Structural bioinformatics

GBPM: GRID-based pharmacophore model: concept and application studies to protein–protein recognition

Francesco Ortuso\(^1\), Thierry Langer\(^2\) and Stefano Alcaro\(^1,\)*

\(^1\)Dipartimento di Scienze Farmacobiologiche Università di Catanzaro ‘Magna Græcia’, Complesso Ninì Barbieri I-88021, Roccetella di Borgia (CZ), Italy and \(^2\)Institute of Pharmacy, Department of Pharmaceutical Chemistry, Innrain 52a, University of Innsbruck, A-6020 Innsbruck, Austria

Received on January 30, 2006; revised on March 3, 2006; accepted on March 22, 2006

Advance Access publication March 27, 2006

Associate Editor: Alfonso Valencia

ABSTRACT

**Motivation:** Automatic procedures to obtain pharmacophore models from experimentally determined macromolecular complexes can help in the drug discovery process, especially when protein–protein recognition plays an important biological role.

**Results:** The GRID-based pharmacophore model (GBPM) is a fully objective method for defining most relevant interaction areas in complexes deriving pharmacophore models from three-dimensional (3D) molecular structure information. It is based on logical and clustering operations with 3D maps computed by the GRID program on structurally known molecular complexes. In this manuscript we describe the concept and discuss application examples regarding protein–protein recognition. In particular two complexes selected in the Protein Data Bank have been tested to evaluate the GBPM capability to identify interaction areas. The results obtained show the capabilities of this new bioinformatic method.

**Availability:** The GBPM method has not been developed as a new computational code. It is based on the combination of existing scientific programs.

**Contact:** alcaro@unicz.it

**INTRODUCTION**

Pharmacophore models are useful tools for drug discovery and lead optimization processes (Milne, 1998). Usually they are created collecting most relevant structural features of biological active molecules. In most cases chemical intuition is necessary for completing the pharmacophore generation, in ambiguous cases possibly leading to erroneous models. One of the most advanced application of pharmacophore model is the virtual screening of large compound databases against multiple macromolecular targets (Langer and Krovat, 2003). This emerging bioinformatic technique, now considered as a new source of novel drug leads (Alvarez, 2004), is especially for bound organic compounds. The GREAT and GRIN modules of the GRID software help to this task and allow the preparation to the GRID mapping procedure. Assuming that the complex has two interacting molecules \(a\) and \(b\), as in the case of protein–protein complexes, the main goal of this first step is to get three structures, respectively, with the \(a + b, a\) and \(b\) subunits keeping the atomic coordinates of the original PDB model (Fig. 1).

The second step performs the GRID calculation with a given probe onto the three subunit models. In order to make the comparison of the map files as easy as possible, the matrix dimension of the GRID box is exactly main-

\[ A \cap B \]

\[ A \cup B \]

\[ n \]

\[ C \]

\[ D \]

\[ E \]

\[ F \]

\[ G \]

\[ H \]

\[ I \]

\[ J \]

\[ K \]

\[ L \]

\[ M \]

\[ N \]

\[ O \]

\[ P \]

\[ Q \]

\[ R \]

\[ S \]

\[ T \]

\[ U \]

\[ V \]

\[ W \]

\[ X \]

\[ Y \]

\[ Z \]

\[ 0 \]

\[ 1 \]

\[ 2 \]

\[ 3 \]

\[ 4 \]

\[ 5 \]

\[ 6 \]

\[ 7 \]

\[ 8 \]

\[ 9 \]

\[ a \]

\[ b \]

\[ c \]

\[ d \]

\[ e \]

\[ f \]

\[ g \]

\[ h \]

\[ i \]

\[ j \]

\[ k \]

\[ l \]

\[ m \]

\[ n \]

\[ o \]

\[ p \]

\[ q \]

\[ r \]

\[ s \]

\[ t \]

\[ u \]

\[ v \]

\[ w \]

\[ x \]

\[ y \]

\[ z \]

\[ A \]

\[ B \]

\[ C \]

\[ D \]

\[ E \]

\[ F \]

\[ G \]

\[ H \]

\[ I \]

\[ J \]

\[ K \]

\[ L \]

\[ M \]

\[ N \]

\[ O \]

\[ P \]

\[ Q \]

\[ R \]

\[ S \]

\[ T \]

\[ U \]

\[ V \]

\[ W \]

\[ X \]

\[ Y \]

\[ Z \]

\[ a \]

\[ b \]

\[ c \]

\[ d \]

\[ e \]

\[ f \]

\[ g \]

\[ h \]

\[ i \]

\[ j \]

\[ k \]

\[ l \]

\[ m \]

\[ n \]

\[ o \]

\[ p \]

\[ q \]

\[ r \]

\[ s \]

\[ t \]

\[ u \]

\[ v \]

\[ w \]

\[ x \]

\[ y \]

\[ z \]

\[ 0 \]

\[ 1 \]

\[ 2 \]

\[ 3 \]

\[ 4 \]

\[ 5 \]

\[ 6 \]

\[ 7 \]

\[ 8 \]

\[ 9 \]

\[ A \]

\[ B \]

\[ C \]

\[ D \]

\[ E \]

\[ F \]

\[ G \]

\[ H \]

\[ I \]

\[ J \]

\[ K \]

\[ L \]

\[ M \]

\[ N \]

\[ O \]

\[ P \]

\[ Q \]

\[ R \]

\[ S \]

\[ T \]

\[ U \]

\[ V \]

\[ W \]

\[ X \]

\[ Y \]

\[ Z \]

\[ a \]

\[ b \]

\[ c \]

\[ d \]

\[ e \]

\[ f \]

\[ g \]

\[ h \]

\[ i \]

\[ j \]

\[ k \]

\[ l \]

\[ m \]

\[ n \]

\[ o \]

\[ p \]

\[ q \]

\[ r \]

\[ s \]

\[ t \]

\[ u \]

\[ v \]

\[ w \]

\[ x \]

\[ y \]

\[ z \]

\[ 0 \]

\[ 1 \]

\[ 2 \]

\[ 3 \]

\[ 4 \]

\[ 5 \]

\[ 6 \]

\[ 7 \]

\[ 8 \]

\[ 9 \]

\[ A \]

\[ B \]

\[ C \]

\[ D \]

\[ E \]

\[ F \]

\[ G \]

\[ H \]

\[ I \]

\[ J \]

\[ K \]

\[ L \]

\[ M \]

\[ N \]
The fourth step is dedicated to the identification of most relevant interaction points of map E. This task is carried out using the MINIM utility included into the GRID program. Such a program collects all points within a certain energy threshold allowing the interpolation of the closest ones. The choice of an energy threshold value is a biased task per se but, considering a pharmacophore model as a minimum interaction descriptor built by few features, we have generally found appropriate an energy threshold in the range of 5–15% higher than the global minimum value. Actually, such a threshold allows collecting at least one feature for each used probe. Often the above energy threshold yields to complicated pharmacophore models that can be reduced using the GRID energy as cutting criterion.

In order to design a suitable model, all reported operations should be repeated using at least three different probes: the hydrophobic probe (DRY), an hydrogen bond acceptor (HBA) (O) and an hydrogen bond donor (HBD) (N1). This choice allows a basic characterization of most interaction areas, however, more sophisticated and selective models can be obtained by adding other GRID probes such as halogen or charged atoms, for example when those are part of the subunit α or of known interacting ligands.

In the fifth step the information originating from the different probe experiments are simply merged, by a trivial text editing, into a preliminary model stored in PDB file format (multiple features of Fig. 1).

The sixth step is dedicated to the validation of these preliminary informations and eventually their modulation in terms of number of points (features). The quality of the model is tested at least as the capability to recognize the original PDB ligand (α subunit). Technically the evaluation step can be carried out by the Catalyst software (Accelrys, 2003, http://www.accelrys.com) using the CTest fit module which calculates a non-energy weighted fit value. The preliminary model is imported converting the GBPM points into Catalyst features. The GRID energies are also included in the fit analysis as feature weight according to the following Equation (1):

$$w_{Fij} = \frac{EF_i}{AEF_j},$$

where $w_{Fij}$ is the weight for the feature $i$ into the hypothesis $j$, $EF_i$ is the GRID energy for the features $i$ and $AEF_j$ is the average GRID energy value computed along all probes of the hypothesis $j$. This approach allows to have a maximum fit value (MFV) equal to the total number of features available for the hypothesis $j$. Taking into account the GRID energies several preliminary models (hypotheses) can be designed modulating the number of features. Unfortunately due to the high variability, i.e. extension and interaction type, of the α–β subunit interface the number of preliminary models cannot be predefined. So, in order to identify the best one, all possible models are submitted to CTest fit.

A fit index (FI), defined as ratio between the CTest fit and MFV is used for the evaluation of each hypothesis and as choice criterion for the identification of the best GBPM.

Moreover, the FI descriptor, which makes possible the comparison among models with different number of features, can be used to extend the evaluation step including other molecules known to interact with the same β subunit binding site. Such an eventuality strongly improves the quality of the final model.

RESULTS

In order to check the versatility of the GBPM method, two different protein–protein complexes selected from PDB were extensively tested. In terms of complexity this kind of interaction is definitively more critical than other recognition phenomena such those of ligand–enzyme or drug–DNA complexes.

X-linked inhibitor of apoptosis

The protein–protein case study is related to the X-linked inhibitor of apoptosis (XIAP). Its third baculovirus IAP repeat domain (BIR3) recognizes molecules 1–5 shown in Figure 3.
The structure of peptide 1 has been determined by NMR and is deposited in 1G3F PDB entry (Liu et al., 2000). Protein 2 conformations were isolated from 1XB0 and 1XB1 models (Shin et al., 2005). Five conformations of molecule 3 were considered from the 1XB0 model. Another conformation of this peptide was extracted from the chimera 1TW6 structure (Vucic et al., 2005). From the same PDB model also molecule 4 was obtained. The conformation of the peptidomimetic 5 was obtained from the 1TFQ model (Oost et al., 2004). Finally the Smac protein 6, complexed with XIAP, was considered using the 1G73 PDB crystallographic model (resolution 2.00 Å). Since 6 is significantly larger than molecules 1–5 the recognition with XIAP is not exclusive of the BIR3 domain, but involves additional regions. A pharmacophore model able to describe entirely this kind of recognition is technically feasible but useless for the virtual screening of 3D databases, because it will eventually be appropriate to identify high molecular weight molecules with improbable drug-likeness properties. So the GBPM was derived using the 1G3F original PDB complex in which is reported the recognition of the relatively small synthetic peptide 1 and XIAP BIR3 domain. Moreover 1 is the largest among 1–5 ligands allowing a more exhaustive description of the interaction area.

The XIAP recognition area with caspase 9 is quite extended, about 700 Å² (Huang et al., 2001), several hydrophobic, electrostatic and hydrogen bond interactions are involved, so a simple receptor-based pharmacophore model results relatively hard to derive. Moreover the very low number of molecules, known to recognize the XIAP BIR3 domain, does not allow a rigorous classic ligand-based approach. For these reasons GBPM represents a potentially useful tool for the XIAP case study.

The computational work has been carried out following the flow chart reported in Figure 1. After the pretreatment step 1, the peptide 1 has been considered as subunit α and the XIAP as subunit β. The 1G3F PDB complex has been used to compute GRID molecular interaction field with O, N1 and DRY probes (maps A). These procedures, using the same complex box dimensions, have been repeated separately onto the α (maps B) and the β (maps C) subunits maintaining, in both cases, their original complex atom coordinates (Step 2). Maps A and B have been compared by GRAB algorithm obtaining the maps D that have been used, together with maps C, to obtain maps E (Step 3). The three maps E have been submitted to MINIM with interpolation option, using an energy threshold of 1 kcal/mol above the global minimum. This approach results into four features with one N1 probe, three with DRY and

![Fig. 2. GRAB operations (third step) applied to a generic protein-ligand PDB complex, respectively, related to subunit β (yellow CPK model) and subunit α (red polypeptide rendering). All maps A–E are displayed with blue solid rendering. The colour version of this figure is available as Supplementary data.](image-url)
The GRID probe N task has been carried out including in the GBPM flow chart (Fig. 1). For molecules 1–5, we introduced the positive ionizable feature, POS. This accounts for the presence of a positively charged N-terminus on molecules, adjusting the weight of the remaining ones. With the phophore model proceeded by iterative elimination of the less relevant ligand regions, despite the compounds share a common fragment. The explanation can be likely addressed to the Catalyst fit algorithm, that in presence of identical types of close chemical groups, prefers the best score overlap obtaining differences even among similar derivatives. The appropriate modulation of features, basing on GRID energy values, allows in the XIAP case study the identification of a GBPM able to recognize with high FIs all crystallographic ligands of the BIR3 domain. Notably this result has been obtained working with the 1G3F PDB model only.

### Interleukin 8 dimer

The GBPM evaluation in protein–protein case studies has been extended analyzing the interleukin 8 (IL8) dimer. This cytokine plays a relevant role in immune cells trafficking and in host defense against infection. It is known that IL8 can exist in both dimer and monomer forms, but this last only is able to productively interact with the CXCR receptors (Fernando et al., 2004). Then the equilibrium between dimer and monomer forms can be considered as an interesting target for modulating the IL8 activity. In the present study we designed a GBPM pharmacophore model for the IL8 dimer interface which could be useful for searching molecules modulating the equilibrium between active and inactive forms of IL8.

Our study has been carried out using the NMR-derived PDB model 1IL8 which represents an average structure of the IL8 homodimer in solution (Clore et al., 1990). To apply the GBPM approach, taking into account the general scheme reported in Figure 1, the IL8 chain A has been considered as subunit \( \beta \) while the chain B was the \( \alpha \) one (Fig. 5).

GRID probes O, N1 and DRY have been used in the 1IL8 case study. The preliminary pharmacophore model (Hyp1-1IL8) for the IL8 dimer interface has been designed selecting for each probe the chemically closest feature as follows: the HBD for N1, the generic hydrophobic (HPB) for DRY and the HBA for O. The weight of each features has been scaled taking into account the GRID interaction energies using Equation (1).

### Ligand and PDB data

<table>
<thead>
<tr>
<th>#</th>
<th>R</th>
<th>Conformations</th>
<th>entry ID</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-Ile-Ala-Gly-Lys-Ser-Glu</td>
<td>1</td>
<td>1G3F</td>
<td>NMR</td>
</tr>
<tr>
<td>2</td>
<td>-Ile-Ala</td>
<td>6</td>
<td>1XB1</td>
<td>X-RAY (2.70 Å)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>1XB0</td>
<td>X-RAY (2.20 Å)</td>
</tr>
<tr>
<td>3</td>
<td>-Ile</td>
<td>5</td>
<td>1TW6</td>
<td>X-RAY (1.71 Å)</td>
</tr>
<tr>
<td>4</td>
<td>-Ile-Ala-Gln</td>
<td>1</td>
<td>1TFQ</td>
<td>NMR</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 3. Chemical structure of ligands 1–5 interacting with BIR3 domain of XIAP and their PDB record, the number of conformation, the source and the resolution of the crystallographic models.

only one with O. The preliminary model (Hyp1-1G3F) has been imported into the Catalyst program using for each probe the chemically closest feature as follows: the HBD for N1, the generic hydrophobic (HPB) for DRY and the HBA for O. The resulting model has been tested, with both rigid and flexible CICtest algorithm, using ligand 1. Since the resulting FI index, equal to 0.02, revealed a poor recognition of peptide 1 Hyp1-1G3F was simplified by removing the less relevant HBD HDB3 and HDB4 features and rescaling the weight of the remaining ones as reported for Hyp1-1G3F. The evaluation of the new seven-features model, Hyp2-1G3F, indicated a better fit with 1 remaining ones as reported for Hyp1-1G3F. The evaluation of a new features model for screening, but useful for characterizing the complex interactions. We have included Hyp1-1IL8 into the discussion of the IL8 case study because no ligands are known for their interaction with the IL8 dimer interface, then we have considered the recognition of the first model to evaluate the applicability of the GBPM onto this protein–protein case study. As reported in Figure 6, our approach has been able to recognize the target. Notably several residues, indicated in Figure 6 and reported by Clore into the IL8 paper for their relevance into the IL8 dimerisation, have the location of this point was coincident with the HDB1 one, which has showed an interaction energy equal to \(-7.31\) kcal/mol. Consequently we built a new model, Hyp3-1G3F, substituting HDB1 with POS rescaling the feature weights, with superior FI value equal to 0.95.

Hypt3-1G3F then has been evaluated with respect to 2–5. With all molecules, taking into account all their experimental determined conformations, FI values higher than 0.90 have been reached confirming the high degree of recognition of known molecules interacting with XIAP BIR3 domain of Hyp3-1G3F (Fig. 4). In some cases, especially with 2, equivalent features overlap different ligand regions, despite the compounds share a common fragment. The explanation can be likely addressed to the Catalyst fit algorithm, that in presence of identical types of close chemical groups, prefers the best score overlap obtaining differences even among similar derivatives. The appropriate modulation of features, basing on GRID energy values, allows in the XIAP case study the identification of a GBPM able to recognize with high FIs all crystallographic ligands of the BIR3 domain. Notably this result has been obtained working with the 1G3F PDB model only.
been identified. Moreover GBPM identified other aminoacids located to the dimer interface that could contribute to the equilibrium between the active and inactive forms of the IL8. These observations have allowed us to positively consider the application of GBPM to the IL8 case study. In order to design a more suitable pharmacophore model we have reduced the total number of features of Hyp1-1IL8 removing the points with higher interaction energy. The best of the models obtained by Hyp1-1IL8 simplification, indicated as Hyp2-1IL8, showed an FI equal to 0.80 and a good recognition of the dimerization region.

In Figure 7 the IL8 dimerization interface recognition of Hyp2-1IL8 and its feature composition has been reported. Even if no ligand validation has been possible for the present case study, we considered IL8 as a good application for the GBPM. Since no pharmacophoric information is available for the dimer interface, our method, validated by the recognition of the original ligand, represents a potentially useful model for virtual screening purposes.

**CONCLUSIONS**

The GBPM have been developed with the aim to define a general computational protocol for characterizing interaction between any kind of biomolecules creating pharmacophore models starting from well referenced experimental complex models such as those deposited in the PDB. It is different from traditional pharmacophore model because no information about the activity is explicitly included in its development. The assumption is that if a consistent recognition between two or more biomolecules can occur and be determined by spectroscopic accurate methods, typically as in co-crystal models of the PDB, their high degree of reciprocal affinity

Fig. 4. Hyp3-1G3F model recognition of ligands 1–5. Feature weights are reported in parenthesis. The colour version of this figure is available as Supplementary data.

Fig. 5. 1IL8 PDB structure. Chain A is represented in green cartoons, chain B is reported in stick CPK notation. The colour version of this figure is available as Supplementary data.
should be summarized by relevant features. The aim of the GBPM approach is to convert the affinity information of experimentally known complexes into a new kind of pharmacophore model, implementing an energy-based criterion. The preliminary validation of the methodology in two diverse kind of protein–protein complexes is going to be extended further also to other examples, however, the results obtained so far already indicate the high versatility of the GBPM approach.

The examples described in the present work revealed good capabilities of our method to identify the most relevant host-guest occurring interaction into the analyzed complexes independently from their chemical nature. It is worth noting that when particular physico-chemical features are found in the ligands, as in the former case study with charged moieties, appropriate probes can dramatically influence the results and therefore must be incorporated into the analysis. GBPM has revealed no impediments for its application on macromolecular targets. Due to the large number of GRID probes available, a notable improvement of the generated pharmacophore models can be achieved suggesting also new substituents for known ligands. Consequently, the GBPM approach could represent a useful bioinformatic tool for the recognition analysis of biomolecules. Future work will focus on possible applications of the GBPM method in the field of the drug discovery process.

ACKNOWLEDGEMENTS

This work was supported by the ‘P.O.P.’ of Regione Calabria and the computational resources of ‘CNR-ISN Sezione di Farmacologia’, Catanzaro, Italy.

Conflict of Interest: none declared.

REFERENCES

Accelrys Inc. (2003), San Diego, CA, Catalyst ver. 4.9.
Fig. 7. Hyp2-1IL8 chain B recognition. Feature weights are reported in parenthesis, hydrogen atoms have been hidden for clarity.