Theor Chem Account (2007) 118:673–679
DOI 10.1007/s00214-007-0352-0

REGULAR ARTICLE

Solutions to the Quantum QSPR problem in molecular spaces

Ramon Carbó-Dorca · Sofie Van Damme

Received: 12 December 2006 / Accepted: 11 March 2007 / Published online: 7 June 2007

Abstract The molecular quantum similarity framework is used to present a new set of Quantum Quantitative Structure–Properties Relationship (QQSPR) procedures. The theoretical basis consists of the so-called fundamental QQSPR equation, deducible from quantum mechanical first principles, associated with the quantum mechanical expectation values computation. Approximate solutions of the fundamental QQSPR equation within direct and reciprocal spaces, containing molecular density functions, are studied in a common framework.

Keywords Quantum mechanical expectation values · Molecular density functions · Quantum similarity measures · Quantum QSPR · Fundamental QQSPR equation · Reciprocal density spaces

1 QSPR operators and fundamental QQSPR equation in original and reciprocal basis

1.1 Quantum similarity measures and quantum similarity matrices

The typical QQSPR framework [1] starts being built up with a known basis set of density functions attached to a so-called molecular core set of cardinality $M : P = \{\rho_I(\mathbf{r})\}$. The core set P can be used in turn to construct a metric matrix by means of molecular quantum similarity measures (MQSM) [2], which, in a quite general form, can be defined weighted by a positive definite operator: $W(\mathbf{r})$, according to the algorithm

$$\mathbf{Z} = \left\{ z_{IJ} = \int_D \rho_I(\mathbf{r}) W(\mathbf{r}) \rho_J(\mathbf{r}) \mathrm{d}\mathbf{r} = \langle \rho_I | \, W \, | \rho_J \rangle \right\}. \quad (1)$$

In the usual practical cases, the unit weighting operator $W(\mathbf{r}) = I$ is chosen, so the simplified metric definition

$$\forall I, J : z_{IJ} = \langle \rho_I \mid \rho_J \rangle = \int_D \rho_I(\mathbf{r}) \rho_J(\mathbf{r}) \mathrm{d}\mathbf{r} \quad (2)$$

holds, producing a metric based on overlap quantum similarity measures. In what follows, the metric will be referred to definition (2) for simplicity, but all that is said can be transferred to the general definition (1) without problems, by just substituting the formal metric elements where it is needed.

1.2 Shape functions and quantum similarity measures

A possible alternative computational scheme can be constructed using shape functions instead of density functions; see for example [3]. However, in this case, the similarity measures of Eqs (1) and (2) are multiplied by the Minkowski

---

Contribution to the Serafín Fraga Memorial Issue.

This paper is dedicated to Serafín Fraga and Xavier Gironès, *in memoriam*. They have formed part of my life as a scientist: I had been an apprentice with the first and I tried to be a teacher with the second. I learned from both, though. They will be like stars in my mind, until we can walk together the endless path of nothingness.

R. Carbó-Dorca (✉) · S. Van Damme
Department of Inorganic and Physical Chemistry,
Ghent University, Krijgslaan 281, 9000 Ghent, Belgium
e-mail: quantumqsar@hotmail.com

R. Carbó-Dorca
Institut de Química Computacional, Universitat de Girona,
Girona, 17071 Catalonia, Spain

norms of the elements of the core set, as any shape function belonging to a new core set $S = \{\sigma_I\}$ is constructed as a scaled function, defined by means of the algorithm

$$\forall \rho_I \in P \to \langle \rho_I \rangle = \int_D \rho_I(\mathbf{r})\mathrm{d}\mathbf{r} = N_I \to \sigma_I = N_I^{-1}\rho_I,$$

where $\{N_I\}$ are coincident with the number of particles of the molecules constituting the core set. The use of shape functions in order to obtain similarity measures will, thus, transform the similarity matrix described in Eq (1) into a shape similarity matrix $\mathbf{Z}_S$ by employing the trivial matrix multiplication scheme:

$$\mathbf{N} = \mathrm{Diag}(N_I) \to \mathbf{Z}_S = \mathbf{N}^{-1}\mathbf{Z}\mathbf{N}^{-1}.$$

In fact, this shows how both core set basis set choices are equivalent, and therefore from now on only density functions will be considered, as the corresponding algorithms become the same, except for some trivial scaling factors.

1.3 Quantum similarity operator and fundamental QQSPR equation

QQSPR problems are solved by defining an operator by means of the series [4] associated with the elements of the basis set P:

$$\Omega(\mathbf{r}) = \gamma(\mathbf{r}) + \sum_I x_I \rho_I(\mathbf{r}) + \sum_I \sum_J x_I x_J \rho_I(\mathbf{r})\rho_J(\mathbf{r}) + O(3), \tag{3}$$

where $\gamma(\mathbf{r})$ is a known fixed gauge operator, usually chosen as a null function term, and $|\mathbf{x}\rangle = \{x_I\}$ is a set of coefficients to be determined and ordered as a column vector.

In the usual QQSAR problems, the core set P is also attached to a set of known values of some property, which can be also ordered as a column vector $|\mathbf{p}\rangle = \{p_I\}$. The operator (3) can be used in the usual quantum mechanical fashion to compute the property values of the core set. Thus,

$$\begin{aligned}\forall I : p_I = \langle \Omega\rho_I \rangle &= \int_D \Omega(\mathbf{r})\rho_I(\mathbf{r})\mathrm{d}\mathbf{r}\\ &= \langle \gamma\rho_I \rangle + \sum_J x_J \langle \rho_I\rho_J \rangle + O(2)\\ &\approx g_I + \sum_J x_J z_{IJ}\end{aligned}$$

and so, one arrives at the linear set of equations, which constitute the fundamental QQSPR equation:

$$\forall I : \pi_I = p_I - g_I = \sum_J z_{IJ} x_J \to |\pi\rangle = \mathbf{Z}\,|\mathbf{x}\rangle\,. \tag{4}$$

From this the coefficient vector $|\mathbf{x}\rangle$ can be obtained, owing to the existence of the inverse metric matrix $\mathbf{Z}^{-1}$ always being assured due to the positive definite nature of the metric matrix $\mathbf{Z}$. Thus,

$$|\mathbf{x}\rangle = \mathbf{Z}^{-1}\,|\boldsymbol{\pi}\rangle \to \forall I : x_I = \sum_J z_{IJ}^{(-1)}\pi_I.$$

1.4 Reciprocal space fundamental equation

Also, the inverse metric matrix is positive definite; that is, one can write:

$$\mathbf{Z} > 0 \to \mathbf{Z}^{-1} > 0,$$

because whenever the metric diagonalization is performed

$$\mathbf{U}^{+}\mathbf{Z}\mathbf{U} = \Theta = \mathrm{Diag}(\theta_I) \wedge \mathbf{U}^{+}\mathbf{U} = \mathbf{U}\mathbf{U}^{+} = \mathbf{I} \Rightarrow \forall I : \theta_I \in \mathbf{R}^{+}.$$

It is easy to prove that the following diagonalization for the inverse also holds:

$$\mathbf{U}^{+}\mathbf{Z}^{-1}\mathbf{U} = \Theta^{-1} = \mathrm{Diag}\left(\theta_I^{-1}\right) \Rightarrow \forall I : \theta_I^{-1} \in \mathbf{R}^{+}.$$

As a consequence, the inverse metric matrix $\mathbf{Z}^{-1}$ can be considered, in turn, a metric matrix. The inverse metric $\mathbf{Z}^{-1}$, defining the so-called reciprocal space, associated with the original space with metric $\mathbf{Z}$.

Given the core basis set P, the attached metric can be written by means of the formal equation:

$$\mathbf{Z} = \int_{\mathbf{D}} |P\rangle\,\langle P|\,\mathrm{d}\mathbf{r} \wedge \langle P| = (\rho_1; \rho_2; \ldots \rho_N). \tag{5}$$

The reciprocal core basis set $P^{(-1)} = \left\{\rho_I^{(-1)}\right\}$ is readily obtained by multiplying definition (5) on both sides by the inverse metric matrix, yielding:

$$\begin{aligned}\mathbf{Z}^{-1} &= \int_{\mathbf{D}} \mathbf{Z}^{-1}\,|P\rangle\,\langle P|\,\mathbf{Z}^{-1}\mathrm{d}\mathbf{r} \wedge \langle P^{(-1)}| = \langle P|\,\mathbf{Z}^{-1}\\ &= \left(\rho_1^{(-1)}; \rho_2^{(-1)}; \ldots \rho_N^{(-1)}\right).\end{aligned}$$

Therefore, one can write for the reciprocal core basis set functions definition:

$$\forall I : \rho_I^{(-1)} = \sum_J z_{IJ}^{(-1)}\rho_J.$$

Moreover, the hybrid metric constructed by the original basis and the reciprocal one is the unit matrix, and this produces the following equations:

$$\begin{aligned}\int_D |P\rangle\,\langle P^{(-1)}|\mathrm{d}\mathbf{r} &= \int_D |P\rangle\,\langle P|\mathbf{Z}^{-1}\mathrm{d}\mathbf{r}\\ &= \left(\int_D |P\rangle\,\langle P|\mathrm{d}\mathbf{r}\right)\mathbf{Z}^{-1} = \mathbf{Z}\mathbf{Z}^{-1} = \mathbf{I}.\end{aligned}$$

The same holds for the transpose construction:

$$\begin{aligned}\int_D |P^{(-1)}\rangle\langle P|d\mathbf{r} &= \int_D \mathbf{Z^{-1}}\,|P\rangle\langle P|d\mathbf{r}\\ &= \mathbf{Z^{-1}}\left(\int_D |P\rangle\langle P|d\mathbf{r}\right) = \mathbf{Z^{-1}Z} = \mathbf{I}.\end{aligned}$$

Nothing opposes to the possibility that the property operator can be expressed in terms of the reciprocal basis set, at the same time that every molecule is represented in reciprocal space. That is, one can write the equivalent of the operator (3) in reciprocal space as:

$$\begin{aligned}\Omega^{(-1)}(\mathbf{r}) = \gamma^{(-1)}(\mathbf{r}) &+ \sum_I w_I \rho_I^{(-1)}(\mathbf{r})\\ &+ \sum_I\sum_J w_I w_J \rho_I^{(-1)}(\mathbf{r})\rho_J^{(-1)}(\mathbf{r}) + O(3).\end{aligned}$$

And using the reciprocal basis set for describing molecules in the core set, one can write up to first order terms:

$$\begin{aligned}\forall I: p_I = \langle\Omega^{(-1)}\rho_I^{(-1)}\rangle &= \int_D \Omega^{(-1)}(\mathbf{r})\rho_I^{(-1)}(\mathbf{r})d\mathbf{r}\\ &= \langle\gamma^{(-1)}\rho_I^{(-1)}\rangle + \sum_J w_J\langle\rho_I^{(-1)}\rho_J^{(-1)}\rangle + O(2)\\ &\approx g_I^{(-1)} + \sum_J w_J z_{IJ}^{(-1)}.\end{aligned}$$

So, the first order fundamental QQSAR equation in reciprocal space will be expressed in terms of the linear system:

$$\begin{aligned}\forall I: p_I - g_I^{(-1)} = \pi_I = \sum_J z_{IJ}^{(-1)} w_J &\rightarrow \mathbf{Z}^{-1}|w\rangle = |\pi\rangle\\ \rightarrow |w\rangle = \mathbf{Z}|\pi\rangle &\rightarrow \forall I: w_I = \sum_J z_{IJ}\pi_J.\end{aligned} \tag{6}$$

2 Add-one-molecule-in procedures

2.1 Original space

For the procedures of adding a known molecule with unknown property to the core set, one can use the following partition of the involved matrices:

$$\mathbf{Z} = \begin{pmatrix}\mathbf{Z_0} & |\mathbf{z}\rangle\\ \langle\mathbf{z}| & \theta\end{pmatrix} \wedge \mathbf{Z}^{-1} = \begin{pmatrix}\mathbf{Z}_0^{(-1)} & |\mathbf{z}^{(-1)}\rangle\\ \langle\mathbf{z}^{(-1)}| & \theta^{(-1)}\end{pmatrix}$$

$$|\pi\rangle = \begin{pmatrix}|\pi_0\rangle\\ \pi\end{pmatrix} \wedge |\mathbf{x}\rangle = \begin{pmatrix}|\mathbf{x_0}\rangle\\ x\end{pmatrix} \wedge |\mathbf{w}\rangle = \begin{pmatrix}|\mathbf{w_0}\rangle\\ w\end{pmatrix}.$$

The elements of the inverse matrix can be easily obtained as the following three equations hold:

$$\begin{aligned}\mathbf{Z_0}\mathbf{Z}_0^{(-1)} + |\mathbf{z}\rangle\langle\mathbf{z}^{(-1)}| = \mathbf{I_0} &\rightarrow \mathbf{Z}_0^{(-1)}\\ &= \mathbf{Z}_0^{-1}\left(\mathbf{I_0} - |\mathbf{z}\rangle\langle\mathbf{z}^{(-1)}|\right),\\ \mathbf{Z_0}|\mathbf{z}^{(-1)}\rangle + \theta^{(-1)}|\mathbf{z}\rangle = |\mathbf{0}\rangle &\rightarrow |\mathbf{z}^{(-1)}\rangle = -\theta^{(-1)}\mathbf{Z}_0^{-1}|\mathbf{z}\rangle,\\ \langle\mathbf{z}\mid\mathbf{z}^{(-1)}\rangle + \theta^{(-1)}\theta = 1 &\rightarrow \theta^{(-1)} = \left(\theta - \langle\mathbf{z}|\mathbf{Z}_0^{-1}|\mathbf{z}\rangle\right)^{-1},\end{aligned} \tag{7}$$

which lead in turn to the result:

$$\mathbf{Z}_0^{(-1)} = \mathbf{Z}_0^{-1}\left(\mathbf{Z}_0 + \theta^{(-1)}|\mathbf{z}\rangle\langle\mathbf{z}|\right)\mathbf{Z}_0^{-1}.$$

Equations (4) and (6) when partitioned can be rewritten as:

$$\begin{aligned}&\mathbf{Z_0}|\mathbf{x_0}\rangle + x|\mathbf{z}\rangle = |\pi_0\rangle\\ &\langle\mathbf{z}\mid\mathbf{x_0}\rangle + x\theta = \pi\end{aligned}$$

and

$$\begin{aligned}&\mathbf{Z}_0^{(-1)}|\mathbf{w_0}\rangle + w|\mathbf{z}^{(-1)}\rangle = |\pi_0\rangle\\ &\langle\mathbf{z}^{(-1)}\mid\mathbf{w_0}\rangle + w\theta^{(-1)} = \pi,\end{aligned}$$

respectively.

Therefore, one can rearrange both equations in such a way that the unknown property in the original space could be expressed as:

$$\begin{aligned}|\mathbf{x_0}\rangle = \mathbf{Z}_0^{-1}(|\pi_0\rangle - x|\mathbf{z}\rangle) &\rightarrow \pi\\ &= \langle\mathbf{z}|\mathbf{Z}_0^{-1}|\pi_0\rangle + x\left(\theta - \langle\mathbf{z}|\mathbf{Z}_0^{-1}|\mathbf{z}\rangle\right).\end{aligned} \tag{8}$$

2.2 Reciprocal space

In reciprocal space, using the approximation

$$\mathbf{Z_0^{(-1)}} \approx \mathbf{Z_0^{-1}},$$

it can be written up to first order as:

$$\begin{aligned}|\mathbf{w_0}\rangle = \mathbf{Z}_0\left(|\pi_0\rangle - w|\mathbf{z}^{(-1)}\rangle\right) &\rightarrow \pi\\ &= \langle\mathbf{z}^{(-1)}|\mathbf{Z}_0|\pi_0\rangle + w\left(\theta^{(-1)} - \langle\mathbf{z}^{(-1)}|\mathbf{Z}_0|\mathbf{z}^{(-1)}\rangle\right).\end{aligned}$$

However, taking into account the remnant elements of the reciprocal augmented metric as written in Eq. (7),

$$\pi = \theta^{(-1)}\left(w\left(1 - \theta^{(-1)}\langle\mathbf{z}|\mathbf{Z}_0^{-1}|\mathbf{z}\rangle\right) - \langle\mathbf{z}\mid\pi_0\rangle\right).$$

This expression is slightly different from Eq (8), the one obtained within the original space settings. However, the exact form in reciprocal space can be written in the following manner: calling $\mathbf{A} = \mathbf{Z_0^{(-1)}} \rightarrow \mathbf{A}^{-1} = \left(\mathbf{Z_0^{(-1)}}\right)^{-1}$, then

$$|\mathbf{w_0}\rangle = \mathbf{A}^{-1}\Big(|\pi_0\rangle - w|\mathbf{z}^{(-1)}\rangle\Big) \to \pi$$
$$= w\Big(\langle \mathbf{z}^{(-1)}|\mathbf{A}^{-1}|\mathbf{z}^{(-1)}\rangle - \theta^{(-1)}\Big) - \langle \mathbf{z}^{(-1)}|\mathbf{A}^{-1}|\pi_0\rangle. \quad (9)$$

This possesses a form equivalent to Eq (8) with an obvious change of sign affecting the whole expression.

3 Formulation of the optimization problem

In any of both direct and reciprocal space cases, as expected from the linear structure of the fundamental equations used and provided that $\lambda \in \mathbf{R}$, for the unknown sought property it can be written as:

$$\pi = a + b\lambda. \quad (10)$$

Also, the equation for the core set unknowns can be written in general as:

$$|\mathbf{u}\rangle = \mathbf{A}(|\pi\rangle - \lambda\,|a\rangle), \quad (11)$$

where **A** is a positive definite matrix.

The unknown property in Eq (10) will be well-defined whenever, using Eq. (11), one is able to obtain a well-defined value of the parameter λ. Because the solution of Eq. (10) corresponds to an infinite collection of real elements, the restricted solution in the case of putting one molecule in is not unique, as from Eq. (11) one can describe several possible ways to obtain optimal values of the parameter λ . For instance:

(I) On defining the difference vector $|d\rangle = |\pi\rangle - \lambda\,|a\rangle$, a difference norm can be constructed:

$$\langle d \mid d\rangle = \langle \pi \mid \pi\rangle - 2\lambda\,\langle \pi \mid a\rangle + \lambda^2\,\langle a \mid a\rangle\,. \quad (12)$$

Optimizing expression (12) with respect to the parameter provides:

$$\lambda^{\text{opt}} = \frac{\langle \pi \mid a\rangle}{\langle a \mid a\rangle}.$$

Moreover, the optimal value of the difference norm will be a minimum, as the second order coefficient in Eq. (12) is a Euclidian norm of a non null vector.

(II) One can consider the norm of vector $|\mathbf{u}\rangle$, as defined in Eq. (11), as the objective function to be optimized; in this case it can be written:

$$\langle \mathbf{u} \mid \mathbf{u}\rangle = \langle \pi|\,\mathbf{A}\,|\pi\rangle - 2\lambda\,\langle \pi|\,\mathbf{A}\,|a\rangle + \lambda^2\,\langle a|\,\mathbf{A}\,|a\rangle\,.$$

So the optimal value of the parameter is now:

$$\lambda^{\text{opt}} = \frac{\langle \pi|\,\mathbf{A}\,|a\rangle}{\langle a|\,\mathbf{A}\,|a\rangle},$$

which provides a similar form as in the previous procedure, weighted by the transformation matrix **A**.

(III) The scalar product of the vectors $\{|\mathbf{u}\rangle\,;\,|\pi\rangle\}$ can be optimized, the objective function is now

$$|\langle \pi \mid \mathbf{u}\rangle|^2 = |\langle \pi|\,\mathbf{A}\,|\pi\rangle - \lambda\,\langle \pi|\,\mathbf{A}\,|a\rangle|^2$$
$$= |\langle \pi|\,\mathbf{A}\,|\pi\rangle|^2 - 2\lambda\,\langle \pi|\,\mathbf{A}\,|\pi\rangle\,\langle \pi|\,\mathbf{A}\,|a\rangle$$
$$+\lambda^2|\langle a|\,\mathbf{T}\,|a\rangle|^2,$$

producing

$$\lambda^{\text{opt}} = \frac{\langle \pi|\,\mathbf{A}\,|\pi\rangle}{\langle \pi|\,\mathbf{A}\,|a\rangle}.$$

(IV) The scalar product of the vectors $\{|\mathbf{u}\rangle\,;\,|a\rangle\}$ can be now optimized, in an equivalent way as in the previous procedure, that is, using the objective function:

$$|\langle t \mid \mathbf{u}\rangle|^2 = |\langle a|\,\mathbf{A}\,|\pi\rangle - \lambda\,\langle a|\,\mathbf{A}\,|a\rangle|^2$$
$$= |\langle a|\,\mathbf{A}\,|\pi\rangle|^2 - 2\lambda\,\langle a|\,\mathbf{A}\,|\pi\rangle\,\langle a|\,\mathbf{A}\,|a\rangle$$
$$+\lambda^2|\langle a|\,\mathbf{A}\,|a\rangle|^2,$$

which permits in obtaining the optimal value:

$$\lambda^{\text{opt}} = \frac{\langle a|\,\mathbf{A}\,|\pi\rangle}{\langle a|\,\mathbf{A}\,|a\rangle}.$$

This result, however, corresponds to the same restriction as the one previously studied in procedure (II). Thus, optimizing the norm $\langle \mathbf{u} \mid \mathbf{u}\rangle$ seems to be equivalent to optimizing the squared module $|\langle a \mid \mathbf{u}\rangle|^2$.

4 Quadratic restriction

In fact, all the previous procedures amount to the same as obtaining the optimal value of a second order polynomial in the parameter λ:

$$p(\lambda) = a + b\lambda + c\lambda^2. \quad (13)$$

Procedures (I)–(IV) correspond to obtaining unrestricted solutions of the optimal parameter. Thus, one can think of optimizing the polynomial (13) taking into account an additional restriction, which, in principle can be written, as well, as a second order polynomial:

$$r(\lambda) = \alpha + \beta\lambda + \gamma\lambda^2.$$

The problem can be solved by using an undetermined Lagrange multiplier procedure. Defining an augmented function like:

$$L[\lambda, \eta] = p(\lambda) - \eta[r(\lambda) - k],$$

where η is a Lagrange multiplier and k a constant the stationary condition with respect to the parameter λ provides the Euler equation:

$$\lambda^{\text{opt}} = \frac{\eta\beta - b}{2(c - \eta\gamma)}$$

and the Lagrange multiplier can be obtained from the equation:

$$\begin{aligned} r\left(\lambda^{\text{opt}}\right) = k \rightarrow \alpha + \beta\left(\frac{\eta\beta - b}{2(c-\eta\gamma)}\right) + \gamma\left(\frac{\eta\beta - b}{2(c-\eta\gamma)}\right)^2 \\ = k, \end{aligned} \tag{14}$$

yielding a null second order polynomial:

$$q(\eta) = t_0 + t_1\eta + t_2\eta^2 = 0,$$

with the coefficients defined as:

$$\begin{aligned} t_0 &= \gamma b^2 - 2(2(k-\alpha)c + \beta b)c \\ t_1 &= 2(4(k-\alpha)\gamma c + \beta(\beta c - (\gamma + 1)b)) \\ t_2 &= \gamma(4(\alpha - k)\gamma - \beta^2). \end{aligned}$$

Using elementary algebra,

$$\eta = (2t_2)^{-1}\left(-t_1 \pm \sqrt{t_1^2 - 4t_2t_0}\right)$$

will be obtained. Thus to assure a real result, the following has to be fulfilled:

$$t_1^2 \geq 4t_2t_0.$$

4.1 Simplifications

(1) A simpler result is obtained if, instead of condition (14), one just forces the second derivative of the augmented function to be a non-zero constant, say $2s$, then

$$\eta = \frac{c - s}{\gamma} \rightarrow \lambda^{\text{opt}} = \frac{(\beta c - \gamma b) - \beta s}{2\gamma s}$$

is obtained.

(2) Also, a simpler form can be obtained whenever $k = \alpha$ holds, then two values of the Lagrange multiplier are possible, as in the general case:

$$\eta_1 = \frac{b}{\beta} \wedge \eta_2 = 2\frac{c}{\gamma} - \frac{b}{\beta}.$$

Of course, in the former simplified case (1), the η_1 value is obtained when choosing:

$$s_1 = c - \frac{\gamma b}{\beta}$$

providing a trivial optimal parameter value

$$\lambda^{\text{opt}} = 0;$$

however, the η_2 value is obtained when choosing

$$s_2 = \frac{\gamma b}{\beta} - c = -s_1,$$

thus yielding an optimal parameter

$$\lambda^{\text{opt}} = -\frac{\beta}{\gamma}.$$

(3) When choosing the restriction polynomial as a first order one, then $\gamma = 0$, and thus

$$\begin{aligned} t_0 &= -2(2(k-\alpha)c + \beta b)c \\ t_1 &= 2\beta(\beta c - b) \\ t_2 &= 0. \end{aligned}$$

However, this will mean that

$$\eta = -\frac{t_0}{t_1} = \frac{(2(k-\alpha)c + \beta b)c}{\beta(\beta c - b)}.$$

So the optimal parameter can be written as:

$$\lambda^{\text{opt}} = \frac{2(k-\alpha)c^2 - b^2}{2(\beta c - b)c}.$$

Taking $k = \alpha$ will further simplify the result to

$$\lambda^{\text{opt}} = \frac{b^2}{2(b - \beta c)c}.$$

5 Concluding remarks

The previous computational scheme for the space, generated by the original core set, has been used within a preliminary rough test for a molecular core set presenting aquatic toxicity. Such a test will be schematically explained below as a practical way to describe the procedure, which will be implemented in the future.

The employed molecular set was proposed some years ago by Verhaar et al. in an initial paper [5] and studied under quantum similarity descriptor QSPR procedures in our laboratory

[6]. As the essential purpose of the present contribution is just setting up the appropriate theoretical framework under which novel QQSPR models can be easily obtained, the authors felt that this molecular core set could be an interesting and simple application example, which will be extended elsewhere and in company of other molecular core sets. The proposed molecular core set constitutes a toxicity case example, containing a series of 92 compounds, made up of benzene, toluene, xylene, phenol and aniline derivatives. The geometries of all the studied compounds have been previously optimized at the Hartree–Fock computational level with the program Gaussian 03 [7].

A necessary condition for the QQSPR algorithm to be applied is that the MQSM matrix has to be positive definite, otherwise the molecular core basis set will not be linearly independent and the inverse similarity matrix could be even singular. In order to obtain such a similarity matrix structure, previous to the similarity calculations, the molecules in the core set have to be oriented in space by a template-based TGSA search [8,9] or by any procedure, which can assure both optimal similarity measures and unique superposition geometries for every molecule [10].

Afterwards, in the practical example, Coulomb MQSM have been calculated according to a slightly modified Eq. (1), using as weight the Coulomb operator and by employing the geometries of the template oriented molecules. The restriction procedures I, II and III, mentioned in the present work, have been implemented for the purpose of obtaining a numerical estimate of the procedure performance. The estimated toxicity results have been obtained within an add-one-molecule-in scheme; thus, the only user-dependent parameter is the chosen number of molecules present in the core set in order to predict a given molecular toxicity.

The criterion, which can be used for deciding the optimal value of the core subset cardinality, is the r-square between the experimental values and the QQSPR predicted values, as it has been chosen in the present example. In the light of the obtained r-square values for the studied sample molecular set, one can conclude that the optimal number of molecules in the core-set becomes 3, ensuring in this way a slightly better result for each of the three employed procedures. Thus, for the molecular set in this case example, 91 toxicities are predicted, based on all possible combinations of 3 molecules out of the remaining 90 molecules. The predicted values adopted simply correspond to the arithmetic mean value obtained from these C(90;3) predictions. The r-square values for the three employed restriction algorithms are: $R^2(\mathrm{I}) = 0.71598$, $R^2(\mathrm{II}) = 0.71612$ and $R^2(\mathrm{III}) = 0.71691$. One can conclude, based on the toxicities computed in this manner and taking into account the large number of molecules studied, that a satisfactory predictive model has been reached.

It must be stressed that the present and future results will appear, without further manipulation, from the approximate solutions of the fundamental QQSPR equation, set up in turn from the MQSM matrix, associated with the chosen molecular set. The present practical results are quite remarkable, taking into account that no other data or descriptor manipulation than the raw MQSM over the template oriented molecules are employed to obtain the estimated property, in this particular schematic case: aquatic toxicity.

In the present proposed calculation procedure, the practical molecular descriptors are quantum density functions. Use has been also made of the fact that all the information one can obtain from a molecular quantum system is contained in such an individual density distribution.

It can be concluded that QQSPR, based on the quantum similarity theoretical framework, permits describing simple computational devices, which can be used to obtain adequate predictive models as an alternative to classical statistical procedures.

Acknowledgments One of us (R. C.-D.) expresses his acknowledgment to the Ministerio de Ciencia y Tecnología for the grant: BQU2003-07420-C05-01, which has partially sponsored this work. Also, the author has benefited from a Salvador de Madariaga fellowship, reference: PR2004-0547, which has permitted his stay at the Ghent University during the period October 2004–October 2006. The author finally acknowledges the useful discussions with Prof. Patrick Bultinck from Ghent University, without whose deep remarks on the fundamental QQSPR equation, this work would not have even started.

The research of one of the authors (S. V. D.) is funded by a PhD. grant of the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen).

References

1. Carbó-Dorca R, Robert D, Amat Ll, Gironés X, Besalú E (2000) Molecular quantum similarity in QSAR and drug design. In: Lecture notes in chemistry, vol 73. Springer, Berlin
2. Bultinck P, Girones X, Carbó-Dorca R (2005) Molecular quantum similarity: theory and applications. In: Lipkowitz KB, Larter R, Cundari T (eds) Reviews in computational chemistry, vol 21. Wiley, Hoboken (USA), pp 127–207
3. Bultinck P, Carbo-Dorca R (2004) J Math Chem 36: 191–200
4. Carbó-Dorca R (2004) J Math Chem 36(3): 241–260
5. Verhaar HJM, Mulder W, Hermens JLM, Rorije E, Langenberg JH, Peijnenburg WJGM, Sabljic A, Güsten H, Eriksson L, Sjöström M, Müller M, Hansen B, Nouwen J, Karcher W (1995) Overview of structure–activity relationships for environmental endpoints. Part 1: general outline and procedure. In: QSAR for predicting fate and effects of chemicals in the environment. Report EU-DG-XII Project. Contract EV5V-CT92–0211. European Union, Brussels, Belgium
6. Robert D, Carbó-Dorca R (1999) SAR QSAR Environ Res 10: 401–422
7. Gaussian 03, Revision B.05, Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Bakken V, Adamo C,

Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2004) Gaussian, Inc, Wallingford CT

8. Gironés X, Robert D, Carbó-Dorca R (2001) J Comp Chem 22(2): 255–263
9. Gironés X, Carbó-Dorca R (2004) J Comp Chem 25: 153–159
10. Carbó-Dorca R (2007) A quantum similarity matrix (QSM) Aufbau procedure. J Math Chem (submitted)