A predicted chemo-polypharmacophoric agent comprising (Propeptide-Fc)/MGF peptide mimicking interactive of high free binding energy properties towards Wnt7a/Fzd7 signalling Akt/mTOR anabolic growth IGF-I/PI3K/Akt -I/MAPK/ERK pathways.

Grigoriadis Ioannis1, Grigoriadis George2 and Grigoriadis Nikolaos3*

1.Department of Computer Drug Discovery Science,

BiogenetoligandorolTM, Thessaloniki, Greece,

2.Department of Stem Cell Bank and ViroGeneaTM,

Biogenea Pharmaceuticals Ltd, Thessaloniki, Greece,

3.Department of IT Computer Aided Personalized Myoncotherapy,

Cartigenea-Cardiogenea, Neurogenea-Cellgenea, Cordigenea-HyperoligandorolTM,

ABSTRACT: The insulin-like growth factor-I (IGF-I) is a key regulator of skeletal muscle growth in vertebrates, promoting mitogenic and anabolic effects through the activation of the MAPK/ERK and the PI3K/Akt signaling pathways. Also, these results show that there is a time-dependent regulation of IGF-I plasma levels and its signaling pathways in muscle. The insulin-like growth factor-I (IGF-I) is a key regulatory hormone that controls growth in vertebrates. Particularly, skeletal muscle growth is strongly stimulated by this hormone. IGFI stimulates both proliferation and differentiation of myoblasts, as well as promoting myotube hypertrophy in vitro and in vivo. The mitogenic and anabolic effects of IGF-I on muscle cells are mediated through specific binding with the IGF-I receptor (IGF-IR). This ligand-receptor interaction promotes the activation of two major intracellular signaling pathways, the mitogen-activated protein kinases (MAPKs), specifically the extracellular signal-regulated kinase (ERK), and the phosphatidylinositol 3 kinase (PI3K)/Akt. The MAPK (RAF/MEK/ERK) is a key signaling pathway in skeletal muscle, where its activation is absolutely indispensable for muscle cell proliferation. Biologically active polypeptides derived from the E domain that forms the C-terminus of the insulin-like growth factor I (IGF-I) splice variant known as mechano growth factor which have been demonstrated neuroprotective and cardioprotective properties, as well as the ability to increase the strength of normal and dystrophic skeletal muscle. Ligands selected from phage-displayed random peptide libraries tend to be directed to biologically relevant sites on the surface of the target protein. Protein-peptide interactions form the basis of many cellular processes. Consequently, peptides derived from library screenings often modulate the target protein's activity in vitro and in vivo and can be used as lead compounds in drug design and as alternatives to antibodies for target validation in both genomics and drug discovery. In this research and science project we for the first time a predicted chemo-polypharmacophoric agent comprising (Propeptide-Fc)/MGF peptide mimicking properties for the possible increasement of the Muscle Mass Fiber Size towards Wnt7a/Fzd7 Signalling to the Akt/mTOR Anabolic Growth IGF-I/PI3K/Akt -I/MAPK/ERK pathways utilising (Propeptide-Fc)/MGF phage-displayed random peptide libraries through a KNIME-RDkit-CDK clustering pipeline.

II METHODS

A Sequential Solution of the Poisson-Boltzmann Equation through a Combination Index Dynamic Unified Theorem for Multiple Entities:

$$s_i(x)=\sum_{j=1}^{n_i}\frac{(x_{i,j}-\bar{X})}{n_i}\bar{X}=\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{n_i}x_{i,j}\qquad \frac{(f_a)_{1,2}}{(f_u)_{1,2}}=\frac{(f_a)_1}{(f_u)_1}+\frac{(f_a)_2}{(f_u)_2}=\frac{(D)_1}{(D_m)_1}+\frac{(D)_2}{(D_m)_2}$$

and when m ≠ 1, then:

$$\left[\frac{(f_a)_{1,2}}{(f_u)_{1,2}}\right]^{1/m}=\left[\frac{(f_a)_1}{(f_u)_1}\right]^{1/m}+\left[\frac{(f_a)_2}{(f_u)_2}\right]^{1/m}=\frac{(D)_1}{(D_m)_1}+\frac{(D)_2}{(D_m)_2}$$

Based on Eqs. 1 and 2, in conjunction with Eq. 4, Chou and Talalay in 1983 introduced the term *combination index* (CI) for quantification of synergism (CI<1), additive effect (CI=1), and antagonism (CI>1) [6,13,14], where at x% inhibition, the general equation for two drugs is given below:

$$CI=\frac{(D)_1}{(D_x)_1}+\frac{(D)_2}{(D_x)_2}=\frac{(D)_1}{(D_m)_1\{(f_a)_1/[1-(f_a)_1]\}^{1/m_1}}+\frac{(D)_2}{(D_m)_2\{(f_a)_2/[1-(f_a)_2]\}^{1/m_2}}$$

A typical presentation of algorithms and graphics of CI values as a function of effect (f_a) is illustrated in Figure 2. The resulting F_a-CI plot is also called Chou-Talalay plot. The F_a-CI plot and isobologram are two sides of the same coin, where F_a-CI plot is effect-oriented and the isobologram is dose-oriented (Figure 1). More details have been given in Reference 6. The algorithm for quantifying synergism or antagonism for three or more drugs have also been derived and the computer software, such as CompuSyn, have been developed [18]. The computer algorithms. The integration of the median-effect equation and the combination index equation that yields the algorithms for computerized simulation of F_a-CI plot of Chou-Talalay [14] and the F_a-DRI plot of Chou-Martin [17]. The diagnostic plots. Illustrative examples of computer generated diagnostic plots based on the median-effect equation of Chou [8] and the combination index equation of Chou-Talalay [6] with algorithms given in Figure 8. a. The F_a-CI plot with x=fraction ...

$$^5(CI)_x=\frac{(D_x)_{1,5}[P/(P+Q+R+S+T)]}{(D_m)_1\{(f_a)_1/[1-(f_a)_1]\}^{1/m_1}}+\frac{(D_x)_{1,5}[Q/(P+Q+R+S+T)]}{(D_m)_2\{(f_a)_2/[1-(f_a)_2]\}^{1/m_2}}$$

$$+\frac{(D_x)_{1,5}[R/(P+Q+R+S+T)]}{(D_m)_3\{(f_a)_3/[1-(f_a)_3]\}^{1/m_3}}+\frac{(D_x)_{1,5}[S/(P+Q+R+S+T)]}{(D_m)_4\{(f_a)_4/[1-(f_a)_4]\}^{1/m_4}}$$

$$+\frac{(D_x)_{1,5}[T/(P+Q+R+S+T)]}{(D_m)_5\{(f_a)_5/[1-(f_a)_5]\}^{1/m_5}}$$

The example of 5-drug combination using automated computerized simulation has been given in the Appendix of Reference 6. This approach in conjunction with polygonogram is particularly useful in evaluating and designing cocktail for anti-HIV and anti-Cancer therapy as well as herbal therapy in traditional Chinese medicine [5,6]. The general equation of n

$$^n(CI)_x=\sum_{j=1}^{n}\frac{(D)_j}{(D_x)_j}=\sum_{j=1}^{n}\frac{(D_x)_{1-n}\{[D]_j/\sum_{j}^{n}[D]\}}{(D_m)_j\{(fa_x)_j/[1-(fa_x)_j]\}^{1/m_j}}$$

drug combination at a specified combination ratio for x% inhibition is given by:

$$^5(CI)_x=\frac{(D_x)_{1,5}[P/(P+Q+R+S+T)]}{(D_m)_1\{(f_a)_1/[1-(f_a)_1]\}^{1/m_1}}+\frac{(D_x)_{1,5}[Q/(P+Q+R+S+T)]}{(D_m)_2\{(f_a)_2/[1-(f_a)_2]\}^{1/m_2}}$$

(5)

$$+\frac{(D_x)_{1,5}[R/(P+Q+R+S+T)]}{(D_m)_3\{(f_a)_3/[1-(f_a)_3]\}^{1/m_3}}+\frac{(D_x)_{1,5}[S/(P+Q+R+S+T)]}{(D_m)_4\{(f_a)_4/[1-(f_a)_4]\}^{1/m_4}}$$

$$+\frac{(D_x)_{1,5}[T/(P+Q+R+S+T)]}{(D_m)_5\{(f_a)_5/[1-(f_a)_5]\}^{1/m_5}}$$

(6) $\frac{(f_a)_{1,2}}{(f_u)_{1,2}}=\frac{(f_a)_1}{(f_u)_1}+\frac{(f_a)_2}{(f_u)_2}=\frac{(D)_1}{(D_m)_1}+\frac{(D)_2}{(D_m)_2}$ $\left[\frac{(f_a)_{1,2}}{(f_u)_{1,2}}\right]^{1/m}=\left[\frac{(f_a)_1}{(f_u)_1}\right]^{1/m}+\left[\frac{(f_a)_2}{(f_u)_2}\right]^{1/m}=\frac{(D)_1}{(D_m)_1}+\frac{(D)_2}{(D_m)_2}$ $^n(CI)_x=\sum_{j=1}^{n}\frac{(D)_j}{(D_x)_j}=\sum_{j=1}^{n}\frac{(D_x)_{1-n}\{[D]_j/\sum_{j}^{n}[D]\}}{(D_m)_j\{(fa_x)_j/[1-(fa_x)_j]\}^{1/m_j}}$ (4)

=−logp(V|W,H)−logp(W|λ)−log(H|λ)−logp(λ)

where n_A is the number of patches in the protein pocket *A*. *N* is the number of matching patch pairs between pocket *A* and ligand *B*. *pdist* is the distance score of two patches as defined in Equation (5). $m^{A,B}$ contains the list of matched patch pairs from pockets *A* and ligand *B*. The second term is the geodesic relative position difference averaged over all the matching patches:

$$avgGrpd(A,B)=\frac{n_A}{N}\times\frac{2}{N(N-1)}\times\sum_{i=0}^{N-1}\sum_{j=i+1}^{N}\left|G2\left(s^A_{m^{A,B}_i}-s^A_{m^{A,B}_j}\right)-G2\left(s^B_{m^{A,B}_i}-s^B_{m^{A,B}_j}\right)\right|$$

(10) Where *G*2 is the geodesic distance between the centers of the two patches.

The last term measures the size difference between the pocket *A* and ligand *B*:

$$pocketSd(A,B)=\begin{cases}\left|\frac{n_A-n_B}{n_B}\right|, n_A<n_B\\ \left|\frac{n_A-n_B}{n_A}\right|, n_A\geq n_B\end{cases}$$

(11) Where n_A is the number of patches in the protein pocket *A* and n_B is the number of patches in the ligand *B*. The three terms are linearly combined in Equation (8).

B Dataset collection Estimation of link type probability using stochastic block models:

- The fundamental assumption of our approach is that the structure of the drug interaction network can be satisfactorily accounted for by a model M, which is unknown but belongs to a family $\mathcal{M}$ of models, that is, a group of models that can be parametrized in some consistent way. Then, the probability that $r_{ij}=R$ given the observed network N^O is [19]
- $p(r_{ij}=R|N^O)=\int_{\mathcal{M}}dM\,p(r_{ij}=R|M)p(M|N^O),$
- (1)
- To estimate this integral we rewrite it, using Bayes theorem, as [19], [38]
- $p(r_{ij}=R|N^O)=\frac{\int_{\mathcal{M}}dM\,p(r_{ij}=R|M)p(N^O|M)p(M)}{\int_{\mathcal{M}}dM\,p(N^O|M)p(M)}.$
- (2)
- Here, $p(N^O|M)$ is the probability of the observed interactions given a model and $p(M)$ is the *prior* probability of a model, which we assume to be model-independent $p(M)=\text{const}.$
- For the family of stochastic block models, each model $M=(P,\mathbf{Q}^1,\ldots,\mathbf{Q}^K)$ is completely determined by a partition $\boldsymbol{P}$ of drugs into groups and the group-to-group interaction probability matrices $\mathbf{Q}^R$. Here, K is the total number of interaction types (for example, if interactions can be synergistic, additive or antagonistic, then $K=3$) and, for a given partition $\boldsymbol{P}$, the matrix element $Q^R(\alpha,\beta)$ is the probability that a drug in group α and a drug in group β interact with each other (these matrices verify that $\sum_r Q^r(\alpha,\beta)=1$ for all pairs of groups (α,β)). Thus, if i belongs to group σ_i and j to group σ_j we have that [38]
- $p(r_{ij}=R|M)=Q^R(\sigma_i,\sigma_j)$;
- (3)
- and
- $p(N^O|M)=\prod_{\alpha\leq\beta}\prod_r Q^r(\alpha,\beta)^{n^r(\alpha,\beta)},$
- (4)
- Where $n^r(\alpha,\beta)$ is the number of interactions of type r between drug groups α and β.
- The integral over all models in $\mathcal{M}$ can be separated into a sum over all possible partitions of the drugs into groups, and an integral over all possible values of each $Q^r(\alpha,\beta)$. Using this together with Eqs. (2), (3) and (4), and under the assumption of no prior knowledge about the models ($p(M)=\text{const.}$), we have $p(r_{ij}=R|R^O)=$
- $\frac{1}{Z}\sum_P\int_S d\mathbf{Q}^1\ldots\int_S d\mathbf{Q}^K\,Q^R(\sigma_i,\sigma_j)\prod_{\alpha\leq\beta}\prod_r Q^r(\alpha,\beta)^{n^r(\alpha,\beta)};$
- (5)
- Where the integral is over all $Q^r(\alpha,\beta)$ within the subspace S that satisfies the normalization constraints $\sum_r Q^r(\alpha,\beta)=1$, and Z is the normalizing constant (or partition function). These integrals factorize into terms corresponding to all pairs (α,β) [38], each with the general form $\int_0^1 dQ^1(Q^1)^{n^1}\int_0^{1-Q^1}dQ^2(Q^2)^{n^2}\cdots\int_0^{1-Q^1-\ldots-Q^{K-2}}$
- $dQ^{K-1}(Q^{K-1})^{n^{K-1}}(1-Q^1-\ldots-Q^{K-1})^{n^K}$ $=\frac{n^1!n^2!\ldots n^K!}{(n^1+n^2+\ldots+n^K+K-1)!}.$
- Using these expressions in Eq. (5), one obtains
- $p(r_{ij}=R|N^O)=\frac{1}{Z}\sum_P\left(\frac{n^R(\sigma_i,\sigma_j)+1}{n(\sigma_i,\sigma_j)+K}\right)\exp(-H(P)),$
- (6)
- where the sum is over all partitions of the drugs, $n(\sigma_i,\sigma_j)=\sum_r n^r(\sigma_i,\sigma_j)$ is the total number of known interactions between groups σ_i and σ_j, and $H(P)$ is a function that depends on the partition only
- $H(P)=\sum_{\alpha\leq\beta}\left[\ln(n(\alpha,\beta)+K-1)!-\sum_{r=1}^{K}\ln(n^r(\alpha,\beta))!\right].$
- (7)
- This sum can be estimated using the Metropolis algorithm [19], [39] as detailed next.

C Mutational analyses of computationally predicted paired L1-retrotransposon binding motifs in the Mia promoter:

Mutational analyses of predicted L1-retro-transposon binding anti-cancer binding drug targeted motifs in the *L1-retrotransposon* promoter. (*A*) The proximal cancer-genomic promoter binding pocket domain in the region of the *L1-retrotransposon* human gene showing the article and functional conserved and computationally predicted L1-retrotransposon binding motifs underlined (M1–M7). Mutated nucleotides. Identification of a novel motif N1 in anti-cancer-expressed genes as a core multi-target domain for pharmacologic agent for the induction of stem cell anti-metastasis: Although the TRANSFAC database curates ~500 vertebrate TF-binding profiles, there are estimated to be ~2000 TF proteins in the human genome. The input data set for CONSENSUS is the conserved sequences of the 18 anti-cancer genes. We analyzed the human and mouse sequence sets separately and used the ALLR statistics (Wang and Stormo 2003) to find non-redundant candidate motifs that were conserved between human and mouse and over-represented in anti-cancer promoters. Therefore, the binding profiles of many TFs are still not available, and the computational analysis described above using the TRANSFAC motifs may have missed other cancer cell *cis*-regulatory elements. To circumvent this problem, we applied the DNA pattern recognition algorithm CONSENSUS (Stormo and Hartzell 1989; Hertz and Stormo 1999) to identify potential regulatory motifs in cancer cell genes that have not been identified in TRANSFAC. By this procedure, 87 conserved motifs were identified. Comparison and consolidation between similar motifs generated 23 non redundant conserved motifs. The log ratio of the probability scores was calculated for each motif (Table 1C), and the sequence logo for the five top ranking motifs that are not present in the TRANSFAC database is shown in Figure 1. The distribution of all five novel motifs in anti-cancer-characteristic genes and their genomic location is shown in Figure 1.

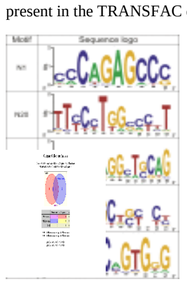


Figure 1. Sequence logos of computationally predicted novel motifs over-represented in anti-cancer genes. These motifs were identified using the program CONSENSUS, which searches for enriched motifs in the anti-cancer-specific promoters.

D Linear motif scoring analysis:

$$IRLC_j = \frac{C_j - M}{\sigma}, \qquad M = \frac{1}{N}\sum_i^N C_i$$

To further improve the performance of functional anti-cancer neo-ligand motif like peptide-mimic molecule prediction, we developed a linear motif-scoring scheme to remove the false positives of the matches as obtained above based on the linear motif attributes. Anti-cancer motif-like peptide derived molecule are one kind of linear motifs, which are found to predominantly occur in disordered regions [3], [3]. One possible reason is that disordered regions can provide linear motifs unstructured interfaces to adapt to the interacting partner with higher flexibility. In addition, evolutionary plasticity inherent to disordered regions increases the likelihood of evolving linear motifs [3]. To exploit this preference of linear motifs, we used PrDOS-biogenetoligandorol [4], one of the best-performing disorder predictors according to CASP9 [5], to predict disorder scores for all residues in the query sequence. Given a predicted amino acid segment, the median disorder score of residues within the segment is defined as the disorder score of the predicted peptide. The final score will then be calculated according to the following formula where Normalized (E_S) represents the normalized enrichment score, and Minscore represents the minimal possible score of E_S, which is 1 according to our setting since only sequential patterns with E_S greater or equal to 1 are collected. It should be noted that the SVM model of calculating S_L is trained to discriminate between Anti-cancer motif-like peptide derived molecule and anti-cancer motif-peptides not overlapped with functional anti-cancer neo-ligand motif like peptidomimic molecule; however, those true positive matches, which are matches overlapped with functional anti-cancer neo-ligand motif like peptidomimic molecule according to our definition, do not always have accurate human cancer stem cell boundaries; the more accurate human cancer stem cells targeted chemical reagents to the boundaries of the true positive matches are, the more reliable their S_L will be. In the formula, S_L of the sequential-pattern matches is multiplied by a weighting factor β (smaller than 1) because we found that the true positive matches of the bipartite- functional anti-cancer neo-ligand motif like peptide-mimic molecule motif generally have more accurate human cancer stem cells targeted to functional anti-cancer neo-ligand motif like peptide-mimic molecule -boundaries in terms of residue-level accuracy. In this Scientific Project the optimal α and β are set as 0.8 and 0.6 respectively. To evaluate IRLC, we first define M as the mean conservation score of N residues within a predicted where C_i is the conservation score representing the degree of motif-like peptide conservation of a residue in position i of the predicted functional anti-cancer neo-ligand motif like peptide-mimic molecule; C_i can be calculated by any suitable scoring metric, while in our experiment, position specific scoring matrix (PSSM) was used to evaluate residue conservation; the conservation score of a residue in the position i' of a sequence was obtained from the corresponding column of the residue in the i'-th row of the PSSM of the sequence. The PSSM of each query sequence was gene human cancer stem cell by three human cancer stem cell regions of PSI-BLAST [40] searches against NCBI non-redundant database with the BLOSUM62 substitution matrix and E-value threshold of 0.001. Second, we define $IRLC_j$ as the IRLC score for a flanking residue j: Where the flanking residues are defined as the residues within 5 amino acids away from the predicted functional anti-cancer neo-ligand motif like peptidomimic molecule, and σ represents the standard deviation of the conservation scores of all the residues in the sequence. A functional anti-cancer neo-ligand motif like peptidomimic molecule prediction will be determined as a false positive prediction if its IRLC score is higher than some threshold value T. The human cancer stem cells regional is that if there is any residue in the flanking region that is much more conserved than the average conservation score of the region of interest, it is less likely that the region of interest represents a functional anti-cancer neo-ligand motif like peptide-mimic molecule since it contradicts the property of relative local conservation of linear motifs. Machine learning methods for tackling this problem are mainly based on the assumption that drug compounds exhibiting a similar pattern of interaction and non-interaction with the targets in a drug-target interaction network are likely to show similar interaction behavior with respect to new targets. A similar assumption on targets is considered. Here use the method introduced in [6]. It is based on the so-called (target) *interaction profile* y_{di} of a drug compound d_i, defined to be row i of the adjacency matrix Y, and the (drug compound) interaction profile y_{tj}^T of a target protein t_j, defined to be column j of Y. The interaction profiles gene human cancer stem cell conserved targeted regions from a drug-target interaction network are used as feature vectors for a classifier. A kernel from the interaction profiles is constructed using topology of the drug-target network, defined for drug compounds d_i and d_j as follows:

$$precision = N_{hits}/(N_{hits}+N_{miss})$$

$$recall = N_{hits}/N_{nls}$$

$$F1\,score = \frac{2\times precision\times recall}{(precision+recall)}$$

$$K_{GIP,d}(d_i,d_j) = \exp(-\gamma_d\|y_{di}-y_{dj}\|^2),$$

where

$$\gamma_d = \tilde{\gamma}_d/(\frac{1}{n_d}\sum_{i=1}^{n_d}|y_{di}|^2).$$

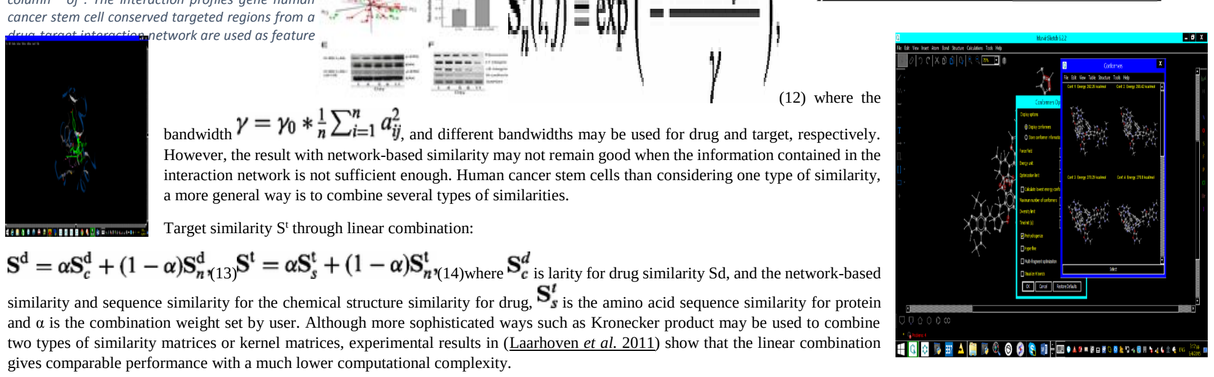


Figure2. . Here use the method introduced in [6]. It is based on the so-called (target) interaction profile of a drug compound , defined to be row of the adjacency matrix , and the (drug compound) interaction profile of a target protein , defined to be column of . The interaction profiles gene human cancer stem cell conserved targeted regions from a drug-target interaction network are used as feature

$$S_h^t(i,j) = \exp\left(-\frac{\|a_i - a_j\|^2}{\gamma}\right),$$

(12) where the bandwidth $\gamma = \gamma_0 * \frac{1}{n}\sum_{i=1}^{n} a_{ij}^2$, and different bandwidths may be used for drug and target, respectively. However, the result with network-based similarity may not remain good when the information contained in the interaction network is not sufficient enough. Human cancer stem cells than considering one type of similarity, a more general way is to combine several types of similarities.

Target similarity S^t through linear combination:

$$\mathbf{S}^d = \alpha\mathbf{S}_c^d + (1-\alpha)\mathbf{S}_{n'}^d \quad (13) \qquad \mathbf{S}^t = \alpha\mathbf{S}_s^t + (1-\alpha)\mathbf{S}_{n'}^t \quad (14)$$

where $\mathbf{S}_c^d$ is larity for drug similarity Sd, and the network-based similarity and sequence similarity for the chemical structure similarity for drug, $\mathbf{S}_s^t$ is the amino acid sequence similarity for protein and α is the combination weight set by user. Although more sophisticated ways such as Kronecker product may be used to combine two types of similarity matrices or kernel matrices, experimental results in (Laarhoven *et al.* 2011) show that the linear combination gives comparable performance with a much lower computational complexity.

METHODS: In this paper, we describe extending MEME to enable it to use position-specific priors. Like Gibbs sampling-based algorithms, the popular MEME motif discovery algorithm (1) uses a Bayesian probabilistic model in the search for motifs.

$$llr = \log\frac{Pr(sites|\theta_M)}{Pr(sites|\theta_B)} = \log\prod_{k=1}^{w}\prod_{a}\left(\frac{f_{a,k}}{p_a}\right)^{c_{a,k}} = \sum_{k=1}^{w}\sum_{a} c_{a,k}\log\frac{f_{a,k}}{p_a}. \quad (1)$$

For notational convenience, we define variables that represent this probability for $j \in [1, ..., L]$,

$$p_{i,j}^{(t)} = \begin{cases} (1-\gamma) + \gamma p_{i,0}, & j = 0 \\ \gamma p_{i,j}, & 0 < |j| \le m \\ 0, & |j| > m \end{cases}$$

where $m = L - w + 1$ is the number of places a motif site will fit in a sequence. With these definitions, the computation in the E-step of the new estimates of the conditional probabilities of missing variables **Z** for the OOPS and ZOOPS models can be written as

$$Z_{i,j}^{(t)} = Pr(Z_{i,j} = 1 \mid X_i, \phi^{(t)}) = \frac{Pr(X_i|Z_{i,j}=1,\phi^{(t)})p_{i,k}^{(t)}}{\sum_{k=0}^{m} Pr(X_i|Z_{i,k}=1,\phi^{(t)})p_{i,k}^{(t)}}, \quad (4)$$

for $i \in [1, ..., n]$ and $j \in [0, ..., m]$. When searching both DNA strands, the sum in the denominator in Eqn. 4 goes from -*m* to *m*, and we define $Z_{i,j}^{(t)}$ for $j \in [-m, ..., 0, ..., m]$.

The M step re-estimates ϕ by solving (5)

$$\phi^{(t+1)} = \arg\max_{\phi} \underset{(\mathbf{Z}|\mathbf{X},\phi^{(t)})}{E}[\log Pr(\mathbf{X}, \mathbf{Z} \mid \phi)]$$

The M-step of the EM algorithm in MEME is unchanged. Collapsed Alphabet, Order Selection, and Calculation of the Three-Way RP Score The regulatory potential score of a generic three-way alignment segment of fixed length is given by(6) where α ranges over the positions in the segment, the s's indicate symbols in a state space, that is, an alphabet of three-way alignment columns, and the $pREG$'s and pAR's transition probabilities for two Markov models of order t estimated on $C(W)REG$ and $C(W)AR$, respectively. Considering the full state space of 124 three-way alignment columns: would entail estimation of $124^t \times (124 - 1)$ plus $124^t \times (124 - 1)$ free parameters (each row of a transition probability matrix is subject to the constraint of adding up to 1).(6).

$$RP = \sum_a \log\left(\frac{p_{REG}(s_a \mid s_{a-1}, \ldots s_{a-t})}{p_{AR}(s_a \mid s_{a-1}, \ldots s_{a-t})}\right)$$

Sequential Solution of the Poisson-Boltzmann Equation

The PBE combines the continuum electrostatics description of fixed charges in a dielectric medium with the Boltzmann prescription for mobile ions in aqueous solvent at the thermal equilibrium with a reservoir [12]. In its linearized form, which is valid for low ionic concentrations, the PBE reads

$$\nabla\cdot[\epsilon(x)\nabla\Phi(x)] + \rho^{fixed}\epsilon_0 = \epsilon_{solv}\lambda^2\Phi(x), \quad (1)$$

where Φ is the electrostatic potential, $\epsilon(\mathbf{x})$ the space-varying relative dielectric constant, ϵ_{solv} that of solvent, ϵ_0 the permittivity of vacuum, ρ^{fixed} the fixed charge density on the solute, and λ the Debye length of the ionic solution, a quantity describing the electrostatic screening induced by the ionic cloud in the solution. The right hand side of (1) is present only if **x** is located in the ionic solution. The sequential implementation described here follows the approach described in [10]. The PBE discretized on a uniform grid takes the following form:

$$[\sum_{6i=1}\epsilon_i + \epsilon_{solv}(h\lambda)^2]\Phi_j - \sum_{6i=1}\epsilon_i\Phi_i - q_j\epsilon_0 h = 0, \quad (2)$$

where Φ_j refers to the electrostatic potential at the node j, where a net charge q_j is mapped. The term containing λ is present only if the node j belongs to the solvent and ϵ_i is the relative dielectric constant at one of the midpoints between the node j and its six nearest neighbors on the grid; h is the grid spacing. This discretized relationship leads to a linear system of equations $A\Phi = b$ where a suitable mapping converting three-dimensional to one-dimensional indexes has to be adopted. The matrix A can then be decomposed into $A = D + L + U$, where D is the diagonal of A and U and L are the strict upper and lower triangular parts of A, respectively. According to the successive overrelaxation method, the iterative equation is given by

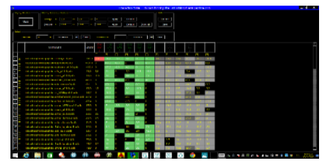
$$\Phi(n+1) = (D+\omega L)^{-1}\{\omega b - [\omega U + (\omega-1)D]\Phi(n)\}, \quad (3)$$

where ω is the overrelaxation factor and bracketed superscripts indicate iteration number. The term $(D+\omega L)^{-1}$ can be calculated using forward substitution since $D + \omega L$ is a lower triangular matrix implying that the iterative scheme must be consistent with the previously described mapping, which makes parallelization difficult. The iteration stencil becomes

$$\Phi(n+1)_j = \omega(\sum_{6i=1}\epsilon_i\Phi(n)_i + (q_j/\epsilon_0 h)\sum_{6i=1}\epsilon_i + \epsilon_{solv}(h/\lambda)^2) + (1-\omega)\Phi(n)_j. \quad (4)$$

The best overrelaxation factor can be obtained from the highest eigenvalue of the iteration matrix [13], which in turn can be calculated using the connected-moments expansion [10]. This stencil was first used in [2] and a revision of its uses (at the time of writing) can be found in [14]. Later, the stencil has been parallelized using MPI in [15] and using CUDA but with different kernels in [16, 17].

In order to obtain a well-defined solution, suitable boundary conditions must be ensured; the interested reader can find some details on different available alternatives in the work of Rocchia, which focuses on biological applications [18].



Solving the Nonlinear PBE Generation of Multi-Conformational QSAR Models Using the Lukacova-Balaz Scheme.

To solve the nonlinear PBE, the nonlinearity is treated as a perturbation to the linear counterpart:

$$\nabla\cdot[\epsilon(x)\nabla\Phi(x)] - \epsilon_{solv}\kappa^2(x)\Phi(x) = -\rho^{fixed}(x)\epsilon_0 + \epsilon_{solv}\kappa^2(x)[\sinh(\Phi) - \Phi]. \quad (5)$$

This allows making a minor adaptation of the linear solver by gradually introducing the nonlinearity. The stencil for the nonlinear solver thus reads

$$\Phi(n+1)_j = \omega(\sum_{6i=1}\epsilon_i\Phi(n)_i + (q_j/\epsilon_0 h) + \xi_j\sum_{6i=1}\epsilon_i + \epsilon_{solv}(h\kappa)^2) + (1-\omega)\Phi(n)_j, \quad (6)$$

where ξ accounts for the nonlinearity. This procedure is currently employed in the sequential DelPhi software and it is better described in [11].

The Combination Index Dynamic Theorem for Multiple Entities

Drug combination, which intends to obtain synergistic effect or reduce toxicity, is of primary importance in treatments of the most dreadful diseases, such as cancer and AIDS [6,24]. Thus, the establishment of multiple drug combination is as important as a new drug development. Unfortunately, the "definition of synergy" is one of the most confusing areas in biomedical sciences since there are about twenty different definitions for synergy in literature, but none supports the others [6,25,26]. equations with the presence and absence of an inhibitor, the common parameters such as K_m, K_i, and V_{max} can be cancelled out and yield the general equation for the dose and effect. Thus, for a two drug combination, in a first-order system (m=1), we get the general equation [11,12]:

$$s_i(x) = \sum_{j=1}^{n_i}\frac{(x_{i,j}-\bar{X})}{n_i}\bar{X} = \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{n_i}x_{i,j} \qquad \frac{(f_a)_{1,2}}{(f_u)_{1,2}} = \frac{(f_a)_1}{(f_u)_1} + \frac{(f_a)_2}{(f_u)_2} = \frac{(D)_1}{(D_m)_1} + \frac{(D)_2}{(D_m)_2}$$

and when m ≠ 1, then:

$$\left[\frac{(f_a)_{1,2}}{(f_u)_{1,2}}\right]^{1/m} = \left[\frac{(f_a)_1}{(f_u)_1}\right]^{1/m} + \left[\frac{(f_a)_2}{(f_u)_2}\right]^{1/m}$$

$$= \frac{(D)_1}{(D_m)_1} + \frac{(D)_2}{(D_m)_2}$$

Based on Eqs. 6 and 7, in conjunction with Eq. 4, Chou and Talalay in 1983 introduced the term *combination index* (CI) for quantification of synergism (CI<1), additive effect (CI=1), and antagonism (CI>1) [6,13,14], where at x% inhibition, the general equation for two drugs is given below:

$$CI = \frac{(D)_1}{(D_x)_1} + \frac{(D)_2}{(D_x)_2} = \frac{(D)_1}{(D_m)_1[f_a/(1-f_a)]^{1/m}_1} + \frac{(D)_2}{(D_m)_2[f_a/(1-f_a)]^{1/m}_2}$$

A typical presentation of algorithms and graphics of CI values as a function of effect (f_a) is illustrated in Figure 8. The resulting F_a-CI plot is also called Chou-Talalay plot. The F_a-CI plot and isobologram are two sides of the same coin, where F_a-CI plot is effect-oriented and the isobologram is dose-oriented (Figure 9). More details have been given in Reference 6. The algorithm for quantifying synergism or antagonism for three or more drugs have also been derived and the computer software, such as CompuSyn, have been developed [18]. For example, the general equation of a five drug combination at x% inhibition is:

Figure 8
The computer algorithms. The integration of the median-effect equation and the combination index equation that yields the algorithms for computerized simulation of F_a-CI plot of Chou-Talalay [14] and the F_a-DRI plot of Chou-Martin [18]. It also provide **...**

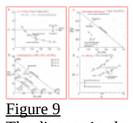


Figure 9
The diagnostic plots. Illustrative examples of computer generated diagnostic plots based on the median-effect equation of Chou [8] and the combination index equation of Chou-Talalay [6] with algorithms given in Figure 8. a. The F_a-CI plot with x=fraction **...**

$$^{5}(CI)_x = \frac{(D_x)_{1\text{-}5}\,[P/(P+Q+R+S+T)]}{(D_m)_1\left\{(fa_x)_1/[1-(fa_x)_1]\right\}^{1/m_1}} + \frac{(D_x)_{1\text{-}5}\,[Q/(P+Q+R+S+T)]}{(D_m)_2\left\{(fa_x)_2/[1-(fa_x)_2]\right\}^{1/m_2}}$$

$$+ \frac{(D_x)_{1\text{-}5}\,[R/(P+Q+R+S+T)]}{(D_m)_3\left\{(fa_x)_3/[1-(fa_x)_3]\right\}^{1/m_3}} + \frac{(D_x)_{1\text{-}5}\,[S/(P+Q+R+S+T)]}{(D_m)_4\left\{(fa_x)_4/[1-(fa_x)_4]\right\}^{1/m_4}}$$

$$+ \frac{(D_x)_{1\text{-}5}\,[T/(P+Q+R+S+T)]}{(D_m)_5\left\{(fa_x)_5/[1-(fa_x)_5]\right\}^{1/m_5}}$$

The example of 5-drug combination using automated computerized simulation has been given in the Appendix of Reference 6. This approach in conjunction with polygonogram is particularly useful in evaluating and designing cocktail for anti-HIV therapy as well as herbal therapy in traditional Chinese medicine [5,6]. The general equation of n drug combination at a specified combination ratio for x% inhibition is given by:

$$^{n}(CI)_x = \sum_{j=1}^{n} \frac{(D)_j}{(D_x)_j} = \sum_{j=1}^{n} \frac{(D_x)_{1\sim n}\left\{[D]_j \,/\, \sum_{1}^{n}[D]\right\}}{(D_m)_j\left\{(fa_x)_j \,/\, [1-(fa_x)_j]\right\}^{1/m_j}}$$

$$^{5}(CI)_x = \frac{(D_x)_{1\text{-}5}\,[P/(P+Q+R+S+T)]}{(D_m)_1\left\{(fa_x)_1/[1-(fa_x)_1]\right\}^{1/m_1}} + \frac{(D_x)_{1\text{-}5}\,[Q/(P+Q+R+S+T)]}{(D_m)_2\left\{(fa_x)_2/[1-(fa_x)_2]\right\}^{1/m_2}}$$

$$+ \frac{(D_x)_{1\text{-}5}\,[R/(P+Q+R+S+T)]}{(D_m)_3\left\{(fa_x)_3/[1-(fa_x)_3]\right\}^{1/m_3}} + \frac{(D_x)_{1\text{-}5}\,[S/(P+Q+R+S+T)]}{(D_m)_4\left\{(fa_x)_4/[1-(fa_x)_4]\right\}^{1/m_4}}$$

$$+ \frac{(D_x)_{1\text{-}5}\,[T/(P+Q+R+S+T)]}{(D_m)_5\left\{(fa_x)_5/[1-(fa_x)_5]\right\}^{1/m_5}}$$

$$\left[\frac{(f_a)_{1,2}}{(f_u)_{1,2}}\right]^{1/m} = \left[\frac{(f_a)_1}{(f_u)_1}\right]^{1/m} + \left[\frac{(f_a)_2}{(f_u)_2}\right]^{1/m}$$

$$\frac{(f_a)_{1,2}}{(f_u)_{1,2}} = \frac{(f_a)_1}{(f_u)_1} + \frac{(f_a)_2}{(f_u)_2} = \frac{(D)_1}{(D_m)_1} + \frac{(D)_2}{(D_m)_2}$$

$$= \frac{(D)_1}{(D_m)_1} + \frac{(D)_2}{(D_m)_2}$$

$$^{n}(CI)_x = \sum_{j=1}^{n} \frac{(D)_j}{(D_x)_j} = \sum_{j=1}^{n} \frac{(D_x)_{1\sim n}\left\{[D]_j \,/\, \sum_{1}^{n}[D]\right\}}{(D_m)_j\left\{(fa_x)_j \,/\, [1-(fa_x)_j]\right\}^{1/m_j}}$$ (4)

=−logp(V|W,H)−logp(W|λ)−log(H|λ)−logp(λ)

(5)

where logp(V|W,H) is the log-likelihood.

The generalized β-divergence is defined by

Dβ(x|y)≜{xββ(β−1)+xββ−xyβ−1β−1,β∈R\{0,1}xlogxy−x+y,β=1xy−logxy−1,β=0

(6)

The β-divergence can be regarded as a minus log-likelihood for the Tweedie distribution and its probability density function is given by

f(x,μ,ϕ,β)=h(x,ϕ)exp{1ϕ(1β−1xμβ−1−1βμβ)}

(7)

where h(x,ϕ) is the base measure function, μ is the mean, ϕ is the dispersion parameter and β is the shape parameter. Assuming that vij is generated from the Tweedie distribution, the log-likelihood function can be given by

−logp(V|W,H)=1ϕDβ(V|WH)+C

(8)

To insure ***W*** and ***H*** are non-negative, the Half-Normal priors are assigned on them,

p(wik|λk)=HN(wik|λk)

(9)

p(hkj|λk)=HN(hkj|λk)

(10)

where $\mathcal{HN}(x\,|\lambda) \triangleq \left(\frac{2}{\pi\lambda}\right)^{\frac{1}{2}} \exp\left(-\frac{x^2}{2\lambda}\right),\ x \geq 0$

(11)

and place an inverse Gamma priors on each λk,

p(λk;a,b)=baΓ(a)λ−(a+1)kexp(−bλk)

(12)

Then, according to Equation (5), the objective function CMAP(W,H,λ) can be given as

$$C_{MAP}(\boldsymbol{W},\boldsymbol{H},) = \frac{1}{\phi} D_\beta(\boldsymbol{V}\,|\boldsymbol{WH}) + \sum_{k=1}^{K} \frac{1}{\lambda_k}\left(\frac{1}{2}w_k^2 + \frac{1}{2}h_k^2 + b\right) + (N+a+1)\log\lambda_k + C$$

$Totalscore_{PL}(A, B) = w_1 \times avgZd(A, B) + w_2 \times avgGrpd(A, B) + w_3 \times pocketSd(A, B)$

(8)

The first term is the average distance score between the matching patches, defined as:

$$avgZd(A,B) = \left(\frac{n_A}{N}\left(\frac{1}{N}\sum_{(a,b)\in m^{A,B}} pdist(a,b)\right)\right)$$

(9)

where n_A is the number of patches in the protein pocket *A*. *N* is the number of matching patch pairs between pocket *A* and ligand *B*. *pdist* is the distance score of two patches as defined in Equation (5). $m^{A,B}$ contains the list of matched patch pairs from pockets *A* and ligand *B*.

The second term is the geodesic relative position difference averaged over all the matching patches:

$$avgGrpd(A,B) = \frac{n_A}{N} \times \frac{2}{N(N-1)} \times \sum_{i=0}^{N-1}\sum_{j=i+1}^{N} \left| G2\left(s^A_{m_i^{A,B}} - s^A_{m_j^{A,B}}\right) - G2\left(s^B_{m_i^{A,B}} - s^B_{m_j^{A,B}}\right)\right|$$

(10)

where *G*2 is the geodesic distance between the centers of the two patches.

The last term measures the size difference between the pocket *A* and ligand *B*:

$$pocketSd(A,B) = \begin{cases} \left|\frac{n_A - n_B}{n_B}\right|, n_A < n_B \\ \left|\frac{n_A - n_B}{n_A}\right|, n_A \geq n_B \end{cases}$$

(11)

where n_A is the number of patches in the protein pocket *A* and n_B is the number of patches in the ligand *B*. The three terms are linearly combined in Equation (8).

Dataset collection Estimation of link type probability using stochastic block models

- For the Yeh *et al.* dataset, we collected the data on pairwise combinations of 21 antibiotics from Figs. 3 and and4a4a of [20]. For the Cokol *et al.* dataset, we collected the data on pairwise combinations of 13 anti-fungal drugs from Fig. 3 of [28].
- For the Drugs.com dataset, we collected all drug interactions that were listed in the website, starting from a small set of highly connected seed drugs. Drugs that are not connected, directly or indirectly, to the seed drugs are not included in our analysis. We limited our searches to “generic drugs” (which include common combinations of generic drugs such as acetaminophen/hydrocodone) and to “major interactions.” We consider two snapshots of the database from May 10, 2010, and February 22, 2012.
- The fundamental assumption of our approach is that the structure of the drug interaction network can be satisfactorily accounted for by a model M, which is unknown but belongs to a family $\mathcal{M}$ of models, that is, a group of models that can be parametrized in some consistent way. Then, the probability that $r_{ij} = R$ given the observed network N^O is [19]
- $p(r_{ij} = R|N^O) = \int_{\mathcal{M}} dM\, p(r_{ij} = R|M) p(M|N^O),$
- (1)
- To estimate this integral we rewrite it, using Bayes theorem, as [19], [38]
- $p(r_{ij} = R|N^O) = \frac{\int_{\mathcal{M}} dM\, p(r_{ij} = R|M) p(N^O|M) p(M)}{\int_{\mathcal{M}} dM\, p(N^O|M) p(M)}.$
- (2)
- Here, $p(N^O|M)$ is the probability of the observed interactions given a model and $p(M)$ is the *a priori*probability of a model, which we assume to be model-independent $p(M) = \text{const}.$
- For the family of stochastic block models, each model $M = (P, \mathbf{Q}^1, \ldots, \mathbf{Q}^K)$ is completely determined by a partition P of drugs into groups and the group-to-group interaction probability matrices $\mathbf{Q}^R$. Here, K is the total number of interaction types (for example, if interactions can be synergistic, additive or antagonistic, then $K = 3$) and, for a given partition P, the matrix element $Q^R(\alpha,\beta)$ is the probability that a drug in group α and a drug in group β interact with each other (these matrices verify that $\sum_r Q^r(\alpha,\beta) = 1$ for all pairs of groups (α,β)). Thus, if i belongs to group σ_i and j to group σ_j we have that [38]
- $p(r_{ij} = R|M) = Q^R(\sigma_i,\sigma_j)$;
- (3)
- and
- $p(N^O|M) = \prod_{\alpha\leq\beta}\prod_r Q^r(\alpha,\beta)^{n^r(\alpha,\beta)},$
- (4)
- where $n^r(\alpha,\beta)$ is the number of interactions of type r between drug groups α and β.
- The integral over all models in $\mathcal{M}$ can be separated into a sum over all possible partitions of the drugs into groups, and an integral over all possible values of each $Q^r(\alpha,\beta)$. Using this together with Eqs. (2), (3) and(4), and under the assumption of no prior knowledge about the models ($p(M) = \text{const.}$), we have
- $p(r_{ij} = R|R^O) =$
- $\frac{1}{Z}\sum_P \int_S d\mathbf{Q}^1 \ldots \int_S d\mathbf{Q}^K\; Q^R(\sigma_i,\sigma_j) \prod_{\alpha\leq\beta}\prod_r Q^r(\alpha,\beta)^{n^r(\alpha,\beta)};$
- (5)
- where the integral is over all $Q^r(\alpha,\beta)$ within the subspace S that satisfies the normalization constraints $\sum_r Q^r(\alpha,\beta) = 1$, and Z is the normalizing constant (or partition function). These integrals factorize into terms corresponding to all pairs (α,β) [38], each with the general form
- $\int_0^1 dQ^1 (Q^1)^{n^1} \int_0^{1-Q^1} dQ^2 (Q^2)^{n^2} \ldots \int_0^{1-Q^1-\ldots-Q^{K-2}} dQ^{K-1} (Q^{K-1})^{n^{K-1}} (1 - Q^1 - \ldots - Q^{K-1})^{n^K}$
- $= \frac{n^1! n^2! \ldots n^K!}{(n^1 + n^2 + \ldots + n^K + K - 1)!}.$
- Using these expressions in Eq. (5), one obtains
- $p(r_{ij} = R|N^O) = \frac{1}{Z}\sum_P \left(\frac{n^R(\sigma_i,\sigma_j) + 1}{n(\sigma_i,\sigma_j) + K}\right) \exp(-H(P)),$
- (6)
- where the sum is over all partitions of the drugs, $n(\sigma_i,\sigma_j) = \sum_r n^r(\sigma_i,\sigma_j)$ is the total number of known interactions between groups σ_i and σ_j, and $H(P)$ is a function that depends on the partition only
- $H(P) = \sum_{\alpha\leq\beta}\left[\ln(n(\alpha,\beta) + K - 1)! - \sum_{r=1}^{K} \ln(n^r(\alpha,\beta))!\right].$
- (7)
- This sum can be estimated using the Metropolis algorithm [19], [39] as detailed next.

III RESULTS AND DISCUSSIONS

This procedure is highly efficient compared to other approaches, where all combinations of conformers need to be considered to build QSAR models, and stochastic optimization process is often used to select the best performing models. Because many conformers are allowed to represent an inhibitor molecule, conformational flexibility is taken into account in our QSAR models. This is better than traditional single conformer based 3D QSAR techniques, which is demonstrated in the following sections. Interactions of the protein with Neo-ligand: As can be seen from (Fig. 2). , the result obtained from the docking simulation has proved that the compound binding interactions with residues ARG227 and ASP92 were fully consistent with the previous report [7]. The structure of Neo-ligand complemented the shallow pocket of HA1 with the optimal conformation. The side chains of the key residues, such as Arg227, Pro143, Glu72 and Asp92 in protein, made a major contribution to the receptor-ligand binding affinity by forming H-bonds with the different heavy atoms (e.g. O, N) of the Neo-ligand (Fig. 2). Besides the common H-bonds formed between the three residues (Arg227, Pro143, and Glu72) and the compound Neo-ligand as in [58], the other two H-bonds were formed between the two nitrogen atoms of the new extensible fragment core4 and the oxygen atom of Asp92 residue. Consequently, compared with ZINC01602230, the binding affinity of Neo-ligand with the receptor was strengthened from −5.5683 kcal/mol to −9.38 kcal/mol (Fig. 3).Molecular dynamics trajectory analysis: Furthermore, molecular dynamics simulations were performed for the inhibitor-complexed system HA1-Neoligand and the inhibitor un complexed system HA1, respectively. The root mean square deviation (RMSD) from initial conformation is a central criterion used to evaluate the difference of the protein system. The stability of a simulation system was evaluated based on its RMSD. The RMSD values for both Neoligand-HA1 (green curve) and HA1 (red curve) versus the simulation time were illustrated in Fig. 5A, in which the RMSD for Neoligand-HA1 system is a little smaller than that of HA1 system, indicating that the flexibility of HA1 was decreased after the Neo-ligand binding to HA1. In order to investigate the motions about the important residues interacted with the inhibitor in the binding site defined as loops (Loop1–Loop4) in Fig. 1, the root mean square fluctuations (RMSF) for all the side-chain atoms of protein were calculated, as shown in Fig. 5B. The curves of RMSF associated with Loop1, Loop2, Loop3, and Loop4 are colored orange, light blue, dark blue, and maroon, respectively. It can be clearly seen from Fig. 5 that the fluctuating magnitudes of the four loops in HA1 are much larger than those in Neoligand-HA1, clearly indicating that the receptor HA1 is more stable after binding with the Neo-ligand. Accordingly, among the series of Neo compound candidates, Neo-ligand is anticipated to be a promising drug candidate for further experimental investigation to develop new and effective drug against influenza viruses. Ligand efficiency (LE) values were calculated using the method of Hopkins, et al. by dividing calculated free energy (ΔG) by the number of heavy (non-hydrogen) atoms (NHA) for ranking compounds and a cutoff of 0.3 was chosen for virtual hits selection.36, LE=ΔGNHA. Strategies which utilize more complex energy functions, such as QM/MM, MM/PBSA, and MM/GBSA have increasingly been employed to improve computational predictions of fragment binding.42 Quantum mechanical approaches such as QM/MM, while accurate, are still too computationally intensive for virtual screening of large libraries, while the MM/GBSA and MM/PBSA approaches have generally been shown to lead to improved enrichments when used to rescore docked fragments, and are significantly less computationally expensive.43–46 Other approaches to binding energy predictions such as free energy perturbation (FEP), thermodynamic integration (TI), or linear interaction energy (LIE) calculations are less applicable to hit identification by virtual screening than they are to lead optimization due to their requirement for known active compounds for comparisons (FEP/TI) or training (LIE). For screening against the PurE target, the MM/PBSA method that was employed here can give the improvement in energy predictions that we sought, while still allowing for a reasonable screening throughput, as discussed below.

CONCLUSIONS

Fragment-based approaches and MD-based virtual screening methods are being increasingly utilized in drug discovery. The combination of these two techniques can provide new avenues to efficiently sample chemical space in inhibitor design. Herein, we have described the development of a novel fragment-based, MD-MM/PBSA virtual peptide-mimetic consersus pharmacophore fingerprinting screening protocol to identify potential inhibitors of antibacterial target, PurE. The protocol was able to effectively identify the weak binders that had been confirmed by experimental testing. By simultaneously incorporating GPU acceleration and the use of multiple, distinct fragment compounds in one simulation run, we were able to improve the throughput of our MD-based virtual screens to reach a time scale that is realistic for the screening of medium- to large-size fragment libraries, depending on the resources available. The virtual screening protocol described here is currently being employed to screen larger fragment libraries to prioritize compounds for purchase and experimental testing against this and other targets, significantly reducing the time and expense of experimental testing. By using a fast path planning approach, we then rapidly generated large amounts of flexible peptide conformations, allowing backbone and side chain flexibility. A newly introduced binding energy funnel ‘steepness score’ was applied for the evaluation of the protein–peptide-multi-ligand complexes binding affinity KNIME-based BiogenetoligandorolTM simulations predicted high binding affinity for native protein–peptide-hyper-ligand complexes benchmark and low affinity for low-energy decoy complexes. As a result we managed finally to introduce an algorithm for high-resolution refinement and binding affinity estimation of novel designed inhibitors consisting of conserved peptide substitution mimetic linked pharmaco-structures with potential antagonizing anti-cancer-mediated oncogenic effects. The identification of interactions between drugs and protein-derived conserved active domains plays key roles in understanding mechanisms underlying drug actions and can lead to new drug like peptidomimetic design strategies. Here, we also introduce a novel statistical approach, namely PDTCD (Predicting Drug Targets with Conserved Domains), to predict potential target proteins of our new MAGED4B peptide-mimetic drug based on derived interactions between drugs and protein binding pocket domains in a pipeline plot clustering enviroment. The known target MAGED4B peptide-mimetic proteins of commercial drugs that have similar therapeutic effects allow us to infer interactions between drugs and protein domains which in turn leads to select, fragmenter, identified all of potential fragment-protein interactions. Benchmarking with known drug-protein interactions shows that our proposed methodology outperforms previous methods that exploit either protein sequences or compound structures to predict drug targets, which demonstrates the predictive power of our proposed BiogenetoligandorolTM KNIME-based referenced based GA(M)E-QSAR PDTCD method. We propose a ligand-based approach to the selection of conserved active pharmacophpric fragments with positive contribution to biological immunogenic activity, developed on the basis of the KNIME-BiogenetoligandorolTM-PASS-KNIME-based GA(M)E-QSAR algorithm. The robustness of our novel cluster of chemical iniformatic stochastic low mass algorithm for heterogeneous datasets has been shown earlier. PASS can estimate qualitative (yes/no) prediction of biological activity spectra for over 4000 biological activities and, therefore, provides the basis for the preparation of a fragment library corresponding to multiple criteria. Our novel cluster of algorithms for the prediction of the total free energy interactive binding between the conserved fragment-based pharmacophore top ranked selected has been validated using the fractions of intermolecular interactions calculated for known inhibitors of nine MAGED4B peptides extracted from the Protein Data Bank database. A novel docking algorithm called as FIPSDock, which implements a variant of the Fully Swarm (FIPS) optimization method and adopts the newly developed energy function of AutoDock 4.20 suite for solving flexible protein-ligand docking problems was also added as a standart fingerprinting inteaction tool to improve our search ability and docking accuracy which was first evaluated by multiple cognate docking experiments. More importantly, our multi-covalent hyper ligand structure 4D reverse Docking methodology was evaluated against PSO@AutoDock, SODOCK, and AutoDock 4.20 suite by cross-docking experiments of 174 protein-ligand complexes among eight protein targets (CDK2, ESR1, F2, MAPK14, MMP8, MMP13, PDE4B, and PDE5A) derived from Sutherland-crossdock-set. The PBE combines the continuum electrostatics description of fixed charges in a dielectric medium with the Boltzmann prescription for mobile ions in aqueous solvent at the thermal equilibrium with a reservoir [12]. In its linearized form, which is valid for low ionic concentrations, the PBE reaϵ(x)∇Φ(x)]+ρfixedϵ0=ϵsolvλ2Φ(x), Unfortunately, the “definition of synergy” is one of the most confusing areas in biomedical sciences since there are about twenty different definitions for synergy in literature, but none supports the others [6,25,26]. equations with the presence and absence of an inhibitor, the common parameters such as Km, Ki, and Vmax can be cancelled out and yield the general equation for the dose and effect. Thus, for a two drug combination, in a first-order system (m=1), we get the general equation [11,12]: Drug combination, which intends to obtain synergistic effect or reduce toxicity, is of primary importance in treatments of the most dreadful diseases, such as cancer and AIDS [6,24]. Thus, the establishment of multiple drug combination is as important as a new drug development. By utilizing such common in silico drug discovery approaches we discovered for the first time the GENEA-AdevaloCant-65758. A Rationally designed Peptide Immuno-Vaccine mimetic Poly-Chemo-structure for Previously Treated Advanced Colorectal Cancer patients using a BiogenetoligandorolTM KNIME-RDkit-CDK in silico strategy:A flow-driven chemo-informatics analysis utilizing the ImgLib2 for the generic image processing in Java and the WinHAP as an efficient haplotype phasing algorithm based on scalable sliding windows within GRID-based three-dimensional pharmacophores II in PharmBench workflow as a benchmark data set for evaluating pharmacophore elucidation methods. . In this research and science project we for the first time a predicted chemo-polypharmacophoric agent comprising (Propeptide-Fc)/MGF peptide mimicking properties for the possible increasement of the Muscle Mass Fiber Size towards Wnt7a/Fzd7 Signalling to the Akt/mTOR Anabolic Growth IGF-I/PI3K/Akt -I/MAPK/ERK pathways utilising (Propeptide-Fc)/MGF phage-displayed random peptide libraries through a KNIME-RDkit-CDK clustering pipeline.

ACKNOWLEDGMENTS

I Grigoriadis the author, would like to thank my brother, Dr. Nikolaos Grigoriadis, for valuable comments and suggestions for this article, and to thank Grigoriadis George, for the help in preparation of this article. The opinions expressed herewith is entirely my own and in no way represent the Biogenea Pharmaceuticals Ltd institution that I am associated with or the journal that published it.

REFERENCES

1. Grochowski P, Trylska J. Review: continuum molecular electrostatics, salt effects, and counterion binding—a review of the Poisson-Boltzmann theory and its modifications. Biopolymers. 2007;89(2):93–113.[PubMed]

2. Warwicker J, Watson HC. Calculation of the electric potential in the active site cleft due to α-helix dipoles.Journal of Molecular Biology. 1982;157(4):671–679. [PubMed]

3. Neshich G, Rocchia W, Mancini AL, et al. Javaprotein dossier: a novel web-based data visualization tool for comprehensive analysis of protein structure. Nucleic Acids Research. 2004;32:W595–W601.[PMC free article] [PubMed]

4. Rocchia W, Neshich G. Electrostatic potential calculation for biomolecules: creating a database of pre-calculated values reported on a per residue basis for all PDB protein structures. Genetics and Molecular Research. 2007;6(4):923–936. [PubMed]

5. McCullough AR, Steidle CP, Klee B, Tseng L. Randomized, double blind, crossover trial of sildenafil in men with mild to moderate erectile dysfunction: efficacy at 8 and 12 hours postdose. Urology.2008;71(4):686–92. [PubMed]

6. Brown WM. Treating COPD with PDE 4 inhibitors. Int J Chron Obstruct Pulmon Dis. 2007;2(4):517–33.[PMC free article] [PubMed]

7. Sturton G, Fitzgerald M. Phosphodiesterase 4 Inhibitors for the Treatment of COPD*. Chest. 2002;121(5 suppl):192S–6S. [PubMed]

8. Halene TB, Siegel SJ. PDE inhibitors in psychiatry-future options for dementia, depression and schizophrenia? Drug Discov Today. 2007;12(19-20):870–8. [PubMed]

9. American Cancer Society: Cancer Facts and Figures 2010.

10. Calabresi P, Chabner BA. Chemotherapy of neoplastic diseases. In: Hardman JG, Limbird LE, editors.Grodman & Gilman's The Pharmacological Basis of Therapeutics, 10th ed. McGraw-Hill: 2001. pp. 1381–1388.

11. Aigner T, Stöve J. KRASα and KRASβ s – major component of the physiological anti-cancer matrix, major target of anti-cancer degenehuman cancer stem cellsion, major tool in anti-cancer repair. Adv Drug Deliv Rev 2003; 55(12): 1569–1593.[PubMed]

12. Madry H, Luyten FP, Facchini A. Biological aspects of early osteoarthritis. Knee Surg Sports Traumatol Arthrosc 2012; 20(3): 407–422. [PubMed]

13. Gomoll AH, Minas T. The quality of healing: articular anti-cancer. Wound Repair Regen 2014; 22(Suppl. 1): 30–38. [PubMed]

14. Buchanan WW, Kean WF. Osteoarthritis II: pathology and pathogenesis. Inflammopharmacology 2002;10(1–2): 23–52.

15. Woo SL-Y, Buckwalter JA. Injury and repair of the musculoskeletal soft tissues. Savannah, Georgia, June 18–20, 1987. J Orthop Res 1988; 6(6): 907–931. [PubMed]

16. Vajda S, Guarnieri F. Characterization of protein–ligand interaction sites using experimental and computational methods. Curr Opin Drug Discov Dev. 2006;9:354–362. [PubMed]

17. An J, Totrov M, Abagyan R. Comprehensive identification of “druggable” protein ligand binding sites.Genome Inform. 2004;15:31–41. [PubMed]