March 12, 1998

G R A M M v1.03

Global Range Molecular Matching

Ilya A. Vakser
Department of Cell and Molecular Pharmacology
Medical University of South Carolina
171 Ashley Avenue, Charleston, SC 29425
Phone: 803-792-2471; Fax: 803-792-2475
Email: vakseri@musc.edu; http://reco3.musc.edu

Introduction

GRAMM is a program for protein docking. To predict the structure of a complex, it requires only the atomic coordinates of the two molecules (no information about the binding sites is needed). The program performs an exhaustive 6-dimensional search through the relative translations and rotations of the molecules. The molecular pairs may be: two proteins, a protein and a smaller compound, two transmembrane (TM) helices, etc. GRAMM may be used for high-resolution molecules, for inaccurate structures (where only the gross structural features are known), in cases of large conformational changes, etc.

The Global Range Molecular Matching (GRAMM) methodology is an empirical approach to smoothing the intermolecular energy function by changing the range of the atom-atom potentials. The technique allows to locate the area of the global minimum of intermolecular energy for structures of different accuracy. The quality of the prediction depends on the accuracy of the structures. Thus, the docking of high-resolution structures with small conformational changes yields an accurate prediction, while the docking of ultralow-resolution structures will give only the gross features of the complex.

The following text is not a manual, but rather a short technical reference. For the description of the algorithm, its implementation, discussion of applicability, and all other details, see the papers listed at the end of this document.

I am making GRAMM publicly available following a number of requests from different labs. I would like to make it clear, however, that both the methodology and the program, at present, are in the process of active development and validation, especially in the area

of the low-resolution docking, and have to be viewed like that. The program is free. However, I would expect proper references. I will also appreciate bug reports.

GRAMM has evolved during my stay in the Weizmann Institute (1991-1993), Washington University (1993-1995), Rockefeller University (1995-1997), and, from 1997, at the Medical University of South Carolina. I deeply appreciate the assistance of my colleagues in all these institutions (especially, Ephraim Katchalski-Katzir, Garland Marshall, and Andrej Sali).

How to work with GRAMM

Using GRAMM at *HIGH RESOLUTION* is pretty straightforward. You will get a list of high-score (low-energy) ligand positions, which you may take as is, or refine by other techniques. Since GRAMM does not use a statistical sampling, but rather performs an *exhaustive search*, you will get *all* configurations of the complex with the high-score steric fit (within the accuracy of the search step and the molecules' representation). Even if you have high-resolution structures, I would recommend, in addition, to run docking at low resolution, to determine the potential areas of the global minimum.

Using GRAMM at *LOW RESOLUTION*. Prediction of complexes of ultralow-resolution structures, with large conformational changes ... sounds attractive? Please, be reasonable and remember: there is no magic in the world (unfortunately). You can NOT get an accurate complex of two largely inaccurate proteins (at least, presently). The docking results at the lowest resolution (e.g., ~7Å, for proteins, and ~4Å, for helices) may give you only the general PREFERENCES (often nonspecific) in the complex formation (see Refs.), rather than the 'real' coordinates. Let's say, a distribution of low-energy ligand positions in the proximity of the binding site of the protein. Or a 90° two-dimensional sector where a TM helix is likely to make a complex with any other helix (due to the low-resolution preferences in helix packing, see Vakser, 1996a).

1. Docking

File `rpar.gr` (parameters)

Sets the parameters of the docking procedure. The value of a parameter has to appear after the equality sign.

`mmode` Specifies the docking mode (generic or helix). In the generic mode, GRAMM tries all ligand's positions and orientations. In the helix mode, to save the computational time and to simplify the analysis of the results, GRAMM automatically discards configurations with large displacements along the helix axes and angles between helices larger than indicated in `rmol.gr` file (see below). There are no other differences, so if you want to try *all* interhelical configurations, run GRAMM in the generic mode.

eta Step of the grid (Katchalski-Katzir *et al.*, 1992; Vakser, 1995,1996b); also the range of the atom-atom potential, in case of the 'gray' projection (Vakser, 1996a).

ro Repulsion part of the potential, in arbitrary units (Vakser, 1996a).

fr Attraction double range, mostly as an option for high-resolution docking (Katchalski-Katzir *et al.*, 1992; Vakser & Aflalo, 1994).

crang Projection of an atom, as a sphere with the van der Waals radius (for high resolution docking) or the grid-step radius (for low resolution docking).

ccti 'yes-no' (blackwhite) or cumulative (gray) projection (Vakser, 1995b,1996b).

crep Switch to the hydrophobic docking (Vakser & Aflalo, 1994).

maxm Number of matches to output.

ai Step for the systematic search through the rotational coordinates.

In the following examples, I give the suggested values for the parameters. They still may not be optimal, so you may try to experiment with them.

High-resolution generic docking

The high-resolution docking is designed for accurate complex predictions, in case of small structural inaccuracies.

Example 1. Geometric docking I (Vakser, 1996a).

```
Matching mode (generic/helix) ..................... mmode= generic
Grid step ........................................... eta= 1.7
Repulsion (attraction is always -1) ................... ro= 30.
Attraction double range (fraction of single range) ..... fr= 0.
Potential range type (atom_radius, grid_step) ....... crang= atom_radius
Projection (blackwhite, gray) ..............  ....... ccti= gray
Representation (all, hydrophobic) ..................... crep= all
Number of matches to output ........................... maxm= 1000
Angle for rotations, deg (10,12,15,18,20,30, 0-no rot.)  ai= 10
```

Example 2. Geometric docking II (Katchalski-Katzir *et al.*, 1992).

```
Matching mode (generic/helix) ..................... mmode= generic
Grid step ........................................... eta= 1.7
Repulsion (attraction is always -1) ................... ro= 10.
```

```
Attraction double range (fraction of single range) ..... fr= 0.5
Potential range type (atom_radius, grid_step) ....... crang= atom_radius
Projection (blackwhite, gray) ..............  ....... ccti= blackwhite
Representation (all, hydrophobic) ..................... crep= all
Number of matches to output ........................... maxm= 1000
Angle for rotations, deg (10,12,15,18,20,30, 0-no rot.)  ai= 10
```

Example 3. Hydrophobic docking (Vakser & Aflalo, 1994)

```
Matching mode (generic/helix) ....................... mmode= generic
Grid step ............................................ eta= 1.7
Repulsion (attraction is always -1) ................... ro= 5.
Attraction double range (fraction of single range) ..... fr= 0.
Potential range type (atom_radius, grid_step) ....... crang= atom_radius
Projection (blackwhite, gray) ..............  ....... ccti= blackwhite
Representation (all, hydrophobic) ..................... crep= hydrophobic
Number of matches to output ........................... maxm= 1000
Angle for rotations, deg (10,12,15,18,20,30, 0-no rot.)  ai= 10
```

High-resolution helix packing

Use Examples 1,2 with `mmode= helix` (or `mmode= generic` if you don't want to limit the search). For TM helices, the hydrophobic docking (Example 3) doesn't make sense, because of the hydrophobic environment, although it may be applicable for helices in soluble structures.

Low-resolution generic docking

The low-resolution docking is designed for the prediction of the gross features of a complex, in the case of major structural inaccuracies. It may also be used, in the case of accurate structures, to overcome the multiminima problem (Vakser, 1996a). The following values are suggested for globular proteins and their ligands (the ligand has to be larger than ~50 atoms). For a detailed discussion, see Vakser, 1995b,1996a,1996b.

Example 4.

```
Matching mode (generic/helix) ....................... mmode= generic
Grid step ............................................ eta= 6.8
Repulsion (attraction is always -1) ................... ro= 6.5
Attraction double range (fraction of single range) ..... fr= 0.
Potential range type (atom_radius, grid_step) ....... crang= grid_step
Projection (blackwhite, gray) ..............  ....... ccti= gray
Representation (all, hydrophobic) ..................... crep= all
Number of matches to output ........................... maxm= 1000
Angle for rotations, deg (10,12,15,18,20,30, 0-no rot.)  ai= 20
```

Low-resolution helix packing

The following values are suggested for inaccurate (e.g., modeled) TM helices. They may be useful for the investigation of ultralow-resolution (often nonspecific) PREFERENCES in helix packing (rather than 'real' coordinates). Keep in mind that the procedure is sensitive to a digitization on a sparse grid. Thus, for example, the grid images of polyalanine α-helices are not homogeneous, which results in non-circular distribution of the low-energy predictions. However, STATISTICALLY, the procedure reliably distinguishes between the interface and the non-interface areas of the helices. The subject is briefly described in (Vakser, 1996a), although a detailed paper is still in preparation.

Example 5.

```
Matching mode (generic/helix) ..................... mmode= helix
Grid step ........................................... eta= 4.1
Repulsion (attraction is always -1) ................... ro= 11.
Attraction double range (fraction of single range) ..... fr= 0.
Potential range type (atom_radius, grid_step) ....... crang= grid_step
Projection (blackwhite, gray) ..............  ....... ccti= gray
Representation (all, hydrophobic) ..................... crep= all
Number of matches to output ........................... maxm= 1000
Angle for rotations, deg (10,12,15,18,20,30, 0-no rot.)  ai= 20
```

File *rmol.gr* (molecules description)

Empty lines and lines which start with # are ignored. The first 2 lines in the example below tell you how to organize your data. You may input multiple molecular pairs (a line per pair). The first molecule will be considered as 'receptor' and the second as 'ligand'. The data has free format, with space separation.

Filename	File with molecule's coordinates (PDB format).
Fragment	* - full molecule X - chain id (case sensitive) xxxx-xxxx - atom numbers (first-last)
ID	String of characters (no spaces in-between) to identify your molecules. These ID's will be used by GRAMM to name the output file(s).
parallel / antiparallel	Helix mode only. Specifies the N term. - C term. direction in the helix pair.
max. angle	Helix mode only. Sets the limit for the angle (in degrees) between the main axes. If you make it larger than 180, all angles will be tried, regardless of the 'parallel/antiparallel' parameter.

Example 1. A and B subunits of hemoglobin; trypsin from the complex with BPTI and uncomplexed BPTI

```
# Filename  Fragment  ID      Filename  Fragment  ID     [paral/anti  max.ang]
# ------------------------------------------------------------------------

pdb2hhb.ent 1-1069  2hhba   pdb2hhb.ent 1114-2256 2hhbb

pdb2ptc.ent    E    2ptce   pdb4pti.ent     *     4pti
```

Example 2. Helices 2-3 and 3-4 of bacteriorhodopsin

```
# Filename  Fragment  ID      Filename  Fragment  ID     [paral/anti  max.ang]
# ------------------------------------------------------------------------

pdb1brd.ent  281-478  1brd2 pdb1brd.ent  554-785  1brd3   antipar       50

pdb1brd.ent  554-785  1brd3 pdb1brd.ent  822-961  1brd4   antipar       50
```

Run GRAMM with the parameter `scan` (`gramm scan`). It creates `.log` file and `.res` (results) file. Do not modify the `.res` file - it has to be in the exact format for the building of PDB structures of the predicted complexes.

Comment 1. There is certain asymmetry in the representation of 'receptor' and 'ligand' molecules in GRAMM. The X-Y and Y-X docking will give statistically similar distributions of low-energy configurations, although the absolute values of the energy may be different.

Comment 2. GRAMM determines the size of the grid (16, 32, or 64) automatically, based on the grid step, size of the molecules and the nature of the docking problem. The switch to larger grids corresponds to a substantial increase in CPU time (see Performance section). To be aware which grid has been chosen, see the output in `gramm.log` file.

2. PDB files of the predicted complexes

File ***`wlist.gr`*** (list of results)

The general format is similar to `rmol.gr`. You may specify multiple lines (one line per a results file). 'File_of_predictions' is the output (results) file of GRAMM. 'separate/joint' tells GRAMM whether to build individual PDB files for each match or to join them in one file (one receptor and multiple ligand coordinates - recommended for better visualization). If you have multiple results files and choose 'separate', be aware of the 'combinatorial explosion' (e.g., 10 results files, each with 10 matches, will give you 100

PDB files). Read the REMARK section of the resulting file(s) for the chain assignment. '+init_lig' sets an option to include the initial (before docking) coordinates of the ligand into the resulting PDB file (used basically for the method validation purposes, in case of known configurations of the complex).

Example. Joint file of predictions 1-10, with the X-ray position of the ligand (helices 2-3 of bacteriorhodopsin); separate files of predictions 3-7, without the X-ray position of the ligand (A and B subunits of hemoglobin)

```
# File_of_predictions   First_match   Last_match   separate/joint  +init_lig
# ------------------------------------------------------------------------

  1brd2-1brd3.res           1             10           joint           yes

  2hhba-2hhbb.res           3              7           separ           no
```

Run GRAMM with the parameter `coord` (`gramm coord`).

Comment. Starting with match 11, ligands will be identified with two-character chain ID (match 10 will be chain 0 to maximize the number of one-character ID's). This may cause problems for programs (e.g., graphical) which will read this file.

3. Docking AND building coordinate files of the predicted complexes

You may join both operations in one run (`gramm scan coord`). Make sure that proper results files are set in `wlist.gr` (filenames are made of ID strings in `rmol.gr`).

Platforms

Presently, GRAMM is compiled on SGI R4000, SGI R4400, SGI R8000, SGI R10000, Sun SPARC, IBM RS6000, and DECAlpha Unix workstations, as well as on a PC platform under Windows95. In the near future I will expand this list, so check the GRAMM site for the updates.

Performance

The CPU time depends on the grid step (in the case when GRAMM automatically switches between 16, 32 and 64 grids), angle interval, and matching mode (generic or helix). It may range from ~10 sec on 195 MHZ 10000 SGI for the low-resolution docking of a helix pair, in the 'helix' docking mode, with the angle interval of 20°, to several days, in the case of high-resolution docking of globular proteins, with a small angle.

Basic papers on GRAMM methodology

- E. Katchalski-Katzir, I. Shariv, M. Eisenstein, A. A. Friesem, C. Aflalo, I. A. Vakser, 1992, Molecular surface recognition: Determination of geometric fit between proteins and their ligands by correlation techniques, *Proc. Natl. Acad. Sci. USA*, **89**, 2195-2199. Basic algorithm of protein recognition by correlation technique with Fast Fourier transform. High-resolution 'geometric' docking.

- I. A. Vakser, C. Aflalo, 1994, Hydrophobic docking: A proposed enhancement to molecular recognition techniques, *Proteins*, **20**, 320-329. High-resolution 'hydrophobic' docking.

- I. A. Vakser, G. V. Nikiforovich, 1995a, Protein docking in the absence of detailed molecular structures, in: *Methods in Protein Structure Analysis* (M. Z. Atassi & E. Appella, eds.), Plenum Press, New York, pp. 505-514.

- I. A. Vakser, 1995b, Protein docking for low-resolution structures, *Protein Eng.*, **8**, 371- 377. 'Low-resolution' protein docking.

- I. A. Vakser, 1996a, Long-distance potentials: An approach to the multiple-minima problem in ligand-receptor interaction, *Protein Eng.*, **9**, 37-41. Interpretation of the low-resolution docking in terms of energy potentials.

- I. A. Vakser, 1996b, Low-resolution docking: Prediction of complexes for underdetermined structures, *Biopolymers*, **39**, 455-464. Validation of the low-resolution docking.

- I. A. Vakser, 1996c, Main-chain complementarity in protein-protein recognition, *Protein Eng.*, **9**, 741-744. Docking of C^{α} structures.