ELSEVIER

Principles for 3D/4D QSAR classification of drugs

Vladimir Potemkin and Maria Grishina

Chelyabinsk State University, Br. Kashirinych 129, Chelyabinsk 454021, Russian Federation

The principles for the 3D/4D classification of drugs are introduced in this article. Based on these principles, new techniques for the reconstruction of complementary selfconsistent receptor fields for the classification of drugs, taking into account their multitautomeric and multiconformational states, are created. The series of examples of classification of drugs by their activity (active or nonactive), by their mechanisms of action, by their target and binding site and by the most important stages of their action are given. Prospects for rational drug design are highlighted.

Introduction

There are many methods that can be used for the classification of drugs. A lot of them such as pattern recognition, cluster analysis, discriminant analysis, neural network and so on, have been used for the analysis and prediction of biological action of molecules [1–10].

Generally, the methodology of all the methods relies upon the following approaches: topological, geometrical, quantum, ADMET (adsorption, distribution, metabolism, excretion and toxicity) and so on. Molecular descriptors are computed for a set of molecules with known activities (a training set). The algorithm creates a model or a decision rule which recognizes active and nonactive molecules. The decision rule (the model) obtained from this process can then be used for the prediction of the bioactivities of new compounds. Sometimes we cannot understand the physical meaning of some models, their relationship to chemical or biological phenomena, with the mechanism of their action but they can be used, since they have good agreement with experimental activities.

At present, the newest 3D/4D descriptors include, for example, PEST (Property-Encoded Surface Translator) [11], WHIM (Weighted Holistic Invariant Molecular) descriptors [12], QuaSAR, descriptors [13], Shape Signatures [14] and so on. Moreover, new additions for CoMFA [15], CoMSIA [16], GRID [17], Raptor [18] and some other applications for classification have been developed (e.g. [19,20]).

Corresponding author: Potemkin, V. (pva@csu.ru)

Usually, the 3D/4D classification methods give crossvalidation quality of less than 0.80 for several datasets (e.g. [18,21]). For some cases the crossvalidation quality is significantly lower (e.g. 0.665 [18], 0.077 and −0.024 [22]). Moreover, even for the best cases we see a realization of the Kubinyi paradox [23] that prevents the usage of the methods/descriptors in the practical classification. Although, the human apparatus of recognition is very powerful, in that we can distinguish differences in the superfine features, it is difficult for us to distinguish a drug in billions of compounds and it is also a difficult thing to teach a machine to do.

The reasons why the recognition of molecules is difficult lie both in the algorithms and in the descriptors. Could you recognize a person in a crowd knowing only their weight, height and the size of their shoes? The probability of error is very high when you know only a few common characteristics.

With respect to recognizing a person in a crowd: do you imagine that you would be more successful in recognizing a person in a crowd by looking at a photograph? Most would agree that the effectiveness of your search will be much better. Why should this be the case? It is because you would be looking at their most recognizable feature – the face. Moreover, you see not only individual features, such as eyes, nose, mouth and ears separately, but also their mutual alignment. The picture is far superior to any verbal or digital description. It is significant that when one has a photo, there is no need to presuppose the exact height, weight and the size of the person's shoes, but we can recognize the person very easily.

1359-6446/06/$ - see front matter doi:10.1016/j.drudis.2008.07.006

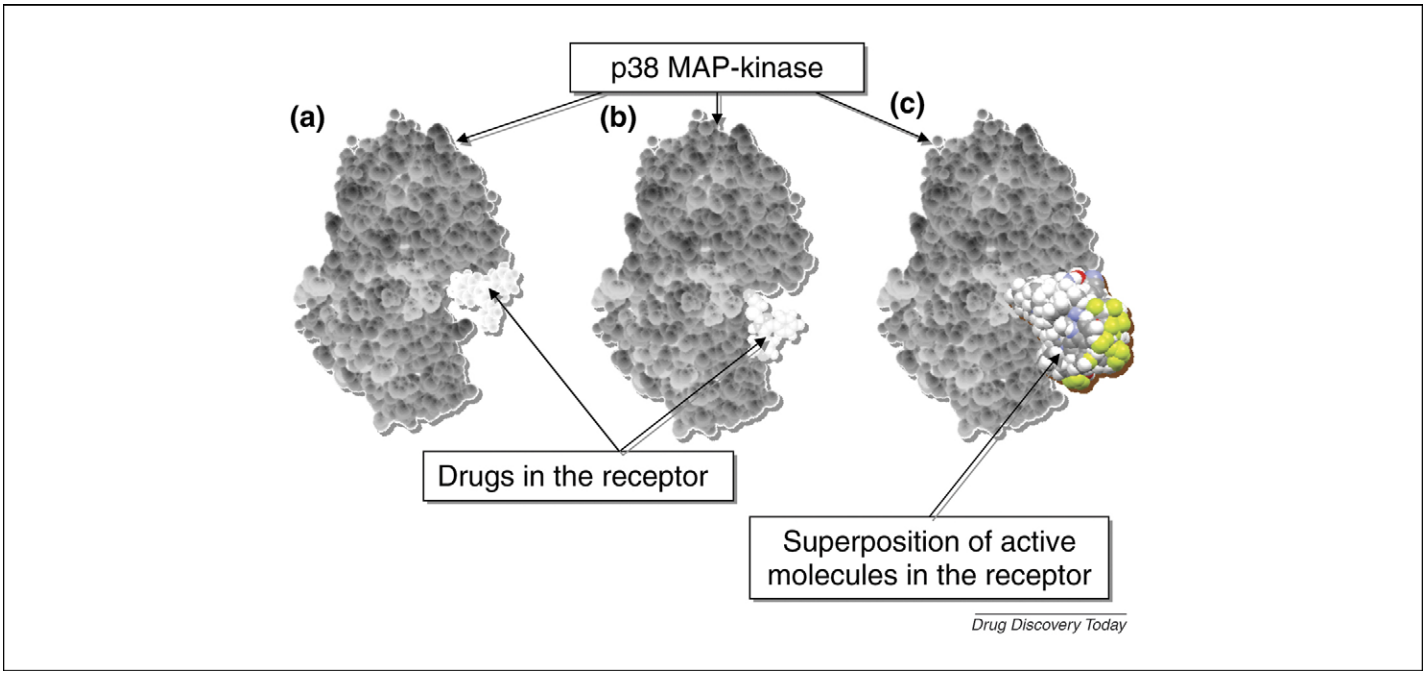

FIGURE 1

Complexes of p38 MAP-kinase with its inhibitors. (**a**) The inhibitor aligning to the top part of the receptor pocket; (**b**) the inhibitor aligning to the bottom part of the receptor pocket; (**c**) the superposition of the active molecules in the receptor cavity.

In an analogous manner, can the effectiveness of drug classification be high if we only use its verbal or digital description with common molecular characteristics? It may be better to look at the 'photograph' of a molecule, consider the whole image and the mutual bracing of its parts.

A molecule seems to be quite a simple thing. It consists of nuclei and electron clouds and we can describe the location of the nuclei and the characteristic points of electron density. The location of the nuclei is very important information, since it defines molecular geometry, which can be either appropriate or inappropriate for a receptor.

Principles for 3D/4D classification

Biological action results from the interaction of a molecule with some component of an organism (ferments, receptors, peptides, enzymes, DNA, water, membrane and so on). Therefore, it is necessary to identify potentially active centers of molecular interactions. It is possible to determine the potential of electrostatic, Van der Waals interactions, hydrogen bonding and so on, at each point of the molecular surface. These potentials determine a molecular field. Based on the complementarity concept [24] we can suppose that the field of a good drug must be complementary to the receptor. The more complementary a molecule is to the receptor, the more likely it is to be active. Therefore, the molecular field of an ideal drug is, ideally, complementary to the field of the receptor. Thus, we can reconstruct the field of the receptor as a complementary one to the field of the ideal drug (as its mould, its negative). It is possible then to recognize a new prospective drug as a complementary one to the derived mould. We do not, however, have an ideal drug. Moreover, there seems little use in creating a new drug if one already has the ideal one. What we do have, however, are real drugs and it is necessary to reconstruct the field of the receptor using the fields of these real molecules.

It should be noted that the field of a real molecule can define only a part of the receptor active site. An example is given in Fig. 1, in which one molecule aligns to the top part of the receptor pocket (Fig. 1a) while another aligns to the bottom (Fig. 1b). At the same time, both molecules are effective p38 MAP kinase inhibitors. Therefore, it can be assumed that a generalized set of active molecules can completely reconstruct the receptor active site as a complementary field (Fig. 1c). We need to bear in mind that a molecule includes active (pharmacophoric) fragments interacting with the receptor, nonactive fragments that do not take part in the interaction and fragments disturbing the interaction (e.g. providing steric barriers). In addition, we must also bear in mind the flexibility of a molecule and the flexibility of the receptor. Consequently, not only does a molecule have to adjust to the receptor, but the receptor also has to adjust to the molecule, due to their inherent flexibilities.

So, the principles for 3D/4D classification are as follows:

- The geometry and the distribution of the interaction centers of an active molecule must be appropriate to the target.
- The classification method must find molecules with complementary geometry and active centers of distribution provided by their conformational and tautomeric state.
- The 'photograph' of a molecule must be representative. The method must recognize an active molecule, irrespective of its foreshortening, its conformational and tautomeric state.

Algorithms

Based on these assumptions a new method for classification of compounds has been introduced. 3D/4D-QSAR BiS algorithm (*Bi*ological *S*ubstrate *S*earch) [25–35] underlies the new method.

The method includes the following main blocks:

1. The structural input of a set of molecules.
2. The creation of a pseudo-atomic environment.

3. The orientation and alignment of the molecules.
4. The adjustment of the potentials of the pseudo-atomic field.
5. The adjustment of pseudo-atomic characteristics and positions.
6. The determination of energies, forces and force constants of the interaction in the 'pseudo-receptor–ligand' complex.
7. The definition of a decision rule as a function of the interaction parameters.

As a result we have aligned and oriented molecules, produced a mould/negative of the active molecules for the classification of compounds, aligned and oriented molecules, and recommendations for rational drug design.

At the first stage a molecule is represented as a space-filling model with atomic radii computed by MERA (*M*odel of *E*ffective *R*adii of *A*toms) [36,37]. The electrostatic and Van der Waals potentials are computed [37] in each point of the molecular surface using the equations:

$$\varphi_m^q = \sum_{i=1}^{N} \frac{q_i}{R_{im}} c; \qquad \varphi_m^{\mathrm{VDW}} = -2\sum_{i=1}^{N} V_{im} \frac{2^3 r_i^3}{R_{im}^6}$$

where N is the number of atoms in the molecule; q_i is the charge on the atom i; R_{im} is the distance from a point m to atom i; c is a scaling coefficient; V_{im} is the potential energy minimum of the Lennard–Jones equation; r_i is the Van der Waals radius of atom i. V_{im} and q_i can be calculated with MERA [36,37]. To simplify the calculations, the Van der Waals potential includes only the attractive part.

The complementary field of a receptor must have complementary characteristics in each point that can be modeled using pseudo-atomic simulation. A pseudo-atom can be placed in each point of the molecular surface. The charge and the radius of the pseudo-atom providing the complementary field in the point can be computed by the following equations:

$$q_m = -\frac{\varphi_m^q}{\sum_{i=1}^{N} c/R_{im}}; \qquad r_m = \sqrt[3]{\frac{\varphi_m^{\mathrm{VDW}}}{-2^3 \sum_{i=1}^{N} 2V_{im}/R_{im}^6}}$$

Then, the orientation of the second molecule is optimized in the obtained field by the combined simplex and quasi-Newton methods to reach the minimal overall probability (P) of the contact of its atoms with all the pseudo-atoms:

$$P = 1 - \prod_{m=1}^{M} (1 - p_m),$$

$$p_m = \exp\left(-\frac{E_m}{RT}\right)$$

where M is the total number of pseudo-atoms.

$$E_m = \sum_{i=1}^{N} \left(\frac{cq_i q_m}{R_{im}} - 2V_{im} \frac{(r_m + r_i)^6}{R_{im}^6} + V_{im} \frac{(r_m + r_i)^{12}}{R_{im}^{12}} \right)$$

Then the complementary receptor field is adjusted by the addition of the potentials of the second molecule:

$$\varphi_m^q = \varphi_m^q + \varphi_m^{q'} \quad \text{and} \quad \varphi_m^{\mathrm{VDW}} = \varphi_m^{\mathrm{VDW}} + \varphi_m^{\mathrm{VDW}'}$$

The field potentials $\varphi_m^{q'}$ and $\varphi_m^{\mathrm{VDW}'}$ of the second molecule are calculated similarly to the first molecule.

The locations of the pseudo-atoms are also changed in agreement with the exterior of the second molecule. Therefore, the pseudo-receptor is modeled as a flexible one. Then, the analogous procedures are carried out for the third, fourth and so on, molecules of the dataset. Finally, when the last molecule is optimized we have a generalized field of the full dataset. Since the receptor is flexible, it is possible to compute the force constant κ_m for each pseudo-atom m, knowing its movement amplitude A_m. In agreement with the postulate on the uniform distribution of degrees of freedom:

$$\frac{\kappa_m A_m^2}{2} = \frac{kT}{2}$$

where k is a Boltzmann constant and T is the absolute temperature.

Therefore,

$$\kappa_m = \frac{kT}{A_m^2}$$

Thus, the energies of the interactions in the 'pseudo-receptor–ligand' complexes and the first and the second derivatives of the energies are computed. Therefore, it is possible to model all kinds of molecular movement (translational, rotational and vibrational). Since the initial receptor model has been changed, all the procedures (the orientation of molecules and the adjustment of the potentials) in the newly obtained receptor field must be carried out once again from the first to the last molecule, taking into account the possible movements of the molecules. The iterative sequence is finished when the differences between the pseudo-atomic characteristics (radii, charges and coordinates) of the current and previous iteration are less than accepted values.

Therefore, the initial aim is achieved that is the complementary generalized selfconsistent field is created. Then a decision rule in the form of the desirability function [38,39] depending on interaction characteristics is defined as:

$$P = \exp\left[-\exp\left\{0.834 - 3.08\left(a_1 + a_2\sum_{l=1}^{L} E_l + a_3\sum_{l=1}^{L} F_l\right)\right\}\right]$$

where E_l is the interaction energy of pseudo-atom l with a molecule (L is the total number of the pseudo-atoms) and F_l is the force affecting the pseudo-atom l. F_l is computed by the formula:

$$F_l = \sum_{i=1}^{N} \frac{\mathrm{d}E_{il}}{\mathrm{d}R_{il}},$$

where N is the total number of atoms in molecule; E_{il} is the interaction energy of atom i and pseudo-atom l; R_{il} is the distance between atom i and pseudo-atom l; a_1, a_2 and a_3 are the parameters.

The advantages of the desirability function are as follows:

- A small change of the argument leads to the great change in the function value, that permits the precise distinction of active molecules from inactive ones.
- The value area of the function is [0;1], therefore, the function does not predict the quantitative value of bioactivity but gives probability of activity (P). In the case where $P_k > 0.5$, a molecule is considered to be active (Class = 1), in other cases the molecule would be considered inactive (Class = 0). Thus, the method makes the categorical classification into active and inactive compounds.

Then, we can try to place a new molecule into the obtained pseudo-atomic receptor, orientate it, compute the interaction

characteristics and, using the desirability function, we can decide whether the molecule is active or not.

In general, the geometry of molecules in receptor pockets and in aqueous solution is different from their global minimum geometry [37]. Therefore, it is necessary to take into account the flexibility of the molecules, their conformational multiformity and their ability to adjust to the receptor. In fact, BiS works as a multiconformational approach. For this aim, a combination of BiS and multiconformational representation of molecules realized in MultiGen [37] software is used. In this software, a conformational search is performed along vibrational modes within accepted energy range relatively to the lowest energy conformer. Then, the effective superposition of the obtained conformers is created using BiS. The biological activity of a compound is determined as a superposition of conformers partial activities (A_i) taking into account the probabilities (p_i) of their existence.

$$\mathrm{BA} = \sum_{i=1}^{K} p_i A_i$$

where K is the total number of conformers.

The probabilities are proportional to the exponent of their energies [37].

$$p_i = \frac{\exp(-E_i/kT)}{\sum_{j=1}^{K} \exp(-E_j/kT)}$$

It has been demonstrated that the results of the conformational search are in complete agreement with NMR and X-ray data [31]. Thus, the multiconformational representation of a molecule as a set of atomic coordinates for each conformer and probabilities of location of each atom in these points of space can be constructed. So, each atom of each conformer is characterized by four parameters: three Cartesian coordinates along with the probability. From this point of view, a molecule has a four-dimensional representation. Using this approach we can choose the most active conformers (the conformer with maximal A_i value) for more detailed consideration of mechanism of the action. Moreover, it has been demonstrated that the multiconformational approach gives significantly better results in predictions of bioactivity [27,32].

Molecules can very often exist in various tautomeric forms and it is necessary to take into account both their tautomeric and conformational states. In such cases, the multiconformational analysis must be applied for each tautomeric form. Then, the bioactivity of a compound is determined as a superposition of the partial activities of the conformers for all tautomeric forms of the molecule taking into account the probabilities of their existence:

$$\mathrm{BA} = \sum_{j=1}^{T} \sum_{i=1}^{K_j} p_{ij} A_{ij}$$

where T is the total number of tautomeric forms of a compound; K_j is the total number of conformers for tautomer j; p_{ij} is the probability of existence of conformer i in tautomeric form j; A_{ij} is the partial activity of tautomer–conformational form ij.

The software gives not only high classification quality, but also the orientations of molecules in the pseudo-receptor that agrees with experimental data. The orientation obtained within the algorithm is compared to the X-ray data of the 'receptor–ligand' complexes for the inhibitors of p38 MAP kinase, CDK2, thermolysin, DHFR, HRV14, viral neuraminidase and for DNA-antimetabolites. A pure correlation between the computed and the experimental coordinates has been demonstrated. The root mean square deviations (RMSD) between them are usually less than 1 Å [31,33]. For the same datasets, other approaches give significantly worse results, for example, the RMSD presented in the paper [40] for some of the above-listed datasets exceeds 18 Å. The good orientation results obtained by BiS show the ability of its usage as a molecular docking method.

There are many different variants of classification. Compounds can be classified by bioactivity (active or nonactive), by the mechanisms of action, by the binding site, by the most important stage of the action and so on. All the above-listed kinds of classification were performed using the BiS algorithm. The algorithm has been tested using more than a hundred datasets of compounds of various kinds of bioactivity [25–35]. The crossvalidation quality of the classification is usually not less than 0.9 for several datasets. Some examples of the classifications are presented in the next item.

3D/4D classification

The study of anti-inflammatory agents with different mechanisms of action (cyclooxygenase-1 (COX-1) inhibitors, cyclooxygenase-2 (COX-2) inhibitors and lipooxygenase (LOX) inhibitors) [41,42] shows the following: the characteristic features of the complementary fields (Fig. 2a–d) for all of the anti-inflammatory agents are as follows: the left, negatively charged part, lipophilic and positively charged central parts. At the same time, there are some differences in the fields: cyclo- and lipooxygenase inhibitors have the negatively charged area at the top right part of the model receptor, while p38 MAP-kinase inhibitors have the lipophilic area at the very same part. The right bottom part of the receptors is very different: it is positive for COX-1 and p38 MAP-kinase inhibitors, lipophilic for COX-2 inhibitors and negative for LOX inhibitors. The pseudo-receptor for COX-2 inhibitors is the most compact while their positive central part is the largest. Thus, it is possible to classify anti-inflammatory drugs by their mechanisms of action using the obtained pseudo-receptor. One can place an anti-inflammatory drug into the obtained pseudo-receptors, estimate energies and forces of the interactions and find out the most appropriate pseudo-receptor for the compound. The crossvalidation qualities of bioactivity classification for all the mechanisms are not less than 0.97 (or 97%).

Having the force constants of the interactions with pseudo-atoms it is possible to determine the rigid and the flexible part of the pseudo-receptor (Fig. 2e). This is very important information, since it shows the fragments of the molecules that cannot be changed without loss of bioactivity. These fragments are located near rigid (green) parts. As opposed to that, the fragments lying near soft (yellow) parts can be changed to increase bioactivity. Moreover, it is clear what replacements can be performed. The soft parts coincide with the negatively charged areas of the pseudo-receptor and we can improve the bioactivity by substituting these positions by positively charged fragments. Moreover, the method determines those pseudo-atoms producing incremental increases in the interaction energy. These pseudo-atoms are represented in Fig. 2f,j. The gaps in the receptor are also evident: these parts of the molecule have no influence on the bioactivity and they can be

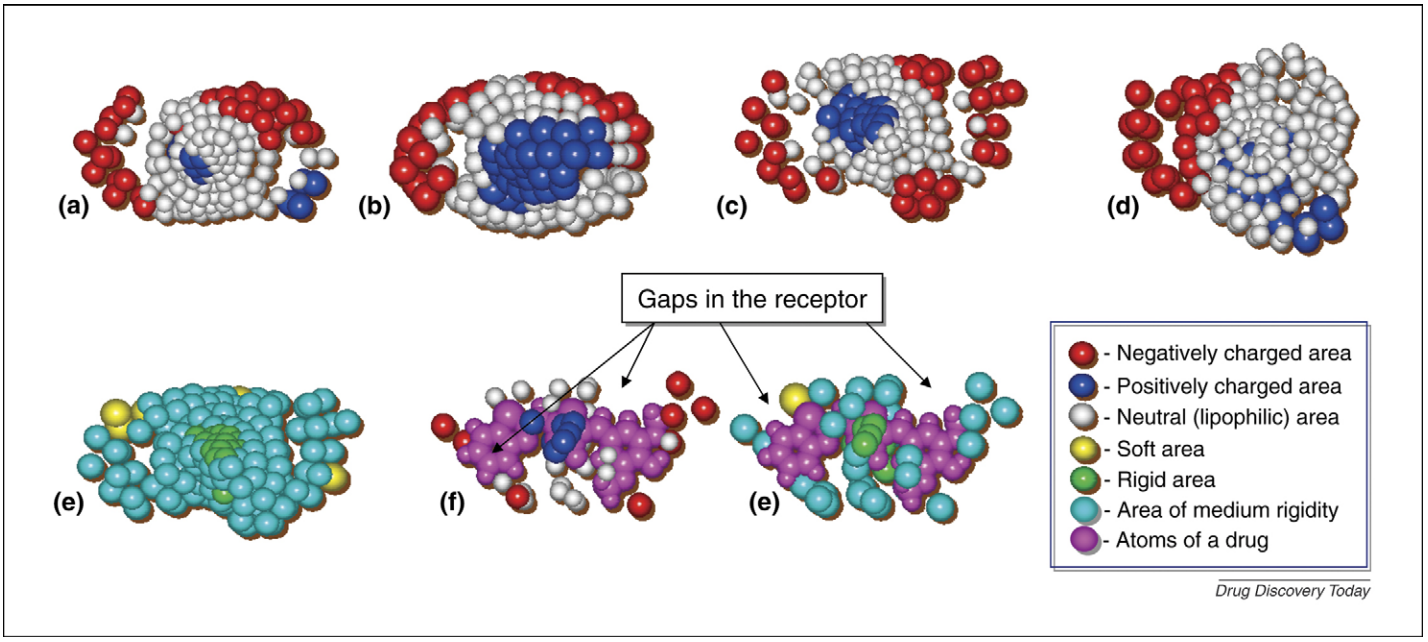

FIGURE 2

The complementary selfconsistent fields for: (**a**) cyclooxygenase-1 inhibitors; (**b**) cyclooxygenase-2 inhibitors; (**c**) lipooxygenase inhibitors; (**d**) p38 MAP-kinase inhibitors; (**e**) rigid and soft parts of pseudo-receptor for lipooxygenase inhibitors; (**f**) the pseudo-atoms incrementing mostly in the interaction energy (charge distribution); (**g**) the pseudo-atoms incrementing mostly in the interaction energy (rigidity).

modified to obtain a drug with improved ADMET-characteristics, for example. Therefore, a drug design procedure is realized in the software. The software suggests the variants of substitutes for each atom in the molecule to create more active compounds and the variants of the substitutes to create less active compounds if it is necessary to suppress an undesirable activity. The software makes the replacements, places new molecules into the pseudo-receptor and computes their activities. These replacements are made automatically and, in a few minutes, one has the information for further synthesis of prospective drugs.

Frequently, molecules of a dataset can interact with different binding sites. This situation can be illustrated using the example of

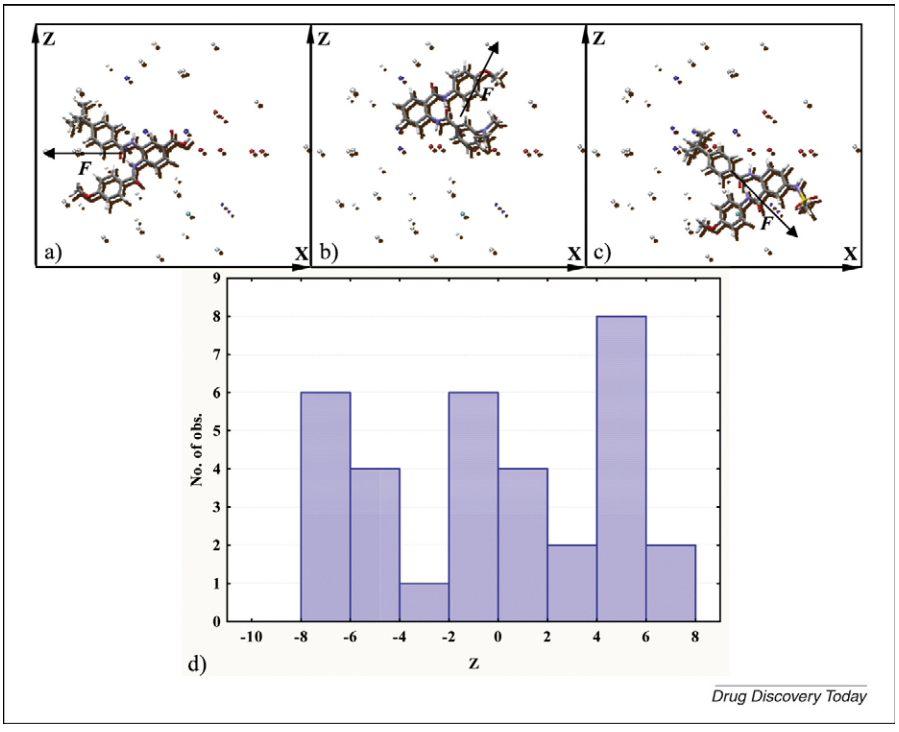

FIGURE 3

Human factor Xa inhibitors aligning: (**a**) to the left part of pseudo-receptor; (**b**) to the right top part of pseudo-receptor; (**c**) to the right bottom part of pseudo-receptor; (**d**) the distribution of centers of masses of human factor Xa inhibitors along Z-axis. *F* – directions of forces.

the human factor Xa inhibitors [43,44]. The study of the dataset within BiS shows that the molecules are located at three various sites of the pseudo-receptor (Fig. 3). It is possible to classify them simply either by coordinates or by interaction energies or by forces. Actually, the human factor Xa has three different binding sites which were experimentally confirmed and BiS gives faultless classification [25]. Thus, one pseudo-receptor can simulate an interaction with various binding sites and with various receptors, including nontarget competing ones.

The presented method can be used for consideration of multistage bioprocesses, since the algorithm models not only the 'receptor–ligand' interaction, but also the dynamics of molecular movement. The most illustrative example is the study of the biological action of tuberculostatic dihydrofolatereductase inhibitors [33,29] (Fig. 4).

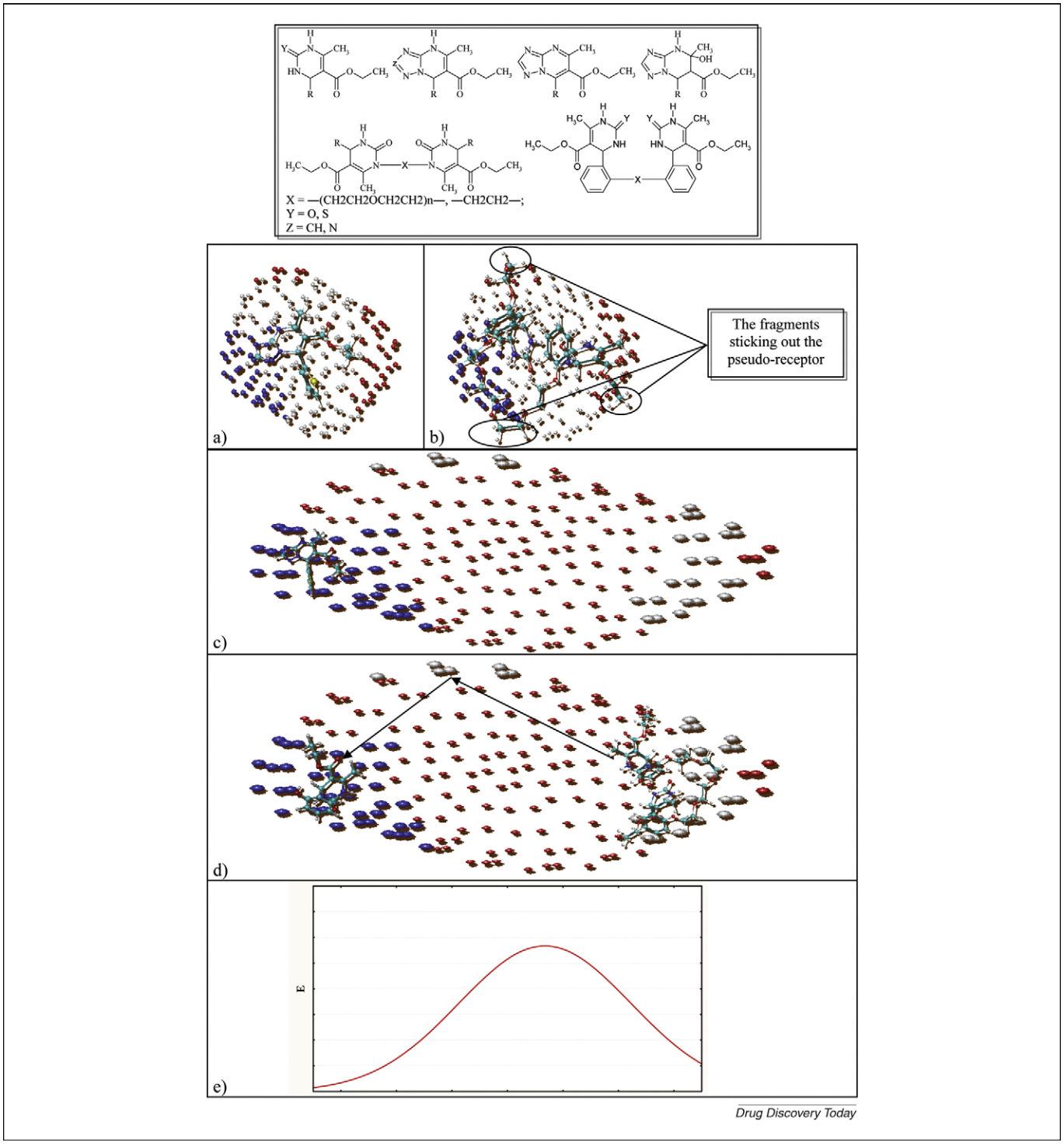

FIGURE 4

"Pseudo-receptor–ligand" complexes for dihydrofolatereductase inhibitors. (**a**) The typical complex for active non-podand inhibitors; (**b**) "pseudo-receptor–ligand" complex for molecules with podand chain; (**c**) complex of extended pseudo-receptor with non-podand molecule; (**d**) "extended pseudo-receptor–ligand" complex for molecules with podand chain (arrows show the pathway to binding site); (**e**) energy profile for movement of molecules with podand chain.

The dataset includes Biginelli compounds, their derivatives with podand chain and their pyridine analogs. The modeling with BiS shows that some of the podands do not fit well within the model receptor. Some fragments of the chain stick out from the receptor (Fig. 4b). The results have been confirmed using molecular docking: some active molecules cannot be fitted in the receptor pocket. Therefore, there exists a paradox: a molecule may have high activity, but cannot be placed within the receptor pocket. An explanation has, however, been proposed. The left (blue) part of the receptor provides the most strong interactions with the most active molecules. It is possible to stretch the remaining part. The extended pseudo-receptor has very interesting characteristic features: the lipophilic parts at the right side and at the top part of the model receptor. The central part is represented by a more polar region. The left hand side shown in the diagram represents the binding site obtained at the first stage of modeling. The active, nonpodand molecules are located at the binding site (Fig. 4c) while the podands are concentrated at the right lipophilic part (Fig. 4d) and their movement into the binding site is unfavorable (Fig. 4e). Molecules surmounting the obstacles are moved to the top lipophilic area. Metabolism of these compounds may occur in this area, since the forces affecting the podand chain and the heterocyclic fragment act in opposite directions. The heterocyclic part of the molecules must reside at the binding site. In this case, the podands act as prodrugs. It is also presupposed that the right side represents the cell membrane. In this case, the podand chain plays a membranotropic role. Then, it undergoes elimination and the active molecule is produced (Fig. 4d). An additional experimental confirmation of the hypothesis is that some podands are active in *Mycobacteria tuberculosis* cells, but not active on isolated dihydrofolatereductase.

On the other hand, some small molecules are active at dihydrofolatereductase but inactive in cells. This phenomenon can be easily explained: some small polar molecules have a barrier, too, but it is a barrier for their penetration through membranes. These molecules must be active *in vitro* at dihydrofolatereductase but inactive in the cells. These results are in complete agreement with the experimental data. Thus, the method can be used for the study of multistage bioprocesses.

Conclusions

Thus, the principles for 3D/4D classification of drugs have been introduced. Based on these principles, the new multitautomer–conformational methodologies for drug classification have been created. Using the series of examples it has been shown that the proposed methods allow the classification of drugs by their bioactivity (active or nonactive), mechanisms of action, their target and binding site and the most important stages of their action. The proposed algorithm can also be used as the molecular docking approach for the study of the interactions in 'receptor–ligand' complexes. Therefore, the proposed principles and methodologies for 3D/4D classification of drugs give new facilities for the fundamental analysis of the mechanisms of biological action, for practical virtual screening and for rational drug design.

Acknowledgements

Russian Foundation for Basic Researches (grants 07-03-96041, 07-04-96053, 08-03-08015 and 08-03-08016), SKIF – GRID Supercomputer Initiative, 'Human Capital Foundation' (London, Great Britain).

References

1 Zupan, J. and Gasteiger, J. (1999) *Neural Networks in Chemistry and Drug Design: An Introduction*. Wiley
2 Tetko, I.V. *et al.* (1999) Pharmaceutical fingerprinting in phase space. 2. Pattern recognition. *Anal. Chem.* 71, 2431–2439
3 Cruciani, G., Mannhold, R., Kubinyi, H., Folkers, G., eds) (2005) *Molecular Interaction Fields: Applications in Drug Discovery and ADME Prediction*, Wiley
4 (2008) *Molecular Design: Concepts and Applications*. Wiley
5 Qin, L. *et al.* (2008) Genetic algorithm based neural classifiers for factor subset extraction. *Soft Comput. Fusion Found. Methodol. Applic.* 12, 623–632
6 El-Solh, A. *et al.* (2004) Predicting active pulmonary tuberculosis using an artificial neural network. *Medinfo* 116, 968–973
7 Zhang, P. *et al.* (2005) Neural vs. statistical classifier in conjunction with genetic algorithm based feature selection. *Pattern Recognit. Lett.* 26, 909–919
8 Khandelwal, A. *et al.* (2007) Computational models to assign biopharmaceutics drug disposition classification from molecular structure. *Pharm. Res.* 24, 2249–2262
9 Deconinck, E. *et al.* (2006) Classification tree models for the prediction of blood–brain barrier passage of drugs. *J. Chem. Inf. Model.* 46, 1410–1419
10 Givehchi, A. *et al.* (2004) Impact of descriptor vector scaling on the classification of drugs and nondrugs with artificial neural networks. *J. Mol. Mod.* 10, 204–211
11 Breneman, C.M. *et al.* (2004) New developments in PEST shape/property hybrid descriptors. *J. Comput. Aid. Mol. Des.* 17, 231–240
12 Gramatica, P. (2006) WHIM descriptors of shape. *QSAR Combinat. Sci.* 25, 327–332
13 Manoj, K.P. *et al.* (2006) Quantitative structure–activity relationships of selective antagonists of glucagon receptor using QuaSAR descriptors. *Chem. Pharm. Bull.* 54, 1586–1591
14 Meek, P.J. *et al.* (2006) Shape signatures: speeding up computer aided drug discovery. *Drug Discov. Today* 11, 895–904
15 Advanced CoMFA, Tripos
16 QSAR with CoMFA, Tripos
17 Von Itzstein, M. *et al.* (1993) Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature* 363, 418–423
18 Vedani, A. *et al.* (2006) Simulating α/β selectivity at the human thyroid hormone receptor: consensus scoring using multidimensional QSAR. *ChemMedChem* 2, 78–87
19 Kasabov, N. (2007) Global, local and personalised modeling and pattern discovery in bioinformatics: an integrated approach. *Pattern Recognit. Lett.* 28, 673–685
20 Davies, E.K. *et al.* (2002) Pattern recognition and massively distributed computing. *Int. J. Comput. Chem.* 23, 1544–1550
21 Papa, E. *et al.* (2008) Prediction of PAH mutagenicity in human cells by QSAR classification. *SAR QSAR Environ. Res.* 19, 115–127
22 Ohgaru, T. *et al.* (2008) Ordinal classification using comparative molecular field analysis. *J. Chem. Inf. Model.* 48, 207–212
23 Van Drie, J.H. (2003) Pharmacophore discovery – lessons learned. *Curr. Pharm. Des.* 9, 1649–1664
24 Bohr, N.H.D. (1948) On the notions of causality and complementarity. *Dialectica* 2, 312–319
25 Potemkin, V.A. *et al.* (2001) An alternative method for 3D-QSAR and the alignment of molecular structures: BiS (Biological Substrate Search). In *Rational Approaches to Drug Design* (Holtje, H.-D. and Sippl, W., eds), pp. 349–353, Prous Science Publishers
26 Potemkin, V.A. *et al.* (2002) Quantitative relationship between structure and antibacterial activity of quinolone derivatives. *Pharm. Chem. J.* 36, 22–25
27 Potemkin, V.A. *et al.* (2002) Multiconformational method for analyzing the biological activity of molecular structures. *J. Struct. Chem.* 43, 1045–1049
28 Mikuchina, K. *et al.* (2002) 3D QSAR analysis and pharmacophore modelling of p38 MAP kinase inhibitors using BiS algorithm. *Arch. Pharm. Pharm. Med. Chem.* 335, C74
29 Potemkin, V.A. *et al.* (2003) Theoretical investigation of the antituberculous activity of membranotropic podands. *Pharm. Chem. J.* 37, 468–472

30 Grishina, M.A. *et al.* (2005) Theoretical study of the substrate specificity of cytochrome P-450 isoforms. *Pharm. Chem. J.* 39, 509–513
31 Potemkin, V.A. *et al.* (2007) Modeling of drug molecules orientation within a receptor cavity in the BiS algorithm framework. *J. Struct. Chem.* 48, 155–160
32 Grishina, M.A. *et al.* (2007) A novel techniques for virtual discovery for study of multistage bioprocesses. *Drugs Future* 32 (Suppl. A), 27
33 Pereyaslavskaya, E.S. *et al.* (2007) The analysis of pharmacophoric parts of DHFR – inhibitors using 3D-QSAR algorithm BiS/MC and X-ray data. *Drugs Future* 32 (Suppl. A), 87
34 Grishina, M.A. *et al.* (2007) A novel approach to pattern recognition for drugs. *Drugs Future* 32 (Suppl. A), 112
35 Potemkin, V.A. *et al.* (2008) A new paradigm for pattern recognition of drugs. *J. Comput. Aid. Mol. Des.* 10.1007/s10822-008-9203-x http://www.springerlink.com
36 Potemkin, V.A. *et al.* (1996) A new approach to predicting the thermodynamic parameters of substances from molecular characteristics. *Rus. J. Phys. Chem. A* 70, 411–415
37 Bartashevich, E.V. *et al.* (2002) Method for multiconformational modeling of the three dimensional shape of a molecule. *J. Struct. Chem.* 43, 1033–1039
38 Harrington, E.C. (1965) The desirability function. *Ind. Qual. Control* 21, 494–498
39 Hsieh, K.-L. *et al.* (2005) Optimization of a multi-response problem in Taguchi's dynamic system. *Comput. Ind. Eng.* 49, 556–571
40 Cosgrove, D.A. *et al.* (2000) A novel method of aligning molecules by local surface shape similarity. *J. Comput. Aid. Mol. Des.* 14, 573–591
41 Geronikaki, A.A. *et al.* (2008) Computer-aided discovery of anti-inflammatory thiazolidinones with dual cyclooxygenase/lipoxygenase inhibition. *J. Med. Chem.* 51, 1601–1609
42 Geronikaki, A.A. *et al.* (2006) Antioxidants and inflammatory disease: synthetic and natural antioxidants with anti-inflammatory activity. *Comb. Chem. High Throughput Screen.* 9, 425–442
43 Herron, D.K. *et al.* (2000) 1,2-Dibenzamidobenzene inhibitors of human factor Xa. *J. Med. Chem.* 43, 859–872
44 Yee, Y.K. *et al.* (2000) N^2-aroylanthranilamide inhibitors of human factor Xa. *J. Med. Chem.* 43, 873–882