact that while molecular volume is a magnitude that describes the

molecule in a global way, the molecular similarity measure reflects the local differences present in such a volume, in terms of electronic density concentration.

CONCLUSIONS

The conceptual difference between molecular volume and similarity measure, hereafter taken as first-order steric molecular descriptors, has been clarified. Moreover, it has been shown that the use of several approximations to the exact value of these magnitudes is a valid way to reduce computational costs and, as a consequence, deal with large molecules. More research in this direction and in the development of higher order molecular descriptors is underway in our laboratory.

ACKNOWLEDGEMENTS

We gratefully acknowledge Dr. D.C. Rohrer for introduce us in the single gaussian approach and Dr. J.M. Oliva for providing us the values of the atomic similarity measures used in this work.

Bibliography

1. RICHARDS, W.G., "Quantum Pharmacology", Butterworths, London, 1983.
2. CONNOLLY, Y.C., KUTTER, E., and AUSTEL, V., Ed., "Modern Drug Research", Marcel Dekker Inc., New York, 1989.
3. KROGSGAARD-LARSEN, P., and BUNDGAARD, H., Eds., "A Textbook of Drug Design and Development", Harwood academic publishers, Chur, 1992.
4. FRANKE, R., "Theoretical Drug Design Methods", Elsevier, Amsterdam, 1984.
5. CONNOLLY, Y.C., "Quantitative Drug Design", Marcel Dekker Inc., New York, 1989.
6. MESTRES, J., SOLÀ, M., DURAN, M., and CARBÓ, R., J. Comp. Chem. *1994*, 15, 1113-1120.
7. SOLÀ, M., MESTRES, J., CARBÓ, R., and DURAN, M., J. Am. Chem. Soc. *1994*, 116, 5909-5915.
8. SOLÀ, M., MESTRES, J., DURAN, M., and CARBÓ, R., J. Chem. Inf. Comp. Sci. *1994*, 34, 1047-1053.
9. GAUSSIAN 92, Revision A, Frisch, M.J., Trucks, G.W., Head-Gordon, M., Gill, P.M.W., Wong, M.W., Foresman, J.B., Johnson, B.G., Schlegel, H.B., Robb, M.A., Replogle, E.S., Gomperts, R., Andres, J.L., Raghavachari, K., Binkley, J.S., Gonzalez, C., Martin, R.L., Fox, D.J., Defrees, D.J., Baker, J., Stewart, J.J.P., and Pople, J.A., Gaussian, Inc., Pittsburgh PA, 1992.
10. ELECTRA: An Electronic Analysis Program, Mestres, J., IQC-UdG, Girona, CAT, 1994.
11. SOLÀ, M., MESTRES, J., OLIVA, J.M., DURAN, M., CARBÓ, R., Int. J.Quantum Chem., in press.
12. MESSEM: A Density-based Molecular Similarity Program, Mestres, J., Solà, M., Besalú, E., Duran, M., and Carbó, R., IQC-UdG, Girona, CAT, 1994.

Table I. Volume (VOL_{vdw} and VOL_{den} in au³), molecular similarity measures (MSM_{am1}, MSM_{gau}, MSM_{fit}, MSM_{scf} and SASM in atomic units) and energies (in hartrees) for the different isomeric systems studied. See text (methodology section) for label definition.

Isomerism	Formula	Isomer	VOL_{vdw}	VOL_{den}	MSM_{am1}	MSM_{gau}	MSM_{fit}	MSM_{scf}	SASM	Energy
Conformational	C_2H_6	eclipsed	299.42	234.05	1.9157	6.1706	62.7028	62.7967	63.9252	-78.789565
	(ethane)	staggered	297.00	232.13	1.9177	6.1975	62.7133	62.8044	63.9252	-78.793948
	C_4H_8	cis	479.12	423.18	3.4093	12.9650	125.1227	125.3086	127.6912	-155.243830
	(2-butene)	trans	477.28	420.08	3.4085	12.9736	125.1336	125.3169	127.6912	-155.246332
Configurational	$C_4H_{10}O_2$	rr,ss	640.82	540.69	9.4643	28.2524	285.3259	285.6742	291.4984	-305.300536
	(2,3-butanediol)	rs,sr	639.10	540.29	9.4937	29.2729	285.3360	285.6849	291.4984	-305.307280
Constitutional	C_2H_5NO	acetamide	398.36	333.33	6.2494	19.7620	194.3870	194.5589	198.4982	-206.814673
		acetaldoxime	400.55	340.00	5.9670	19.4173	194.3973	194.6547	198.4982	-206.711172
	C_4H_4O	furan	429.65	376.43	5.7417	21.6359	204.7184	205.0225	209.3958	-227.350075
		2-butinal	460.22	397.40	5.7610	20.6755	204.8752	205.2052	209.3958	-227.310110

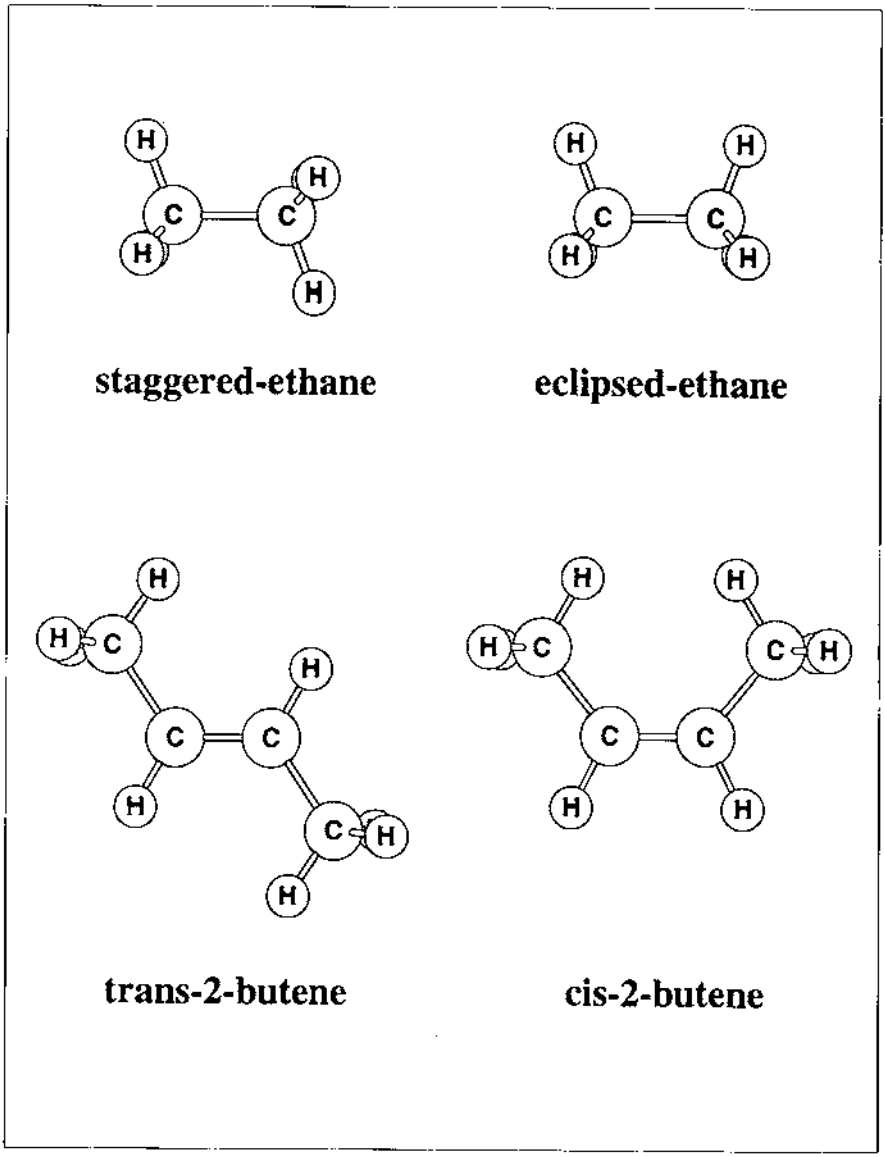

Figure 1. The two sets of conformational isomers: ethane (eclipsed and staggered, up) and 2-butene (cis and trans, down)

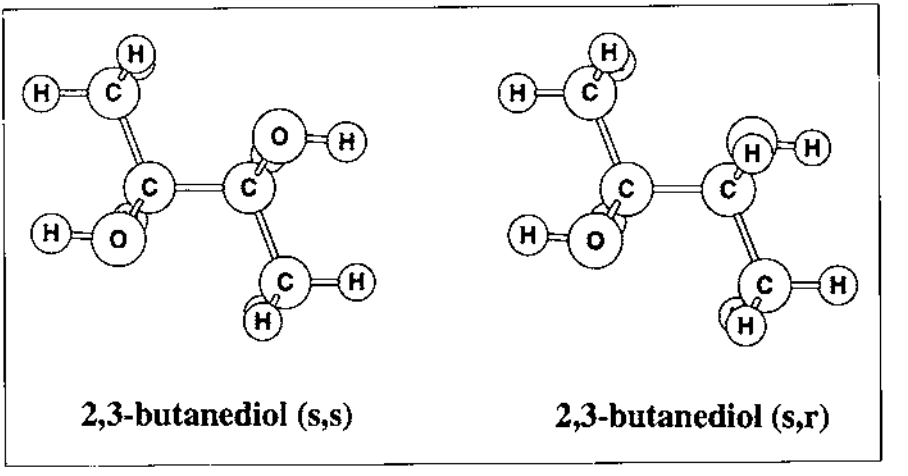

Figure 2. The set of configurational isomers: 2,3-butanediol (ss and sr)

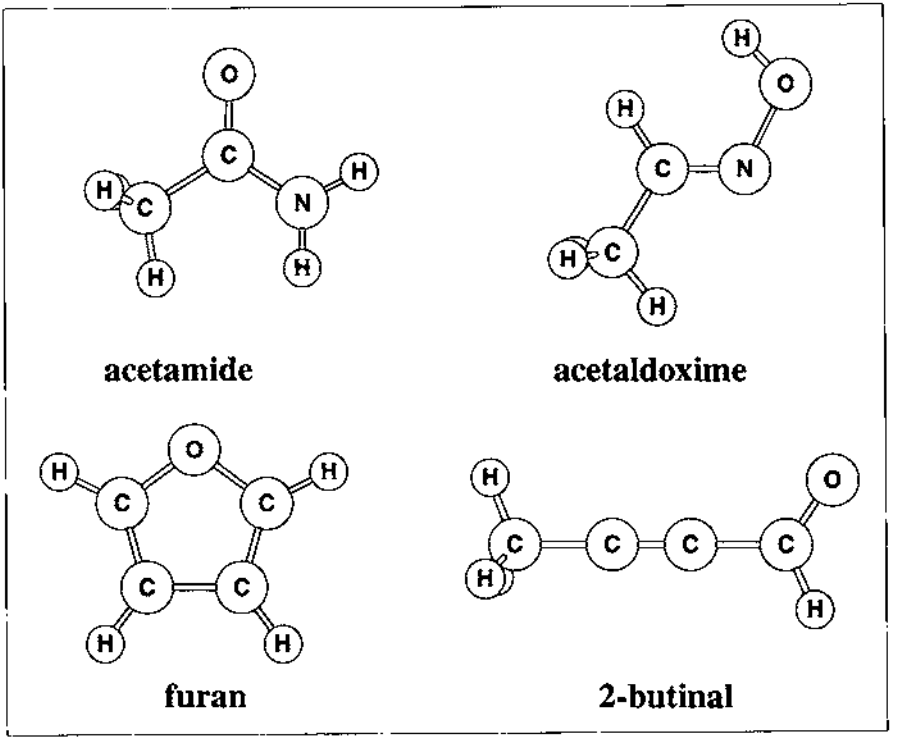

Figure 3. The two sets of constitutional isomers: acetamide and acetaldoxime (up) and furan and 2-butinal (down)