Journal of Computational Methods in Molecular Design, 2015, 5 (2):98-108

Scholars Research Library
(http://scholarsresearchlibrary.com/archive.html)

ISSN : 2231- 3176
CODEN (USA): JCMMDA

Determination of the relationship between the docking studies and the biological activity of δ-selective enkephalin analogues

Fatima Sapundzhi[1*], Tatyana Dzimbova[3], Nevena Pencheva[2] and Peter Milanov[1,4]

[1]*Faculty of Mathematics and Natural Sciences, South West University "Neofit Rilski", 2700 Blagoevgrad, Bulgaria*
[2]*Faculty of Public Health and Sports, South West University "Neofit Rilski", 2700 Blagoevgrad, Bulgaria*
[3]*Institute of Molecular Biology, Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria*
[4]*Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria*

ABSTRACT

Delta opioid receptor (DOR) is an attractive object for docking experiments, because is the target of enkephalins, which are endogenous opioid pentapeptides with DOR preferences. The aim of this study is to establish optimal fitting polynomial function for modelling of the structure-activity relationship of a series of δ-opioid selective enkephalin analogues, basing on the quantitative parameters of in vitro bioassay (efficacy, affinity and potency) and the results of the molecular docking with theoretical model of DOR (PDBe:1ozc). The relationship of efficacy with the GoldScore scoring function and with the total energy was modelled with first- to fourth-degree polynomials and surface fitted method. The polynomial surface of the third order has the best fit, assessed by method of least squares. The finding, established in this study, suggests that: (1) the third order polynomial could be successfully used for modeling of the relationship between the efficacy of δ-selective enkephalin analogues and results from docking procedure; and (2) the combination of ligand-based and structure-based approaches of virtual screening is a reliable search of effective δ-selective enkephalin candidates.

Keywords: Surface fitting, Docking, Delta opioid receptor, Affinity, Efficacy

---

INTRODUCTION

The discovery of potent and selective ligands to the delta-opioid receptor (DOR) is related with a large amount of investigations with enkephalin analogues, because enkephalins are endogenous opioid pentapeptides [1, 2, 3] with DOR preferences [4]. Enkephalin analogues with improved δ-properties are an attractive purpose because they: - mediate analgesia [5]; - share morphine's positive therapeutic effects, but reduce negative properties [6, 7]; and – reveal a minimum potential for development of physical dependence [8]. The structure modifications, having a key role in the affinity and selectivity of enkephalin ligands, are the substitution at position 2 and the alterations in the C terminal. It was established in our previous *in vitro* studies [9, 10], that incorporation of hydrophilic Cys(O_2NH_2) at position 2 in the enkephalin molecule greatly increases the potency and selectivity of the respective analogues at DOR. Moreover, basing on the *in vitro* bioassay data and the so-called hyperbolic model of a partial agonism [11], it is possible to calculate with the explicit formulas not only the potency (IC_{50}, or concentration, which produce 50 % of the maximal response of the tissue) and the affinity (dissociation constant) of the respective analogues, but their relative efficacy, as well.

During the last years the computer drug design and virtual screening has become an integral part of the drug discovery process. Although, the various computational techniques are suggested, they would be divided into at least two broad categories of virtual screening [12]: a ligand based, or the so-called pharmacophore model, where the quantitative structure-activity relationship (QSAR) [13] studies are used to screen a respective base of molecules; and a structure-based, which involves docking of candidate ligands into a protein target or receptor, ensuring the

scoring function to estimate the likelihood that the compound will bind to the receptor; the molecular docking as a powerful computational tool have played an increasing role in the functional study of receptors, proteins at all and structure-based drug design [14, 15]. Thus, it could be very useful to find a relationship between the quantitative parameters of the *in vitro* tests (efficacy, affinity, potency) and the results of studies of the molecular docking (the minimum energy of the ligand-receptor complex, the binding affinity values of the scoring functions, etc.) in order to predict biological activity.

However, the data with virtual screening of selective enkephalin analogues and opioid receptors at all are scares in the pertinent literature [12]. On the other hand, in the docking programs an important problem is the development of energy scoring function that can quickly and accurately describe protein-ligand interaction. Binding affinity is calculated by docking, based on different kinds of scoring functions such as PLP, ASP, G-score, F-score, LigScore, GoldScore, ChemScore, ChemPLP etc. In our previous study [16] docking was performed with published theoretical model of DOR (PDBe:1ozc) using Molegro Molecular Docker. Data obtained with this software did not lead to significant correlation with biological activity. This was the reason to use different software with different scoring functions. Then docking experiments were performed with GOLD 5.2 and it was proved that the GoldScore scoring function reveals the best correlation with biological activity [17, 18].

The aim of the current study is to apply the model that describes relationship between the biological activity of δ-selective enkephalin analogues and docking results. In order to achieve this purpose the following three problems should be solved: (1) performance of molecular docking calculations with theoretical model of DOR (PDBe:1ozc) and δ-selective enkephalin analogues, and also calculation of the total energies of ligand-receptor complex after docking experiments; (2) determination of a correlation between the quantitative parameters of the *in vitro* tests (efficacy, affinity, potency) and the results of the molecular docking (scoring function); and (3) finding a function $z = f(x, y)$ from some class polynomials, that can fit given n distinct data points $\{(x_i, y_i, z_i)\}_{i=1}^{n}$ on a surface M in R^3.

MATERIALS AND METHODS

1.1.Objects

- *Receptor - DOR*

The theoretical model of the δ-opioid receptor (DOR) published in RCSB Protein Data Base (PDBe accession code 1ozc) was used. The protein is composed of 268 amino acids in a chain.

- *Ligands*

Eleven ligands, investigated for their potency, selectivity and efficacy to DOR with *in vitro* bioassay in a previous study [9, 10] were selected for docking studies. The following generally accepted terms for *in vitro* assay are used: A-opioid agonist; IC_{50} (potency) – concentration of an agonist A (ligand), which produce 50% of the maximal response of the tissue; $1/K_A$ (affinity) – K_A is dissociation constant with units A of the ligand; e_{rel} (efficacy) – relative efficacy of the agonist A, which is unitless [10, 11]. Their primary structure, including that of selective ligand DPDPE ([D-Pen^{2,5}]-enkephalin, selective δ-opioid receptor agonist) [19] and endogenous enkephalins ([Leu^5]- and [Met$^{5]}$-enkephalin) and their analogues are presented in Table 1.

Table 1: Ligands used in this study

Primary structure	Mouse vas deferens Ligand	IC_{50} (nM)	K_A (nM)	e_{rel}
Tyr-D-Pen-Gly-Phe-D-Pen	DPDPE	6.18±1.17	180±35	30.2±10.0
Tyr-Gly-Gly-Phe-Leu	[Leu^5]-enk	11.45±2.06	54.9±13.1	5.8±1.0
Tyr-Gly-Gly-Phe-Met	[Met^5]-enk	18.91±2.15	48.4±7.5	3.6±0.3
Tyr-Cys(Bzl)-Gly-Phe-Leu	[Cys(Bzl)^2, Leu^5]-enk	8.30±1.40	68.5±29.7	9.3±3.2
Tyr-Cys(Bzl)-Gly-Phe-Met	[Cys(Bzl)^2, Met^5]-enk	9.53±1.20	23.8±3.0	3.5±0.3
Tyr-Cys(O2NH2)-Gly-Phe-Leu	[Cys(O_2NH_2)^2, Leu^5]-enk	1.29±0.31	36.4±16.4	29.2±9.5
Tyr-Cys(O2NH2)-Gly-Phe-Met	[Cys(O_2NH_2)^2, Met^5]-enk	2.22±0.45	14.1±5.4	7.3±2.0
Tyr-D-Cys(O2NH2)-Gly-Phe-Leu	[DCys(O_2NH_2)^2, Leu^5]-enk	11.40±2.01	73.4±12.7	7.4±1.9
Tyr-D-Cys(O2NH2)-Gly-Phe-Met	[DCys(O_2NH_2)^2, Met^5]-enk	75.96±11.67	463±161	7.1±1.8
Tyr-HCys(O2NH2)-Gly-Phe-Leu	[HCys(O_2NH_2)^2, Leu^5]-enk	31.92±5.10	76.4±7.1	3.4±0.2
Tyr-HCys(O2NH2)-Gly-Phe-Met	[HCys(O_2NH_2)^2, Met^5]-enk	16.09±1.90	55.7±6.1	4.5±0.3

***Target**: Human ɗ-opioid receptor (DOR), published in PDBe (id: 1OZC), [20].*

1.2.Docking procedure

The docking experiments were carried out with software GOLD 5.2 (Genetic Optimisation for Ligand Docking), which uses a genetic algorithm and considers full ligand conformational flexibility and partial protein flexibility [15, 21, 22]. Docking experiments were carried out using all four GOLD scoring functions, GoldScore, ChemScore,

Astex Statistical Potential (ASP) and ChemPLP, which makes it possible to verify the binding ability of the appropriate ligand with the receptor.

In this article it was described the implementation of the ColdScore function as a scoring function for GOLD 5.2 and its usefulness to perform docking precisely, to predict the binding energies, and to realise the biological effects of investigated compounds.

GoldScore scoring function (Eq. 1) is a molecular mechanics–like function with four terms. The fitness score is taken as the negative of the sum of the component energy terms, so that larger fitness scores are better.

(1) $$GoldScore = S_{hb_{ext}} + S_{vdw_{ext}} + S_{hb_{int}} + S_{vdw_{int}},$$

where
S_{hb_ext} - protein–ligand hydrogen-bond energy (external H-bond);
S_{vdw_ext} - protein-ligand van der Waals (vdw) energy (external vdw);
S_{hb_int} - intramolecular hydrogen bonds in the ligand;
S_{vdw_int} - intramolecular strain in the ligand.

The total energies of ligand-receptor complex were calculated by Molegro Molecular Viewer (http://molegro-molecular-viewer.software.informer.com), (MMV Version 2.5) [23], using MolDock algorithm. For analyzing the docking results it was used Ligand Energy Inspector tool of MMV. It allows getting detailed information about the energy interactions for the protein-ligand complex. In this study, in order to calculate the total energy of the ligand-receptor complexes it was used MolDock scoring function (Eq. 2).

(2) $$E_{score} = E_{inter} + E_{intra}$$

where E_{score} is a docking scoring function, E_{inter}- ligand-protein interaction energy, and E_{intra} - internal energy of the ligand [3].

1.3. Correlation and fitting methods

- *Correlations*

Determination of a correlation between the quantitative parameters of the *in vitro* tests (efficacy, affinity, potency) and the results of the molecular docking (scoring function) was carried out by software GraphPad Prism 3.0 (http://www.graphpad.com/scientific-software/prism). In this study was used this software for calculating the Pearson correlation coefficient (Eq.3). It is a measure of the correlation (linear dependence) between both variables. Concerning the choice of the criterion it has to be kept in mind: that the Spearman correlations are based on ranks, not the actual values, and so it could be assumed that in this investigation, the proper criterion would be that of Pearson.

(3) $$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Where $(x_i - \bar{x})$ is each *x*-value (quantitative parameters of the *in vitro* tests) minus the mean of *x;*$(y_i - \bar{y})$ is each *y*-value (results of the molecular docking) minus the mean y.

- *MATLAB®*

The fitting of experimental data can be presented as follows (Eq:4):

(4) $$\underset{(a_{00},\dots,a_{0n})}{\text{minimize}}\, F(a_{00},\dots,a_{0n}) = \sum_{s=1}^{m}\left(z_s - \sum_{0\le i+j\le n} a_{ij} x_s^i\, y_s^j\right)^2$$

where

(5) $$z = \sum_{0\le i+j\le n} a_{ij} x^i\, y^j$$

In (Eq.5) *s* is the number of points; *m* is the number of ligand-receptor complexes; *z* is a dependent variable, x and y are independent variables. The values of $z_1, z_2 \dots, z_n$ represent the efficacy as *in vitro* parameter; the values of $x_1, x_2 \dots, x_n$ represent the result from the docking procedure (scoring function); the values of $y_1, y_2 \dots, y_n$ represent the total energy for ligand-receptor complex; a_{ij} are the parameters of the model. Here *n* is the degree of the

polynomial, $0 \leq i + j \leq n$. The degree gives the number of coefficients to be fit and the highest power of the predictor variable.

To study the fitting behavior of several polynomial degree functions, it has been performed series of fittings, starting from the first degree until the fourth one. The Surface Fitting Tool of MATLAB® (http://www.mathworks.com/products/matlab) [24] was applied when we analyzed the behavior of one variable which depended on more independent variables and the individual model could be interpreted as a surface fitting function of the experimental data by method of least squares (Eq.4). All analyses were done using MATLAB and by Surface Curve Fitting Toolbox. All models for data fitting were tested (fit a linear, quadratic, cubic) and it was evaluated how well those models fitted the data and how precisely they could predict. Surface Fitting Tool allows us to visually explore one or more data sets and fits as scatter plots, graphically evaluate the goodness of fit using residuals and prediction bounds, access additional interfaces for fitting data, and compare fits and data set.
Mathematically, the residual for a specific predictor value is the difference between the response value y and the predicted response value $\hat{y}$ (residual $=$ data $-$ fit).

- *Parameters to evaluate the goodness of fit*

After using graphical methods to evaluate the goodness of fit it should be examined the goodness of fit statistics. The *Surface Fitting Toolbox* software supports these goodness-of-fit statistics for parametric models.
-*SSE* (Sum of squares due to error) - the total deviation of the response values from the fit to the response values or the summed square of residuals (Eq.5). A SSE value near to 0 shows that the model has a smaller random error component and then the fit will be more useful for prediction.

(6) $$SSE = \sum_{i=1}^{n} \omega_i (y_i - \hat{y}_i)^2$$

Here y_i is the observed data value, $\hat{y}_i$ is the predicted value from the fitting curve and ω_i is the weighting applied to each data point, usually $\omega_i = 1$.
-*R-Square* (r^2) - measures how successful the fit is in explaining the variation of the data and it is the square of the correlation between the response values and the predicted response values (Eq.7). It is called the square of the multiple correlation coefficients. r^2 is defined as the ratio of the sum of squares of the regression (SSR) and the total sum of squares (SST) about the mean (Eq.8). R-square can take on any value between 0 and 1, with a value closer to 1 indicating that a greater proportion of variance is accounted for by the model. R-square value of 1.0 means that the fit explains 100% of the total variation in the data about the average.
-

(7) $$r^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

(8) $SSR = \sum_{i=1}^{n} \omega_i (\hat{y}_i - \bar{y}_i)^2$, $SST = \sum_{i=1}^{n} \omega_i (y_i - \bar{y}_i)^2$, $SST = SSR + SSE$

-*Degrees of Freedom* - the number of response values n minus the number of fitted coefficients m estimated from the response values ($v = n - m$). Here v is the number of independent pieces of information involving the n data points that are required to calculate the sum of squares.
-*Adjusted R-square* (Eq. 9) - the best indicator of the fit quality when compare two models that are nested - that is, a series of models each of which adds additional coefficients to the previous model. The $adjusted\ r^2$ statistic can take on any value less than or equal to 1, with a value closer to 1 indicating a better fit. Negative values can occur when the model contains terms that do not help to predict the response

(9) $$adjusted\ r^2 = 1 - \frac{SSE(n-1)}{SST(v)}$$

-$RMSE$ (Root Mean Squared Error) - the standard error of the regression and an estimate of the standard deviation of the random component in the data (Eq.10). A RMSE value closer to 0 indicates a fit that is more useful for prediction.

(10) $$RMSE = s = \sqrt{MSE}$$

MSE - the mean square error or the residual mean square ($MSE = SSE/v$).
-*# Coeff* - the number of coefficients in the model.

- *Optimal solution and guarantee for global extreme*

By using the method of least squares we found the parameter valuesa_i, i = 1, . . . , k. This method is a mathematical procedure for finding the best-fitting curve to a given set of points by minimizing the sum of the squares of the offsets (residuals) of the points from the curve. In order to have high probability for the solving to be the global minimum it was applied the following procedure recommended in MATLAB: (1) the procedure was started with a large number of different initial parameter values $a_i, i = 1, \dots, k$, If the found solution is: $(a_{00}^{opt}, \dots, a_{0k}^{opt})$, then again it was started the procedure from points $(a_{00}^{opt} \pm \varepsilon_1, \dots, a_{0k}^{opt} \pm \varepsilon_k)$, where $\varepsilon_t \in [0, 200], t = 1, \dots, k$, where ε_t are high enough and again was reached $(a_{00}^{opt}, \dots, a_{0k}^{opt})$ as a problem solution. (2) In order to avoid an eventual local extreme it was applied a procedure resembling the "simulated annealing" method (a global optimization technique that is used to avoid falling into a local extreme), the optimization procedures was repeated many times, by starting from point $(a_{00}^{opt} \pm \varepsilon_1, \dots, a_{0k}^{opt} \pm \varepsilon_k)$, where the initial $\varepsilon_t, t = 1, \dots, k$ are not high and we reached again $(a_{00}^{opt}, \dots, a_{0k}^{opt})$ as a problem solution.

- *Criteria for comparing the classes of models*

After finding the best model which gives an optimal fitting of data we apply the Akaike's information criteria (AIC) (Eq.11) [25, 26, 27] and Bayesian information criteria (BIC) (Eq.12) [1,14,30] to select one of the polynomial models, according to the criteria of optimal selection. For calculation and comparison of the criteria values of AIC and BIC it was used the program "Comparing Models" [28]. Given a set of candidate models for the data, the preferred model is the one with the minimum AIC value.

$$(11)\quad AIC = \begin{cases} n * ln\left(\frac{RSS}{n}\right) + 2 * k, & \frac{n}{k} \geq 40 \\ n * ln\left(\frac{RSS}{n}\right) + 2 * k + \frac{2 * k * (k+1)}{n-k-1}, & \frac{n}{k} < 40 \end{cases}$$

$$(12)\quad BIC = n * ln\left(\frac{RSS}{n}\right) + k * ln(n),$$

Where n is the number of data points; k is the number of parameters fit by the regression plus one (since regression is "an estimating" of the sum-of-squares as well as the values of the parameters). Here RSS (residual sum of squares), $RSS = \sum_{i=1}^{n}(y_i - f(x_i)^2$ is the sum of the squares of the vertical deviations from each data point to the fitted line.

RESULTS AND DISCUSSION

Docking results and calculation of total energy

For surface fitting of the relationship between efficacies of enkephalin analogues, total energy calculates by MMV and GoldScore scoring function were applied with methods described in Section 2. The results are presented in Table 2.

Table 2: The values of the main parameters used for surface fitting: GoldScore scoring function calculated by GOLD 5.2, total energy calculated by MMV and e_{rel} obtained by *in vitro* bioassay

Ligands	GoldScore function	total energy	e_{rel}
[Cys(Bzl)^2-Leu^5]-enk	64,68	-107.022	9.3
[Cys(Bzl)^2-Met^5]-enk	81,49	-89.091	3.5
[Cys(O_2NH_2)^2-Leu^5]-enk	67,72	-97.619	29.2
[Cys(O_2NH_2)^2-Met^5]-enk	73,91	-91.246	7.3
[DCys(O_2NH_2)^2-Leu^5]-enk	74,73	-84.852	7.4
[DCys(O_2NH_2)^2-Met^5]-enk	75,13	-86.221	7.1
DPDPE	57,67	-109.709	30.2
[HCys(O_2NH_2)^2-Leu^5]-enk	68,43	-62.774	3.4
[HCys(O_2NH_2)^2-Met^5]-enk	78,65	-93.301	4.5
[Leu^5]-enkephalin	73,42	-81.869	5.8
[Met^5]-enkephalin	73,26	-118.971	3.6

Docking was performed with the theoretical model of DOR (PDBe:1ozc) and all 11 ligands in GOLD 5.2 [15, 21]. The binding site of the receptor is known from the literature. It is the residues within 10 Å around an aspartic acid residue, Asp128. (For example in Figure 1 it is shown the ligand-receptor complex between theoretical model of DOR and ligand [Leu^5]-enkephalin, obtained in MMV). Given a protein target in our case DOR (PDBe:1ozc), molecular docking with GOLD 5.2 generates several probable ligand binding orientations/conformations at the active site around the receptor. A GoldScore fitness function is used to rank these ligand orientations/conformations by evaluating the binding density of each of the probable complexes. The results of the docking (Tab. 2) show the

relative pose prediction performance of GOLD 5.2 by the GoldScore fitness function [18]. These values are calculated using only the best scored pose out for each binding site (solution with the highest score).

The total energy of the ligand-receptor complex of theoretical model of DOR (PDBe:1ozc) and the best pose of the corresponding ligand was calculated by MolDock algorithm [23] in MMV (Tab.2).

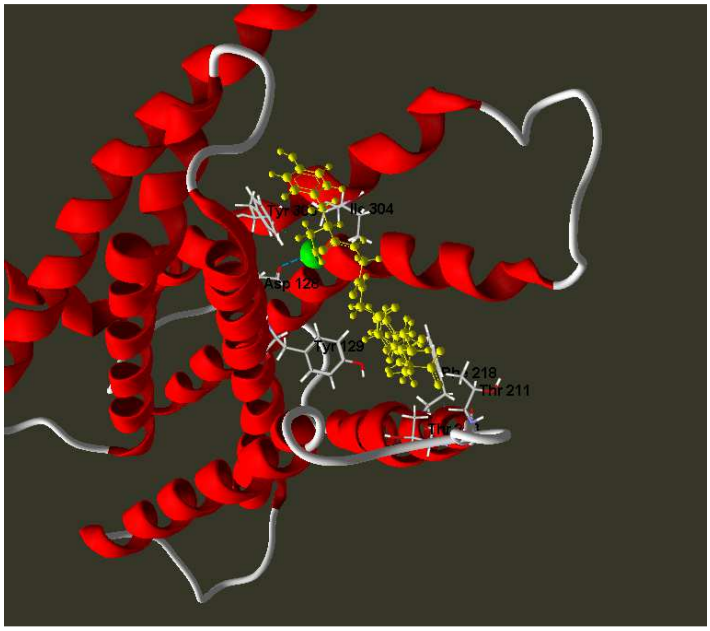


Figure 1: Schematic diagram of ligand-receptor complex between DOR (PDBe:1ozc) and endogenous ligand [Leu^5]-enkephalin. The receptor is presented in ribbons and helixes. The ligand is presented in yellow circles. This diagram was generated with the MMV

Correlations and fitting

The results of correlations between the data from the docking with all four scoring functions available in GOLD 5.2 and the values of *in vitro* bioassay (IC_{50} , K_A, e_{rel}) [9, 10, 11] were obtained with GraphPad Prism 3.0. The commonly-used criterion for affinity prediction is the Person's correlation between the calculated scores procedure and the experimental data. The results are shown in Table 3.

Table 3: The values of Pearson's correlation between *in vitro* parameters (IC_{50}, K_A, e_{rel}) and all four scoring functions of GOLD 5.2 for the theoretical model of DOR (PDBe:1ozc)

Functions	IC_{50}	K_A	e_{rel}
ASP	-0.09349	-0.2418	-0.2518
ChemPLP	-0.05332	-0.3422	-0.4721
ChemScore	-0.2801	-0.2678	-0.08067
GoldScore	0.2069	-0.09418	**-0.7209**

As it can be seen from Tab. 3 the highest values of the Pearson's correlation coefficient were obtained between GoldScore fitness function and e_{rel}, ($r = -0.7273$, Fig. 2A) [18].The correlations between the other *in vitro* parameters and GoldScore function are low: between GoldScore and IC_{50}, $r = -0.2069$, between GoldScore and K_A, $r = -0.09418$. There is a low correlation between the values of the total energy of ligand-receptor complex and e_{rel} ($r = -0.3748$, Fig. 2B). This data indicate that GOLD 5.2 software gives reliable results in the docking of delta selective enkephalin analogues with theoretical model of DOR (PDBe:1ozc).

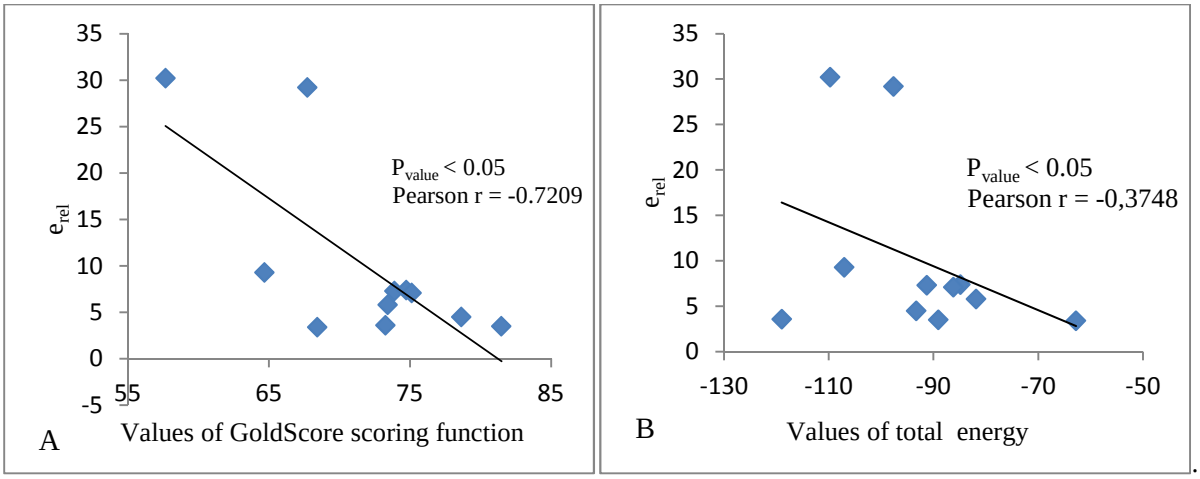


Figure 2: Correlations between the values of the GoldScore scoring function of enkephalin analogues and e_{rel} (A) and the values of the total energy and e_{rel} (B) .

In order to determine of the relationship between biological activity and docking results of the studied objects it was applied the software MATLAB software with its Surface Curve Fitting Toolbox, that provides applications and functions for fitting curves and surfaces to data. To reach this objective it was tested the polynomial models of the first to fourth degree for fitting the experimental data. These models were obtained by the method of least squares [24]. Then they obtained models were evaluated on how well they fitted the data and how precisely they could predict. All of these models were estimated with the statistical criteria of goodness of fit, which were described in Section 2. The results are presented in Table 4.

Table 4: Assessing the goodness of fit for the polynomial models obtained by least squares method

Degree	SSE	R-square	Adj r-square	RMSE	Coefficients
First	443.5817	0.5446	0.4308	7.4463	3
Second	167.1000	0.8285	0.6569	5.7810	6
Third	**0.0092**	**1.0000**	**0.9999**	**0.0960**	**10**
Fourth	Needed at least 15 data points to determine 15 coefficients				

As it can be seen from the results in Tab. 4 the goodness of fit statistics shows that the obtained model for fitting of the data with the third order for x and the third order for y is a good one. The polynomial model of third degree is with the highest value of $r^2 = 1.0$, and the value closer to 1 indicating that a greater proportion of variance is accounted for by the model. The value of $SSE = 0.009207$, which is near to 0. Therefore this value of SSE shows that the model has a smaller random error component and then the fit will be more useful for prediction. The degrees of freedom for the obtained model are 1. The value of $adjusted\ r^2 = 0.9999$ and it is closer to 1. This indicates that the obtained polynomial model for the surface fitting data is a good model. Therefore this model explains a high proportion of the variability in experimental data, and is able to predict new observations with high certainty [30].
For the polynomial model of the third order it was applied techniques for avoiding falling into local extreme that guarantees with a very high probability that the optimal solutions found in the article are global, not local. These techniques were described in Section 2.

After finding the polynomial models of first, second and third degree we apply AIC [25, 26, 27] and BIC [25, 26, 31] criteria to select one of these model, called an “optimal” model, according to the criteria of optimal selection. One of the most commonly used criteria for model selection is AIC. Its idea is to select the model that minimizes the negative likelihood penalized by the number of parameters [25, 26, 27]. The model that minimizes BIC has the highest posterior probability. AIC and BIC criteria differ only in that the coefficient multiplies the number of parameters, i.e. the criteria differ by how strongly they penalize large models. In general, models chosen by BIC will be more parsimonious than those chosen by AIC. The criteria AIC and BIC are calculated for any of the found polynomial models by program “Comparing Models” [28] and the values are presented in Table 5:

Table 5: Assessment of models with AIC and BIC information criterion.

Polynomial models	AIC	BIC
First degree	55.33	50.26
Second degree	81.26	46.71
Third degree	-319.95	-51.57

As it can be seen in Tab. 5 the "optimal" model is the polynomial model which is a model with the lowest value for AIC = -319.95 and BIC = -51.57. Therefore, the optimal model is the model of third degree.

The best results for fitting data according to the results in Table 4 and Table 5 were obtained for surface fitting by a cubic polynomial in three-dimensional with *x* degree of 3 and a *y* degree of 3 for determining the relationship between biological activities and docking results presented in Table 3. In this case, the polynomial model may provide a good approximation of the relationship. By using a polynomial least squares surface fitting technique, a third order cubic polynomial was fitted to the data and is represented as the following Eq.(13):

$$(13)\ f(x,y) = a_{00} + a_{10} * x + a_{01} * y + a_{20} * x^2 + a_{11} * x * y + a_{02} * y^2 + \\ + a_{30} * x^3 + a_{21} * x^2 * y + a_{12} * x * y^2 + a_{03} * y^3$$

Here x represents the values of GoldScore scoring function and it is normalized by mean 71.74, with the value of the standard distribution of the data 6.693; y represents the values of total energy and it is normalized by mean -92.97, with the value of the standard distribution of the data 15.3.

Empirical testing found that this third order polynomial surface fit was the best compromise in terms of goodness of fit and the overall representation of the experimental data. Additionally, the mean surface was calculated for all ligands and the polynomial coefficients of this fit are presented in Table 6. The confidence bounds on the coefficients determine their accuracy. In this study they are with 95% confidence bounds and are relevant to evaluate and compare fits.

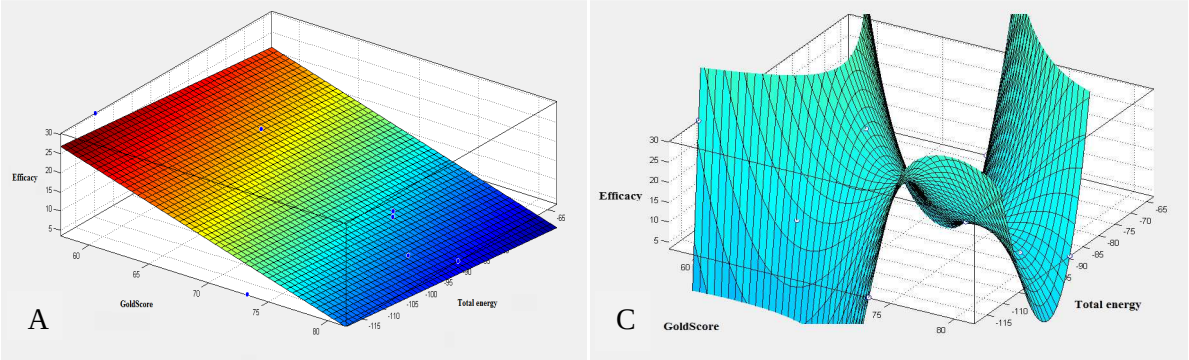
The coefficients of the surface fitting by a cubic polynomial in three dimensions are presented in Table 6.

Table 6: The mean values (confidence bounds) of the coefficients of the third order polynomial model chosen as optimal model

Coefficients	Mean (with 95% confidence bounds)	
a_{00}	11.51	(9.823, 13.19)
a_{10}	-11.07	(-16.13, -6.008)
a_{01}	-22.37	(-33.1, -11.64)
a_{20}	16.71	(14.65, 18.78)
a_{11}	3.451	(-6.742, 13.64)
a_{02}	-0.6185	(-3.866, 2.629)
a_{30}	-12.15	(-14.89, -9.411)
a_{21}	19.03	(11.96, 26.11)
a_{12}	44.7	(29.97, 59.43)
a_{03}	14	(7.377, 20.62)

The efficacy as a function of the values of GoldScore fitness function of docking experiments and the values of the total energy was presented in Figure 3 (A,B,C and D) with a polynomial surface fitting of first to third order in MATLAB.

As it can be seen from the graphical representation of the experimental data (Table 3) the best polynomial surface fitting is obtained for third order of polynomial model (Figure 3 C and D). In Figure 3 (D) it is presented the polynomial surface fitting with the third degree of polynomial after applying the validation with the same experimental data and it was obtained very good results for them again (SEE=0.0092 and RMSE=0.0289).



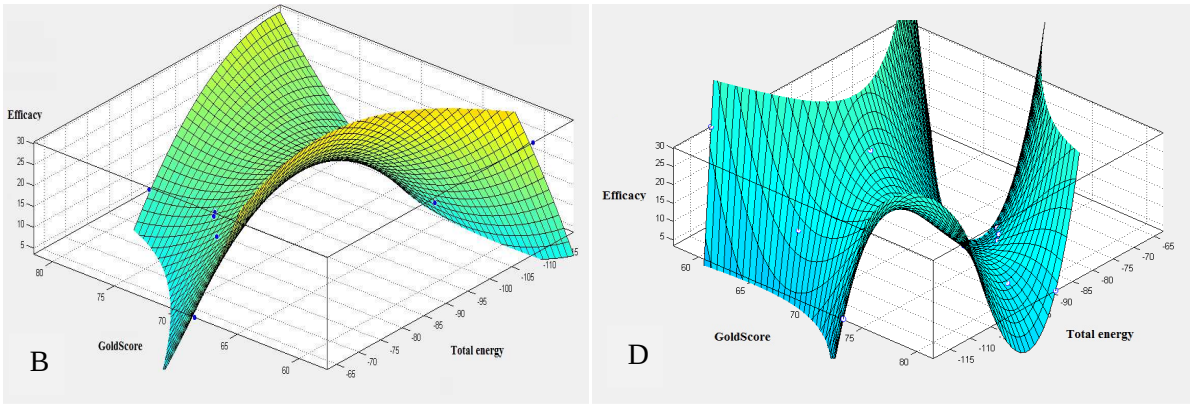


Figure 3: A three-dimensional surface fitting of experimental data with polynomial of third degree

which representing the efficacy as a function of the values of GoldScore scoring function from docking procedure and the values of the total energy for ligand-receptor complex. The surface fitting with the first degree of polynomial model is presented in (A); with the second degree in (B); with the third degree in (C) and in (D) after validation the same experimental data. These diagrams were generated with MATLAB.

A graphic representation of the relationship among the three numeric variables in two dimensions is presented in Figure 4. The values of the GoldScore function and the values of total energy are for X and Y axes, and the values of the efficacy are for contour levels. For the fitting by a cubic polynomial in three-dimensional surface the contour plot (Figure 4) makes it easier to see points that have the same height. As it can be seen in Figure 4 the best results are obtained for the polynomial model of third degree. When compared to the surface plots, they may be less effective to quickly visualize the overall shape of 3D data; however, their main advantage is that they allow for precise examination and analysis of the shape of the surface. Contour plots display a series of undistorted horizontal "cross sections" of the surface.

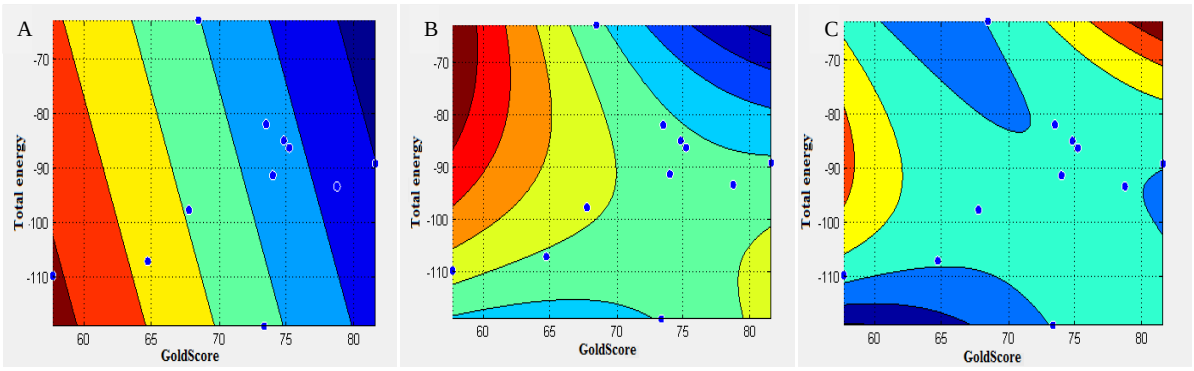


Figure 4: A two-dimensional contour plot of the three-dimensional surface in the Figure 3.

X represents the values GoldScore scoring function and Y represents the values of total energy. The first degree of polynomial fitting is presented in (A). The second degree of polynomial fitting is presented in (B). The third degree of the polynomial fitting is presented in (C).These diagrams were generated with the Matlab.

Figure 5 represents the residual plot for obtained polynomial model. It provides visual displays for assessing how well the model fits the data, for evaluating the distribution of the residuals, and for identifying influential observations. The top plot of residual plot shows that the residuals are calculated as the vertical distance from the data point to the fitted curve. The bottom plot displays the residuals relative to the fit, which is the zero line. As it can be seen in Figure 5 the obtained model is randomly scattered around zero. This indicates that the polynomial model of third degree describes the data in a good way.

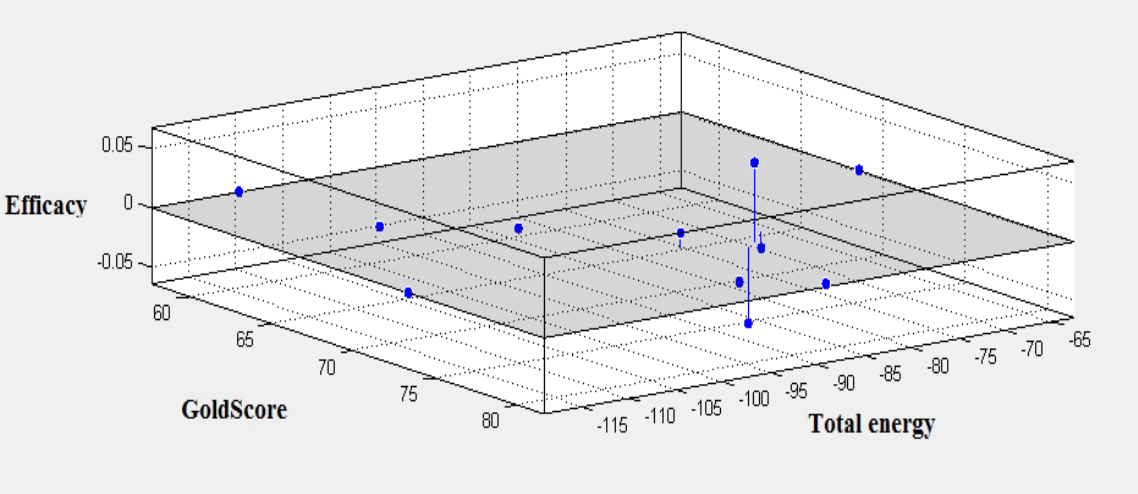


Figure 5: Residuals Plot for the obtained polynomial model.

Polynomial models are among the most frequently used empirical models for curve fitting and they are popular for the following reasons: they have a simple form, well known and understood properties, moderate flexibility of shapes and are computationally easy to use. Polynomials models have two important characteristics: a quantitative – the degrees of the polynomials, respectively the number of parameters of model and a qualitative - the regression function is linear in terms of the unknown parameters a_{ij} (Eq.4). This allows easy to find the optimal regression coefficients using method of least squares.

The obtained model for the experimental data showed good fitting properties and significant predictive ability ($SSE = 0.0092,\ R^2 = 1.0, RMSE = 0.0960$). Therefore this model of third degree is suitable for determination the relationship structure-biological activity. The GoldScore scoring function and total energy obtained from docking could be used for predicting the efficacy of newly designed compounds. This would be helpful in shortening the drug design process.

CONCLUSION

The docking procedure was carried out with theoretical models of DOR and series of of δ-opioid selective enkephalin analogues. The total energies for the resulting ligand-receptor complexes were calculated. All parameters of *in vitro* studies are compared with the results from docking (scoring function) in order to find correlation. GoldScore scoring function correlates (Pearson's correlation coefficient $r = -0.7273$) with the efficacy (e_{rel}) of δ-opioid selective enkephalin analogues, calculated from *in vitro* experiments.

The polynomial surface of the third order has the best fit, assessed by method of least squares. The finding, established in this study, suggests that: (1) the third order polynomial could be successfully used for modeling of the relationship between the efficacy of δ-selective enkephalin analogues and results from docking procedure; and (2) the combination of ligand-based and structure-based approaches of virtual screening is a reliable search of effective δ-selective enkephalin candidates.The polynomial surface fitting model was obtained by Surface Fitting Toolbox in MATLAB.

Analysis and comparison of the data from *in vitro* tests and docking studies could help to understand better the relationship between *in vitro* biological effects and docking studies and to answer whether the models of the biological macromolecules (in our case DOR) correspond to the real 3D structure. On the other hand, obtained model is applicable for predicting the efficacy of compounds with known score and total energy. This opens the scope for further research and the results will be published soon.

Acknowledgments
This work is partially supported by the project of the Bulgarian National Science Fund, entitled: "Bioinformatics research: protein folding, docking and prediction of biological activity", code NSF I02/16, 12.12.14.

REFERENCES

[1] P. Law; Y. Wong; H. Loh. *Ann. Rev. Pharmacol. Toxicol*, **2000,** 40: 389-430.

[2] J. Lord; A. Waterfield; J. Hughes; H. Kosterlitz. *Nature*, **1977**, 267: 495-499.
[3] R. Maldonado; Severini R.; W. Hans; B. Kiefer; P. Melchiorri; L. Negri. *British Journal of Pharmacology*, **2001**, 132: 1485–1492.
[4] R. Bodnar J. *Peptides*, **2013**, 60: 65-95.
[5] R. Quock; T. Burkey; E. Varga; Y. Hosohata; K. Hosohata; S. Cowell; C. State; F. Ehlert; W. Roeske; H. Yamamura. *Phannacol. Rev.* **1999**, 51: 503-532.
[6] A. Dickenson. *Behav. Brain Sci* .**1997**, 20: 392-403.
[7] P. Cheng; D. Wu; J. Decena; Y. Soong; S. Mccabe; H. Szeto. *Eur. 1. Pharmacol.* **1993,** 230: 85-88.
[8] A. Cowan; X. Zhu; H. Mosberg; R. Omnaas; E. Porreca. *J. PhannacoI. Exp. Ther.* **1988,** 246: 950-955.
[9] N. Pencheva; A. Bocheva; E. Dimitrov; C. Ivancheva. *Eur. J. Pharmacol.* **1996**, 304: 99-108.
[10] N. Pencheva; P. Milanov; L. Vezenkov; T. Pajpanova; E. Naydenova. *Eur. J. Pharmacol.* **2004**, 498: 249-256.
[11] P. Milanov; N. Pencheva. *Serdica J. Computing*, **2011**, 5: 333–358.
[12] U. Rester. *Curr Opin Drug Discov Devel*, **2008**, 11:559–68.
[13] H. Kubinyi. *Springer Science & Business Media*, **1994** Netherlands ().
[14] M. Eldridge; C. Murray; T. Auton; G. Paolini; R. Mee. *J. Comput. Aided Mol. Des.*, **1997**, 11: 425-445.
[15] GOLD, version 5.2 UserGuide; CCDC Software Ltd.: Cambridge, UK, **2010**.
[16] T. Dzimbova; R. Mavrevski; N. Pencheva; T. Pajpanova; P. Milanov. *Bulgarian Chemical Communications*, **2012**, 44: 242-245.
[17] T. Dzimbova; F. Sapundzhi; N. Pencheva; P. Milanov. *Int. J. Bioautomation*, **2013**, 17: 5-16.
[18] F. Sapundzhi; T. Dzimbova; N. Pencheva; P. Milanov. *Biomath Communications*, 2014, 1: 84.
[19] J. Clark; Y. Itzhak; V. Hruby; H. Yamamura; G. Pasternak. *Eur. J. Pharmacol.*, **1986**, 128: 303-304.
[20] F. Décaillot; K. Befort; D. Filliol; S. Yue; P. Walker; B. Kieffer. *Nat Struct Biol.*, **2003**, 10: 629-936.
[21] M. Verdonk, J. Cole; M. Hartshorn; C. Murray; R. Taylor. *Proteins*, **2003**, 52: 609-623.
[22] R. Reinscheid; H. Nothacker; A. Bourson; A. Ardati; R. Henningsen; J. Bunzow; D. Grandy; H. Langen; F. Monsma; O. Civelli. *Science*, **1995**, 270: 792-794.
[23] R. Thomsen; M. Christensen.*J. Med. Chem.*, **2006**, 49: 3315-3321.
[24] P. Lancaster. *Proc. NATO Adv. Study Inst. Ser.*, Univ. Calgary, Calgary, Alta., **1979,** 49: 91–102.
[25] H. Acquah de-graft. *Journal of Development and Agricultural Economics*, **2010**, 2: 001-006.
[26] H. Akaike. *IEEE Transactions on Automatic Control*, **1974**, 19: 716-723.
[27] J. Kadane; N. Lazar. *Am. Statist. Assoc*. **2004**, 99: 279–290.
[28] R. Mavrevski. *C. R. Acad. Bulg. Sci*., **2014**, 67: 1345-1354.
[29] K. Befort; L. Tabbara; S. Bausch; C. Chavkin; C. Evans; B. Kieffer. *Mol Pharmacol*, **1996**, 49: 216-223.
[30] F. Sapundzhi; T. Dzimbova; N. Pencheva; P. Milanov. *Proceedings of the Sixth International Scientific Conference–FMNS 2015, 10 – 14 June*, **2015**, 1: 104-112.
[31] W. Zucchini. *Journal of Mathematical Psychology*, **2000**, 44: 41-61.