SORRISO: Quantum chemistry tool for visualization of electron transfer processes

Ildar Galikov [a] , Alytis Gruodis
Department of General Physics & Spectroscopy, Vilnius University
Saulėtekio 9, korp. 3, Vilnius, Lithuania

Received 3 September 2010, accepted 3 November 2010

Abstract. User friendly program *SORRISO* was created for two-dimentional (2D) viewing of molecular system. Created „from zero“ using object-oriented programming (OOP) paradigm, *SORRISO* was is able to visualize the positions of nuclears of the molecular system, presented in different formats such as XYZ (cartesian), YES (hexadecimal record), ENT (Brookhaven pdb). Additional possibilities of *SORRISO* includes the visualization of the dynamic process – intermolecular electron transfer (IET) presented in the framework of molecular orbitals (MO) through atomic orbitals (AO). Rate of IET for each intermolecular transition between one-particle states is presented in typical spectral form, and increasing / decreasing of charge density in the surrounding of nuclear is presented dynamically - by blicking of balls.

Citations: Ildar Galikov, Alytis Gruodis. *SORRISO*: Quantum chemistry tool for visualization of electron transfer processes – *Innovative Infotechnologies for Science, Business and Education,* ISSN 2029-1035 – **2(9)** 2010 – Pp. 11-16.

Keywords: Sorriso; Charge transfer; Intermolecular electron transfer; Visualization.
PACS: 82.30.Fi.
Short title: Sorriso - electron transfer.

Introduction

XXI century could be named as century of information in natural sciences. Big flow of different kind of unsorted information makes life difficult for researchers in many fields of molecular physics, applied mathematics, bioinformatics etc. Unsorted information represents some times the wrong image of phenomenon of due to partial, perfunctory view into the problem. But necessity to sort and summarize the initial information always is related to two signidicant factors: i) big expences of computer resources; ii) *know-how* idea - how to present the sophisticated idea in convenient form.

Scientific information represented visually is much easier and quicker to understand in comparison to tabular/row number way. Visualization in chemical physics is always related to the *molecular* model or model of *molecular system*. Distribution of molecular fragments presented in the structural IUPAC form is less less easer that in spatial form (see Fig. 1). Both models are equivalent chemically, but spatial distribution allows us to model the electronic charge redistribution by optical excitation. Moreover, presence of additional parameters to atomic spatial distribution presented in user-frienly form allows us to manipulate not by static representation but by dinamic also. This is the reason, why visualisation of digital information of spectral distributions is very important and difficult problem in physics.

Rapid grow of computer usage in the molecular modelling call the formation of three essential tool types:

i) *molecule editors* as application programs or applets - *plug-in*s for creating and modifying three-dimensional (cartesian, Z-matrix, etc) representations of molecular structures. Editing in graphical mode as well as in text mode requires several knowledges and previous run-in;
ii) *molecular viewers* as application programs or applets - *plug-in*s for imaging of molecular structures;
iii) many applied tolls contain both properties: molecule editors as well as molecular viewers.

Typical examples of molecule editors are follow: ChemDraw [1], BKChem [2], and molecular viewers (usually together with some elements of molecule editors): HyperChem [3], Gausview [4], RasMol [5] etc. These programs are convenient and effective, although sometimes they are not sufficient. In order to fully process data of specific problems, special programs or *plug-in*s must be created.

In the field of many recent problems in chemical physics, intermolecular electron transfer (IET) plays an significante role. Several chemical (Fe corosium, redox etc) as well as protochemical (light harvesting, exciplex absorption etc) reactions in pigmentary pairs of donor-acceptor molecules can be well modeled using quantum-chemistry approaches.

[a]Corresponding author, email: *galikov@gmail.com*

The purpose of this work is visualisation of dynamics of IET parameters through data obtained by using quantum-chemistry packages - Gaussian03 [6], GAMESS [7], Dalton 2.0 [8].

Visualization of IET parameters is presented for several molecular forms of chemical compound N-N-dimethylamino benzylidene indan-(1,3)dione (DMABI) which was very extensively studied last decade [9-12].

SORRISO [ital. *smile*] - is a new and unique program designated for convenient and quick analysis of large amounts of data. C++ source code was written in *object-oriented programming* (OOP) manner, using *Windows.h* [13] and *OpenGL* [14] libraries. *Visual Studio 8.0* [15] was used as an integrated development environment from *Microsoft* corporation. Executable file was prepared as an *Win32-Api*.

The aim of the project *SORRISO* was to create a convenient tool for simulation of molecular view (2D projection), that would allow:

i) static view - representation of positions of atoms (nuclears) using different models: nuclears presented by balls only, by balls with sticks, by sticks only;
ii) dinamic view - recalculation of recent projection of nuclear position "by the way" according to user request, using left and right bootom of mouse manipulator (rotation in plane and rotation in space, respectivelly);
iii) dinamic view - recalculation of recent projection of nuclear position "by the way" according to user request, using left and right bootom of mouse manipulator (magnifying and shrinking, respectivelly);
iv) "pseudo-perspective" viewing of part of molecular system in order to recognize the distance from viewing point until different atoms; small radius of ball creates the effect of long distance;
v) spectral representation of the IET rate on the energy of donor state (number of donor MO);
vi) requesting of the corresponding change of the electron density in donor and acceptor systems and blicking visualization of intermolecular electron transfer.

1. Visualization principle

1.1. Assumptions for molecular system

Creating the model of molecular system it is necessary to understand which essential factors must be estimated during energetic changes. Under energetic changes, particularly, several processes could occur – decreasing / increasing of electronic cloud density; molecular conformational changes; decomposition of aggregate; etc. In mentioned cases molecular system containing by nuclears as well as electrons must be estimated in different aspects.

First assumption of quantum chemistry claims that nuclear positions predetermine the molecular energetics. This means that spatial distribution of nuclears (but not electrons) is essential for describing energetical interactions.

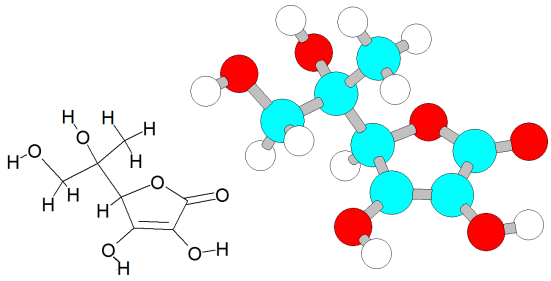

Fig. 1. Different molecular respresentation of vitamin C: structural IUPAC form (left) and spatial form.

According to this assumptions, electrons always adapt (fit) to the positions of nuclears. This assumption allows us to describe molecular geometry as the spatial distribution of nuclears (in 3D coordinates).

Second assumption of quantum chemistry claims that geometries of molecular system in the ground and excited states must be closed to each other. This means that only slightly perturbed molecules (for example, by optical excitation) could be described using quantum chemistry approaches.

According to such two main assumptions, many molecular visualization models operate. Positions of nuclears in 3D speace represents molecular geometry, and this factor is basic and fundamental. Main task of nuclear visualization could be titled as an *projection calculation task*. Additionaly, electronic density related to each atom (in the nearest nuclear surrounding) could be presented in some ways, for example, using molecular orbital (MO) methods and so on.

1.2. Calculation of projections

Molecular geometry or three-dimensional (3D) distribution of atoms (nucleous) is traditionally performed by matrix of different type: cartesian form as well as Z-matrix form, as shown in Fig. 2. Cartesian matrix represents an record of three real numbers representing $\{x,y,z\}$ coordinates for each atom. Z-matrix represents an record of three real numbers representing distance, plane angle and dihedral angle $\{r, \psi, \phi\}$ as spherical coordinates for each atom.

Project *SORRISO* was prepared as tool of molecular viewer. No *plug-in*s for structure editing was inserted, only tools of rotating and zooming are present.

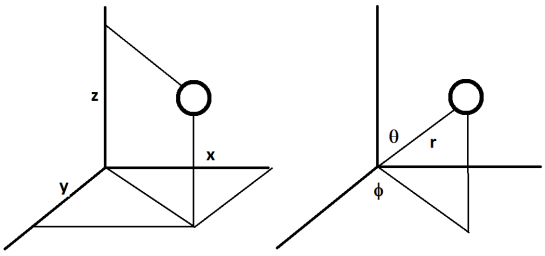

Fig. 2. Cartesian and Z-matrix (spherical) coordinates.

In order to realize smooth 3D rotation and zooming, several mathematical operations were required [16]:
i) transformation of cartesian coordinates to spherical;
ii) factoring of r (zooming operation);
iii) transformation of spherical coordinates to cartesian;
iv) routine calculation of 2D projections from cartesian $\{x,y,z\}$ coordinates (3D rotation operation).

In order to prevent distortion of coordinates, zooming was realized by using spherical coordinates. In order to calculate the zooming, it is sufficient to change the radius r and transform the coordinates back to cartesian. The transformation from spherical to cartesian is performed according following formulae:

$$x = r \cdot \sin\psi \cdot \cos\phi \quad (1)$$

$$y = r \cdot \sin\psi \cdot \sin\phi \quad (2)$$

$$z = r \cdot \cos\psi \quad (3)$$

where variables r, ψ, ϕ are prelimited in presented intervals:

$$r \subset [0,\infty), \psi \subset [0,\pi), \phi \subset [0,2\pi), \quad (4)$$

Mathematically, the problem is solved as following. Lets assume the existing of two coordinate system: usual $\{XYZ\}$ and transformed $\{X'Y'Z'\}$. If the spatial rotation of molecular system is performed (by angle α in XY plane and angle β in $Y'Z$ plane) the new projections in previous coordinate system have to be recalculated. Vector $\mathbf{r}$, describing place of an atom is introduced for that cause in two different coordinate systems:

$$\mathbf{r} = x_1 \cdot \mathbf{n}_1 + x_2 \cdot \mathbf{n}_2 + x_3 \cdot \mathbf{n}_3 \quad (5)$$

$$\mathbf{r} = x'_1 \cdot \mathbf{n}'_1 + x'_2 \cdot \mathbf{n}'_2 + x'_3 \cdot \mathbf{n}'_3 \quad (6)$$

where $\mathbf{n}_1$, $\mathbf{n}_2$, $\mathbf{n}_3$, $\mathbf{n}'_1$, $\mathbf{n}'_2$, $\mathbf{n}'_3$ are unit vectors in two different cartesian systems, respectivelly. Since vector did not change, the equations are equal:

$$x_1 \cdot \mathbf{n}_1 + x_2 \cdot \mathbf{n}_2 + x_3 \cdot \mathbf{n}_3 = x'_1 \cdot \mathbf{n}'_1 + x'_2 \cdot \mathbf{n}'_2 + x'_3 \cdot \mathbf{n}'_3 \quad (7)$$

We then multiply everything by $\mathbf{n}_1$, $\mathbf{n}_2$, $\mathbf{n}_3$, in turn and after applying the orthogonality of vectors, we get a system of equations:

$$x_1 = x'_1(\mathbf{n}_1 \cdot \mathbf{n}'_1) + x'_2(\mathbf{n}_1 \cdot \mathbf{n}'_2) + x'_3(\mathbf{n}_1 \cdot \mathbf{n}'_3) \quad (8)$$

$$x_2 = x'_1(\mathbf{n}_2 \cdot \mathbf{n}'_1) + x'_2(\mathbf{n}_2 \cdot \mathbf{n}'_2) + x'_3(\mathbf{n}_2 \cdot \mathbf{n}'_3) \quad (9)$$

$$x_3 = x'_1(\mathbf{n}_3 \cdot \mathbf{n}'_1) + x'_2(\mathbf{n}_3 \cdot \mathbf{n}'_2) + x'_3(\mathbf{n}_3 \cdot \mathbf{n}'_3) \quad (10)$$

This is the expression of old projections over new ones. Analogically, we may express them vice versa where α_{ik} is an coefficient of transformation:

$$\alpha_{ik} = \mathbf{n}_i \cdot \mathbf{n}'_k = \cos(\mathbf{i} \cdot \mathbf{k}) \quad (11)$$

We now calculate the coefficient matrix α_{ik}:

$$\alpha_{ik} = \begin{Vmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{Vmatrix} \quad (12)$$

In our case we need to solve nine stereometry tasks in order to obtain coefficients:

$$\alpha_{ik} = \begin{Vmatrix} \cos(\alpha) & \sin(\alpha) * \sin(\beta) & -\sin(\alpha) * \cos(\beta) \\ 0 & \cos(\beta) & \sin(\beta) \\ \sin(\alpha) & -\cos(\alpha) * \sin(\beta) & \cos(\alpha) * \cos(\beta) \end{Vmatrix} \quad (13)$$

2. Routine scheme

Libraries used in *SORRISO* are shown in Table 1.

The main file of the program sorriso.cpp consists of several fragments. Description of standard libraries (see Table 1) constitute the first fragment. Following it is the main method of program *WinMain* and auxiliary *CallBack* method. *WinMain* is responsible for creation of dialogue window. All objects are initialized in this method during the execution process. An action listener loop is performed in *WinMain*. The purpose of it is to analyse system messages for specific actions, such as mouse click and selection of menu. *CallBack* methods describe the execution upon specific user-requested actions.

2.1. Organisation of data input

Input of data is realised *via* an easy to access menu, which is written in sorriso.rc file. After selection of an adequate menu item, dialog window with option to select a file of a specific type, appears. The selection of file type is filtered, so users will not be able to choose files of other types. According to well-known information storage rule, it is recommended to read data from file only once, especially large ones.

Fig. 3 shows main window part of *SORRISO* as menu item opened for file input. After selection of the file, all it's contents are copied to a container of vector type object in order to process it and create a neat structure. Vector instead of and ordinary array was chosen because molecules have different number of atoms. Reading of file is performed in two phases:
a) insertion of all contents in file to a vector and calculation of atoms;
b) analysis of vector and creation of a structure consisting of required data.

Table 1. Standard libraries used in *SORRISO*.

OpenGL	glu.h	Advanced graphics
C++	string	Container for character string
	fstream	File input-output operation
	iomanip	Description of stream formats
	vector	Container of universal type
	math	Standard mathematics
C	windows.h	windows tools
	stdio.h	File input-output operation

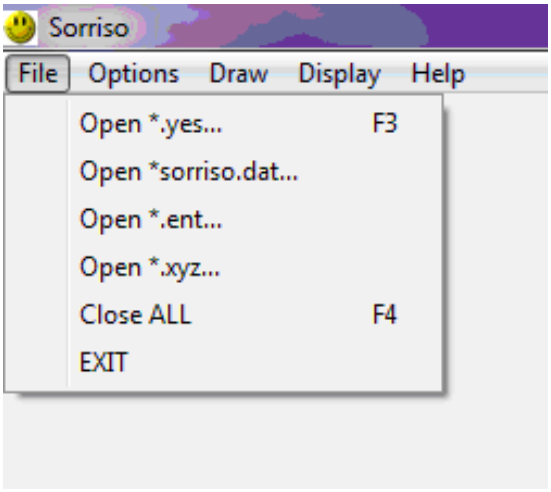

Fig. 3. Main window part of *SORRISO*: menu item for input.

Once the number of atoms is known, fixed-size arrays may be created. Since arrays are processed faster, they are preferred over vectors.

2.2. Possible formats of input data

XYZ file format. This file format (file extension - *.xyz) is known as a so called *chemical file format* [17]. Formal standard is absent, and many IT companies use several variations. Typical XYZ format (presented in Table 2) is devoted for determination of molecule geometry by giving requested lines of records containing atom name (IUPAC standard) and atomic cartesian coordinates separated by space. The units are generally in Ångströms. Precision must be presented as three or more quantities after comma.

YES file format. YES file format (file extension - *.yes) is used for practical purposes as high-precision format. Precision of number is defined as standart *double precision* for real numbers with floating point coded by four byte (4B) coding system. Formal YES standard was created originally for project *NUVOLA* [18] in order to keep the same precision of real numbers by information exchanging between operating modules through file reading/writing operation. Typical YES format (presented in Table 2) is devoted for determination of molecule geometry by giving requested lines of records containing atom name (IUPAC standard) and atomic cartesian coordinates separated by space (the same like in XYZ format). Additional string of 16 characters (ASCII only) represents the coded real number (8 bytes). Each byte (1B=8b) is presented in form of two hexadecimal characters (4 bits into 1 char).

ENT-matrix file format. ENT file format (file extension - *.yes) belongs to the group of the chemical file format. Textual file format for describing the three dimensional structures of molecules is very useful for exchanging purposes [19].

Table 2. Acetylene. XYZ, YES, ENT, Z-matrix data formats.

```
XYZ data format

C            -1.869  -0.737   0.000
C            -0.957   0.027  -0.000
H            -2.686  -1.420  -0.000
H            -0.140   0.710   0.000

YES data format

     6          C           1
     6          C           2
     1          H           3
     1          H           4
BFFDE76C8B439581    -1.869000000000     1 1
BFE795810624DD2F    -0.737000000000     1 2
0000000000000000     0.000000000000     1 3
BFEE9FBE76C8B439    -0.957000000000     2 1
3F9BA5E353F7CED9     0.027000000000     2 2
8000000000000000    -0.000000000000     2 3
C0057CED916872B0    -2.686000000000     3 1
BFF6B851EB851EB8    -1.420000000000     3 2
8000000000000000    -0.000000000000     3 3
BFC1EB851EB851EC    -0.140000000000     4 1
3FE6B851EB851EB8     0.710000000000     4 2
0000000000000000     0.000000000000     4 3
402805B573EAB368    12.011150000000     1
402805B573EAB368    12.011150000000     2
3FF020A5269595FF     1.007970000000     3
3FF020A5269595FF     1.007970000000     4

ENT data format

HETATM    1  H    1   -2.686  -1.420  -0.000
HETATM    2  H    2   -0.140   0.710   0.000
HETATM    3  C    3   -1.869  -0.737   0.000
HETATM    4  C    4   -0.957   0.027  -0.000
CONECT    1    3
CONECT    2    4
CONECT    3    4    1
CONECT    4    3    2
END

Z-MATRIX data format

X1
X2    1    1.8
X3    2    1.8    1   90.0
X4    3    1.8    2   90.0    1  180.0
C5    2    0.665  3   90.0    4  -90.0
C6    2    0.665  3   90.0    4   90.0
H7    5    1.03   2  180.0    3  -90.0
H8    6    1.03   2  180.0    3   90.0
```

ENT format (represented in Table 2) as a variation of Protein Data Bank (pdb) file format is devoted for determination of molecule geometry similar like in XYZ format.

Z-matrix file format. Z-matrix file format (file extension - *.zmt) belongs to the group of the chemical file format. [16] as an internal coordinate representation. Due to the modeling purposes, it is convenient to prepare Z-matrix in structure prelimited terms such as bond length, plane angle, dihedral angle. In that case bonding characteristics will be presented easier in comparison to XYZ. Additional elements of geometry (symmetry plane, axis etc) could be included using dummy atom formalism - see example in Table 2. Fig. 4 represents the spatial distribution of mentioned Z-matrix containing real C,H as well as dummy X atoms.

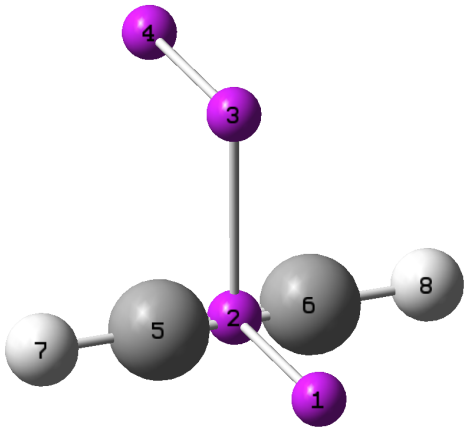

Fig. 4. 3D-distribution of acetylene atoms; dummy atoms X represent the symmetry plane.

2.3. Problems of visualisation

The visualisation task of molecule projection needs an enormous amounts of computer resources. The main problem is the calculation of trigonometrical functions. Mathematical calculation of them is expansion of Taylor's sequence until desired accuracy is achieved. This routine is resource dependent in very high degree, since every new projection of an atom requires at least 6 functions. Graphical function for drawing of a sphere requires at least 720 calls of function. In total, one atom requires 726 expansions of Taylor's sequence. Using of *windows.h* methods that depend only on processor power, even on very powerful workstations was too slow. This circumstance forcing to rewrite all functions using *OpenGL* library, which uses the graphics accelerator as well.

Projection on z axis seems unnecessary, but it is essential to realization of 3D perspective - closer atoms may block ones further away. Mathematically this condition is realised over projective selection of an atom diameter. Atoms further away seem smaller, thus making the view more realistic.

Fig.5-7 represent the DMABI dimmer using *Balls and sticks*, *Balls only*, *Sticks only* methods, respectively. User can decided according his own meaning which type of visualization is the best. Reswitching between three different regimes - *Balls and sticks*, *Balls only*, *Sticks only* - is allowed by standard windows bottoms on the program window.

3. Intermolecular electron transfer.

Calculation of IET parameters was provided using *NUVOLA* [18]. After opening YES file and selecting menu option "Open sorriso.dat" in "File" menu, *SORRISO* extracts information of spectrum which is represented in the main window. For user convenience, required peak of spectrum is selected by a mouse click. The selected peak changes colour to blue, and the selected spectrum is displayed in red (see Fig. 8).

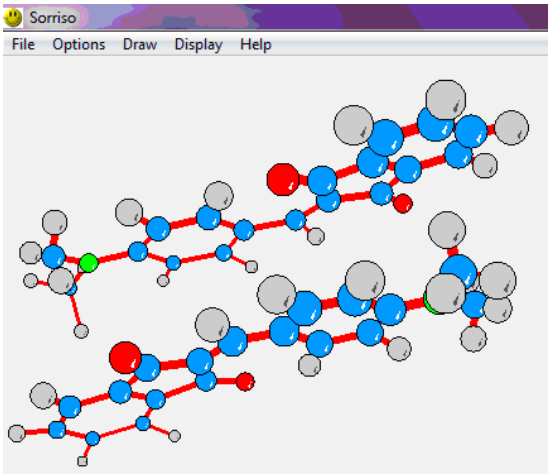

Fig. 5. DMABI dimmer represented in *Balls and sticks* method.

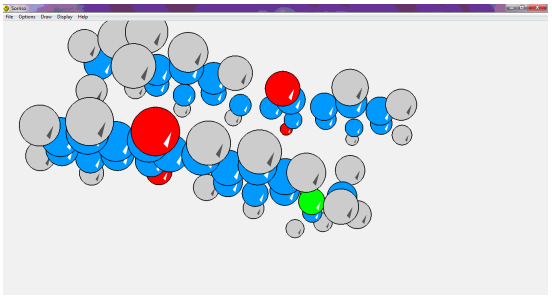

Fig. 6. DMABI dimmer represented in *Balls only* method.

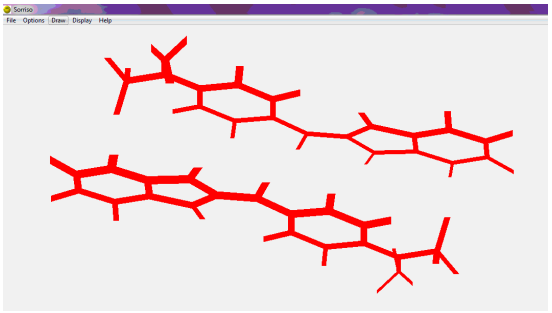

Fig. 7. DMABI dimmer represented in *Sticks only* method.

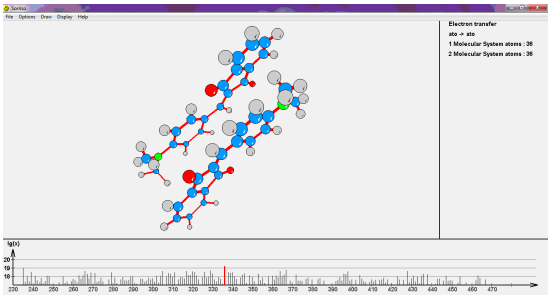

Fig. 8. Molecule and its IET spectrum.

Every peak of spectrum represents information of intermolecular electron transfer, thus by manipulating mouse on the expansion of spectrum on the bottom of the window it is possible to gain visual information of IET in real time - if, for example, user wants to find out the exact atoms participating in the electron transfer (see Fig. 9).

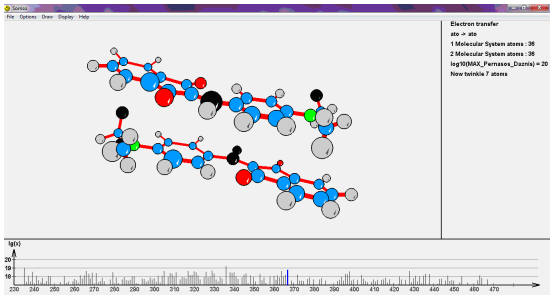

Fig. 9. DMABI dimer and its spectrum with selected peak. At initial time moment, 7 atoms of the highest frequency are displayed in black.

In order to display this, fourth dimension - time is taken into consideration. Program reads required data and sorts the array in bubble sort method. Since only the most active transfers are required, the program displays only seven, most active atoms. These atoms increase twice in size and start blinking. The higher the frequency - the higher probability of a jump the electron has.

Conclusions

Convenient tool *SORRISO* for simulation of molecular view in 3D space was created. *SORRISO* as molecular viewer could be useful for estimating of molecular 2D projections when molecular derivatives are represented in XYZ, YES, ENT data formats. Also *SORRISO* allows simultaneously visualise the parameter of intermolecular electron transfer through local charge redistribution.

References

1. Mills, N. ChemDraw Ultra 10.0. – *J. Am. Chem. Soc.* 128 (41) (2006) 13649–13650.
2. <http://bkchem.ziral.org/index.html>, accessed 2010-01-05.
3. HyperChem™, Hypercube, Inc., <http://www.hallogram.com/science/hyperchem/>, accessed 2010-01-19.
4. Gausview™, Gaussian Inc., <http://www.gaussian.com/>, accessed 2010-01-05.
5. RasMol <http://en.bio-soft.net/3d/rasmol.html>, accessed 2010-01-24.l
6. Gaussian03, Gaussian Inc., <http://www.gaussian.com/>, accessed 2010-01-05.
7. GAMESS, <http://www.msg.ameslab.gov/gamess/index.html>, accessed 2010-01-05.
8. DALTON, a molecular electronic structure program, Release 2.0 (2005), <http://www.kjemi.uio.no/software/dalton/dalton.html>, accessed 2010-01-05.
9. M. Balevicius, E. Stumbrys, B. Sorokolit, A. Gruodis. Theoretical studies of electronic absorption spectra of dimethylaminobenzylidene 1,3-indandione (DMABI) by means of semiempirical quantum chemical calculations. – *Lithuanian Journal of Physics* Vol.35 N.1 (1995) 20-26.
10. V.Gulbinas, G. Kodis, S. Jursenas, L. Valkunas, A. Gruodis, J. C. Mialocq, S. Pommeret, T. Gustavsson. Charge transfer induced excited state twisting of N,N-dimethylaminobenzylidene-1,3-indandione in solution. – *Journal of Physical Chemistry A* 103 (1999) 3969-3980.
11. N. Galikova, M. Kelminskas, A. Gruodis, L. M. Balevičius. Tyrosine – tryptophan complex: intermolecular electron transfer using quantum chemistry approach. – *Computer Modelling and New Technologies* Vol.15 N.1 (2011) 19–23.
12. Nurija Galikova, Mantas Kelminskas, Alytis Gruodis, Leonas Mindaugas Balevičius. Modelling of electron transfer in DMABI dimer by quantum chemistry approach. – *Chemija* 22(2) (2011) 73–79.
13. OPENGL, <http://www.opengl.com>, accessed 2010-04-02.
14. <http://edu.ioffe.ru/register/?doc=physica/lect.ch4.tex>, accessed 2010-01-20.
15. Microsoft Visual Studio 2008. Version 9.0.21022.8 RTM – 2007 Microsoft corporation.
16. Jerod Parsons, J. Bradley Holmes, J. Maurice Rojas, Jerry Tsai, Charlie E. M. Strauss. Practical conversion from torsion space to Cartesian space for in silico protein synthesis – *Journal of Computational Chemistry* 26(10)(2005) 1063–1068.
17. XYZ format specification, <http://openbabel.org/wiki/XYZ_%28format%29>, accessed 2010-04-20.
18. Nurija Galikova, Alytis Gruodis. *NUVOLA*: intermolecular electron transfer properties using quantum chemistry approach. – *Innovative Infotechnologies for Science, Business and Education,* ISSN 2029-1035 – N2(3) (2008) 12.1-12.9.
19. <http://www.wwpdb.org/documentation/format33/v3.3.html>, accessed 2010-04-09.