

# Manual for VLG

## Virtual Library Generator

### Version 1.0

Fabian Dey, Nicolas Majeux and Amedeo Caflisch

To improve this documentation, please send comments and feedback to:

Amedeo Caflisch

E-mail: [caflisch@bioc.unizh.ch](mailto:caflisch@bioc.unizh.ch)

FAX: (+41 44) 635 68 62

November 11, 2008

# Contents

<b>1</b>	<b>Getting started</b>	<b>1</b>
1.1	Concept . . . . .	1
1.2	Files required for running VLG . . . . .	1
1.3	Running VLG . . . . .	2
1.4	Output of VLG . . . . .	2
<b>2</b>	<b>Method</b>	<b>3</b>
2.1	Removal of redundant substituent-vectors . . . . .	3
2.2	Substituent Attachment . . . . .	4
2.3	Energy Calculation . . . . .	4
2.4	Sampling Algorithm . . . . .	5
<b>3</b>	<b>Input File</b>	<b>7</b>
3.1	Input File: Example Using Receptor . . . . .	7
3.2	Description . . . . .	9
3.2.1	Splitting up calculations . . . . .	9
3.2.2	Parameter Filename . . . . .	10
3.2.3	Output Folder . . . . .	10
3.2.4	Output Mode . . . . .	10
3.2.5	Random Seed Number . . . . .	10
3.2.6	Receptor Specifications . . . . .	10
3.2.7	Grid settings . . . . .	11
3.2.8	Symmetry Evaluation . . . . .	11
3.2.9	Sampling Algorithm . . . . .	11
3.2.10	Cutoff . . . . .	11
3.2.11	Scaffold Specifications . . . . .	11
3.2.12	Substituent Specifications . . . . .	11
3.3	Input File Example Without Receptor . . . . .	12
<b>4</b>	<b>Parameter File</b>	<b>13</b>
4.1	Parameter File Structure . . . . .	13
4.2	Description Parameter File . . . . .	14
4.2.1	Atom Section . . . . .	14
4.2.2	Bond Section . . . . .	14
4.2.3	Forbidden Connections . . . . .	14
<b>5</b>	<b>Troubleshooting</b>	<b>15</b>
5.1	What to do if ... . . . . .	15
5.2	Error Messages . . . . .	15
<b>6</b>	<b>Acknowledgments</b>	<b>18</b>

# 1 Getting started

## 1.1 Concept

VLG is a program that adds substituents to a given scaffold (fragment or molecule) which is docked inside a protein binding site (referred to as receptor) or in stand-alone mode without any protein. VLG will decorate scaffolds with user-defined substituent libraries in an exhaustive, combinatorial manner. The list of substituents can be defined individually for every scaffold substitution point. Heavy atom to hydrogen atom vectors serve as the attachment points on both the substituents and the scaffold. If specified by the user, VLG will rotate the added substituents around the attachment axis, calculate van der Waals and coulombic energies within the ligand and, if needed, the interaction energy between ligand and receptor. The conformation or pose with the lowest energy is saved to disk.

For lists of substituents and scaffolds commonly occurring in virtual libraries and drug molecules consult [1] and [2].

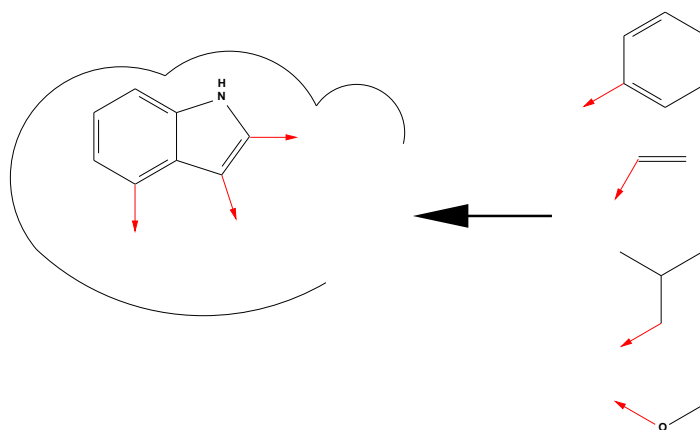


Figure 1: Concept

## 1.2 Files required for running VLG

- *VLG-executable*  
VLG is statically compiled in 32-bit mode using the GNU-compiler suite. It should thus run on any up-to-date Linux operating system.
- *Input file*  
Contains all relevant data which are specific for a certain run. For a detailed description see section 3.
- *Parameter file*  
Contains force-field related information as well as a list of forbidden connections. For a detailed description see section 4.

- *Receptor file*  
The receptor file needs to be in Sybyl MOL2 format containing all atoms (including both polar and non-polar hydrogens) with CHARMM atom types [3] and partial charges assigned.
- *Scaffold file/s*  
Preparation identical to the receptor file.
- *Substituent library/libraries*  
Preparation identical to the receptor file.

### 1.3 Running VLG

VLG is run on a Linux operating system in a shell by typing:

```
vlg [input file] &> [output file]
```

Consult the troubleshooting section (5) if problems arise while trying to run VLG.

### 1.4 Output of VLG

VLG prints information about the run to standard output. In order to save this output, run VLG as described in 1.3. The verbosity of the output can be controlled through a parameter in the input file (section 3.2.4). Apart from this output, VLG stores the decorated scaffolds in a user defined folder (section 3.2.3). The decorated scaffold MOL2 files are named according to the following pattern:

$$[\text{scaffold-name}]_{-}\left\{ [\text{scaffold-vector}]_{-}[\text{substituent}]_{-}[\text{substituent-vector}] \right\}_N .\text{mol2}$$

Where  $N$  is the number of attached substituents per scaffold. Apart from [scaffold-name], the above mentioned values are integers which originate from the appearance of a certain vector or substituent in a list of vectors or substituents, respectively (see section 3.2.11).

## 2 Method

### 2.1 Removal of redundant substituent-vectors

If specified in the input file (section 3.2.8), VLG will try to remove redundant vectors from each individual substituent. Two vectors are non-redundant if the number and atom types of atom neighbors of the origins of the two vectors are not equal. If the aforementioned criterion is met, VLG makes a copy of the substituent and superimposes the hydrogen, heavy and a neighboring atom of the latter of the first vector of the original and the second vector of the copied substituent with one another using the method described by Kabsch [4].

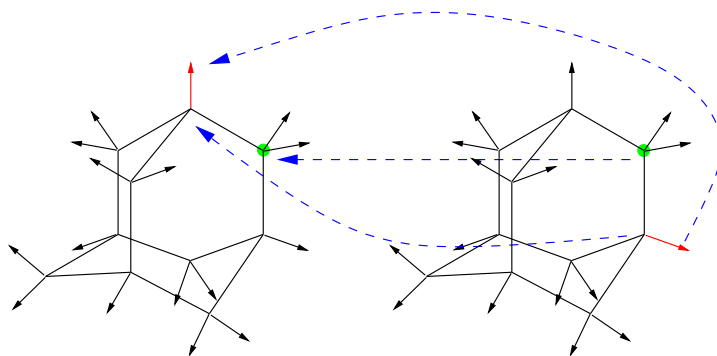


Figure 2: Superpositioning example adamantane: vectors under investigation are red, neighboring atoms are green

The similarity  $S(O, C)$  between the original ( $O$ ) and the superimposed ( $C$ ) substituent is calculated with the following formula (see Ref. [5] and [6]):

$$S(O, C) = \frac{S_{OC}}{\max(S_{OO}, S_{CC})} \quad (1)$$

$$\text{where } S_{XY} = \sum_{i \in X} \sum_{j \in Y} w_{t_i t_j} \exp(-\gamma r_{ij}^2) \quad (2)$$

$$\text{and } S(O, C) \in [0, 1] \quad (3)$$

$r_{ij}$  is the distance in Å between two atoms  $i$  and  $j$ ,  $w_{t_i t_j}$  is an atom element unit matrix and  $\gamma$  is a coefficient which acts on the broadness of the distribution. When comparing two positions of identical conformations of a molecule formula 1 is further reduced to:

$$S(O, C) = \frac{S_{OC}}{S_{OO}} \quad (4)$$

$$\text{since } S_{OO} \equiv S_{CC} \quad (5)$$

If the evaluated similarity is above the given cutoff, one of the two substituent-vectors is redundant and is removed. Due to the geometric approach with which redundancy is evaluated, VLG discriminates between stereochemically distinct vectors (e.g.  $R$ - and  $S$ -vectors). Therefore, attachment vectors of substituents with distorted conformation (e.g. when not properly minimized) might not be recognized as redundant.

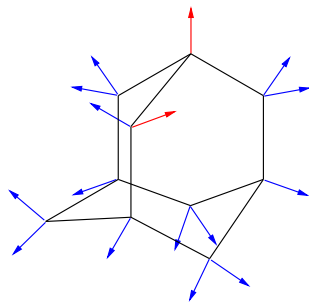


Figure 3: Adamantane after removal of redundant vectors (blue), unique vectors in red

## 2.2 Substituent Attachment

VLG attaches substituents exhaustively to the scaffold. It proceeds in such a manner that for a given set of scaffold vectors all allowed substituent connections are attached in all possible combinations applying the potential restrictions defined by the user (see 3.2.11 and 4.2.3). The user is given the possibility to specify which substituent should be added to a specific scaffold attachment vector. Heavy atom – hydrogen atom vectors serve as possible attachment points on both the scaffold and the substituents. Out of all the possible vectors, the actual attachment vectors to be considered during the VLG-run are specified by the user in the specific input file sections (see 3.2.11 and 3.2.12). The bond length of the newly formed bond is taken from the parameter file. At the moment, neither atom types nor partial charges are updated for the scaffold or the substituent upon merging the two.

## 2.3 Energy Calculation

If requested, VLG computes both the inter, as well as the intra van der Waals and coulombic energies of the generated molecule.

$$E = E_{\text{vdW}}^{\text{intra}} + E_{\text{elect}}^{\text{intra}} + E_{\text{vdW}}^{\text{inter}} + E_{\text{elect}}^{\text{inter}} \quad (6)$$

Internal energies are computed between all scaffold–substituent and substituent–substituent combinations. Interactions of bonded atoms (1–2) and atoms separated by two bonds (1–3) are excluded from the evaluation. The van der Waals energy is calculated according to a 6–12 Lennard–Jones model:

$$E_{\text{vdW}} = \sum_{i < j} \sqrt{\varepsilon_i \varepsilon_j} \left\{ \left( \frac{R_i^{\text{vdW}} + R_j^{\text{vdW}}}{r_{ij}} \right)^{12} - 2 \left( \frac{R_i^{\text{vdW}} + R_j^{\text{vdW}}}{r_{ij}} \right)^6 \right\} \quad (7)$$

Where  $R_i^{\text{vdW}}$  is the van der Waals radius,  $\varepsilon_i$  is the minimum of the van der Waals potential between two atoms of type  $i$  at optimal distance of  $2 \cdot R_i^{\text{vdW}}$ , and  $r_{ij}$  is the distance between atoms  $i$  and  $j$  in Å.

The coulombic energy is calculated with a distance-dependent dielectric model:

$$E_{\text{elect}} = 332 \sum_{i < j} \frac{q_i q_j}{\epsilon r_{ij}^2} \quad (8)$$

Where  $q_i$  and  $q_j$  are the partial charges in electronic units of atoms  $i$  and  $j$ , respectively,  $r_{ij}$  is the interatomic distance in Å and  $\epsilon$  is the dielectric constant.

Additionally van der Waals and coulombic energies between the entire molecule and the receptor are calculated when using a receptor during the VLG-run. In order to save time in the calculation, the potential of the receptor is stored on a grid and has thus to be calculated only once at the beginning of the run. The value of the potential is derived from the eight points of the grid surrounding  $\vec{r}_i$  by the trilinear interpolation method [7]. The energies are calculated according to equations 7 and 8.

The grids are saved in binary format using numbers of 4-bit size. The first three integers specify the number of grid points in the x-,y- and z-dimension. Following are the float values of the potential, in the van der Waals-grid these include both the attractive and the repulsive part of the potential whereas the coulombic grid only contains one value per grid point.

## 2.4 Sampling Algorithm

If specified in the input file, VLG modifies the dihedral angle of the added bond and performs rigid body translation and rotation around the geometric center of the molecule all by random amounts using a combination of two modified Monte Carlo sampling procedures, namely stochastic tunneling [8] and simulated annealing [9]. The substituents as well as the scaffold are treated as rigid bodies during conformational sampling. The main difference of stochastic tunneling compared to a classic Monte Carlo approach lies in the transformation of the energy with:

$$f_{stun}(E) = 1 - \exp(-\gamma(E-E_0)) \quad (9)$$

Where  $\gamma$  is a constant coefficient,  $E$  is the energy at the current step and  $E_0$  is the lowest energy that has been encountered during the optimization run. The transformation of the energy leads to a flattening of the potential energy surface with the proceeding of the sampling run while preserving the original location of all minima [8]. The probability  $p$  of accepting a step  $x_1 \rightarrow x_2$  is defined by the Metropolis criterion [10] using the transformed energy:

$$\begin{aligned} \text{with } \Delta E_{stun} &= f_{stun}(E_2) - f_{stun}(E_1) \\ p &= \begin{cases} 1 & \text{if } \Delta E_{stun} \leq 0 \\ \exp(-\frac{\Delta E_{stun}}{T}) & \text{if } \Delta E_{stun} > 0 \end{cases} \end{aligned} \quad (10)$$

The probability of accepting a move thus depends on the energy difference  $\Delta E_{stun}$  and the temperature  $T$  which is given in  $k_b T$  units. The idea of simulated annealing is to explore local minima by gradually decreasing the temperature over the optimization run. According to the metropolis criterion (equation 10), the lowered temperature leads to fewer unfavorable steps ( $\Delta E > 0$ ) being accepted and thus retaining the molecule within local minima. VLG uses a geometric cooling procedure to calculate the temperature  $T_n$  at step  $n$  of the optimization procedure:

$$T_n = \alpha * T_{n-1} \quad (11)$$

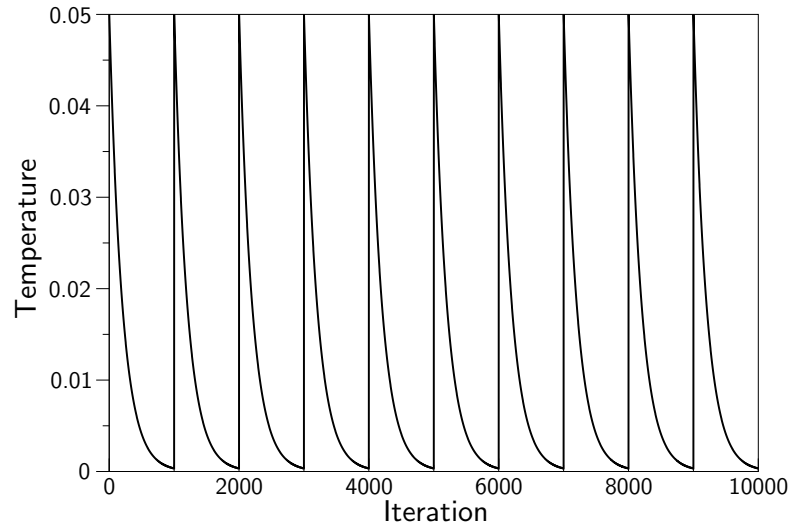


Figure 4: Simulated annealing with restart every 1000 steps:  $\alpha = 0.995$

where  $\alpha$  is a constant, user-defined cooling factor. In order to sample multiple minima, the approach implemented in VLG allows multiple restarts of the cooling cycle in one optimization run (Figure 4).



## 3 Input File

### 3.1 Input File: Example Using Receptor

3.2.1	# number of hosts / host id 10 1
3.2.2	./parameter.vlg
3.2.3	./outputs/
3.2.4	# "silent" (0) - (3) "chatty" 1
3.2.5	# seed number 2345
3.2.6	# use receptor (y/n) / # receptor coordinates (in mol2 format) filename y ./protein.mol2
3.2.7	# number of residues / # binding site residue list 14 34 35 42 55 57 74 78 87 167 103 105 106 109 157 # van der Waals energy: # grid margin / grid spacing / # write (w) or read (r) van der Waals grid / grid filename # 10.0 0.3 w ./Grids/vdwaals.grid

Table 1: VLG input file part 1

3.2.7	# Coulombic energy: # dielectric constant / # distance dependent dielectric (1=yes , 0=no) / # grid margin / grid spacing / # write (w) or read (r) Coulombic grid / grid filename # 2.0 1 10.0 0.5 w ./Grids/Coulombic.grid # # #
3.2.8	# check symmetry of substituents / symmetry cutoff / # exponential factor gamma y 0.95 0.9 # #
3.2.9	# Monte Carlo Variables / # temperature (in kbT units) / cooling factor / cooling restart / # maximal substituent rotation (rad) / # maximal rigid body translation / # maximal rigid body rotation (rad) / # number of iterations / gamma # y 0.05 0.995 1000 0.1 0.1 0.2 10000 0.0003 # #
3.2.10	# Use cutoff (y/n) / cutoff y 1000.0 # # #
3.2.11	# number of scaffolds / # scaffold filename / # restrictions (y/n) / min. / max. number substituents / # number of connection vectors / # list of connection vectors / # substituent set IDs / # number of mandatory connection vectors / # list of mandatory connection vectors # 1 ./indole.mol2 y 2 3 3 6 14 8 15 9 16 1     1     2 1 2 # #

Table 2: VLG input file part 2

### 3.2.12

```
# number of substituents /
# substituent filename /
# substituent set ID /
# number of connection vectors /
# list of connection vectors
#
5
benzene.mol2
1
1
1 7
toluene.mol2
1
1
2 1 2 3 2 4
aniline.mol2
1
1
5 6
ethanol.mol2
2
3
3 7 3 8 3 9
pyridine.mol2
2
1
3 8
#
#
# For lists of commonly occurring fragments see:
# Vieth et al. J. Med. Chem. 2004, 47, 224-232
# Ertl et al. J. Chem. Inf. Comput. Sci. 2003, 43, 374-380
#
```

Table 3: VLG input file part 3

## 3.2 Description

Some helpful hints on the build-up of the input file:

- Lines starting with “#” are comment lines and are ignored.
- The order of the information containing sections is fixed and must not be altered.
- Regions encoding information must not contain any unnecessary information like additional comments. Moreover, if a specific feature such as the removal of redundant vectors (3.2.8) is switched off (“n”), parameters and input values concerning this feature must be removed completely from the input file.

### 3.2.1 Splitting up calculations

Because of the vast chemical space that can be accessed through exhaustively adding substituents to scaffolds, VLG offers the possibility to split up calculations onto separate, serial (non-parallel) runs. As described in section 2.2 VLG takes a given scaffold vector combination and adds substituents to it. If multiple hosts are defined in the input file (which is equal to the number of single processor calculations), VLG evaluates the number of combinations of how substituents can be attached to a specific scaffold vector set. Each host process then treats the total number of possibilities, that is the number of possibilities

how the set of substituents can be attached at all scaffold vector, divided by the number of hosts. The “host id”’s should range from 1  $\rightarrow$   $n$ , where  $n$  is the total number of hosts, in the respective input files.

### 3.2.2 Parameter Filename

Specifies the location and name of the parameter file.

### 3.2.3 Output Folder

VLG will store only the decorated scaffold that pass the cutoff criterion in the output.

### 3.2.4 Output Mode

Defines the verbosity of the output from “silent” (0) to “chatty” (3).

### 3.2.5 Random Seed Number

Used to initialize the internal random number generators which are required for the sampling algorithm described in section 2.4. The random number supplied with the input file is used to initialize a base random number generator, which initializes five additional random number generators. The five independent random number generators are used for:

- rotation angle of the substituents
- rotation angle of the rigid body
- rotation axis and translation direction of the rigid body
- amount of translation of rigid body
- evaluation of metropolis criterion

### 3.2.6 Receptor Specifications

Specifies if VLG is run in stand-alone mode or in conjunction with the receptor. In the latter case, the binding site definition has to be added, where the first line contains the number of amino acids followed by the residue ID’s as specified in the *@<TRIPOS> SUBSTRUCTURE*-section.

The binding site definition of the receptor is used to evaluate the coordinate extrema of the binding site atoms. These values are needed to calculate the boundary of the van der Waals and Coulombic potential energy grid including a user defined security margin (see sections 3.2.7).

### 3.2.7 Grid settings

These sections let the user specify the dimension of the grid and the density of grid points. A sufficiently large margin should be chosen, since molecules for which atoms lie outside of the grid-boundary receive an arbitrarily high energy value and might thus be discarded. For a certain receptor/binding site definition the grid has to be written only once, which reduces the computational cost considerably. For the coulombic section one can specify a distance-dependent dielectric function.

### 3.2.8 Symmetry Evaluation

If enabled, VLG tries to remove redundant substituent vectors (see 2.1 for a description of the input values). This step has to be carried out only once for a given substituent library, as VLG will output the complete list of substituents with their non-redundant set of vectors to standard output.

### 3.2.9 Sampling Algorithm

For an explanation of the input values see 2.4.

### 3.2.10 Cutoff

No molecules with an energy above the cutoff, if specified, are written to disk.

### 3.2.11 Scaffold Specifications

This section includes a list of all scaffolds to be decorated including information of which vectors to be used. Possible vectors are heavy  $\rightarrow$  hydrogen (origin  $\rightarrow$  extension) atom bonds, where the indexes of the origin and the extension of the vectors refer to the first column of the *@<TRIPOS> ATOM* section in the Sybyl MOL2 file (1.1). Restrictions on the set of vectors to be used include the minimal and maximal number of vectors that have to be occupied when adding substituents to a given set of scaffold vectors. Furthermore, the mandatory vectors have to be occupied when substituting multiple sites on the scaffold. "Mandatory" vectors represent a subset of the entire set of vectors (see figure 5 and table 4). Additionally, the set of substituents to be used can be defined individually for every scaffold vector (see substituent set definition in table 4 and section 3.2.12). In the example input file vectors *V1* and *V2* would be modified with substituents from set 1 (benzene, aniline and toluene) and scaffold vector *V3* with substituents from set 2 (ethanol and pyridine).

### 3.2.12 Substituent Specifications

This section comprehends the list of substituents to be attached to the scaffold as well as the set of substituent vectors defined in the same way as described in the scaffold section (3.2.11). Additionally, every substituent is given a set ID with which the user can define multiple sets of substituents to be attached at different scaffold attachment points. The list of substituent set IDs needs to start with "1" and follow a strict consecutive order. The list of set IDs is generated from the individual substituent definitions. Every

substituent has only one set ID. If a given fragment belongs to multiple sets the entire definition of the substituent should be copied and the ID replaced.

### 3.3 Input File Example Without Receptor

In order to run VLG in stand-alone mode, section 3.2.7 has to be removed completely from the input file. Furthermore all information in section 3.2.6, except for the switch stating that no receptor should be used (“n”) and the two variables specifying the use and the value of the dielectric constant, needs to be deleted.

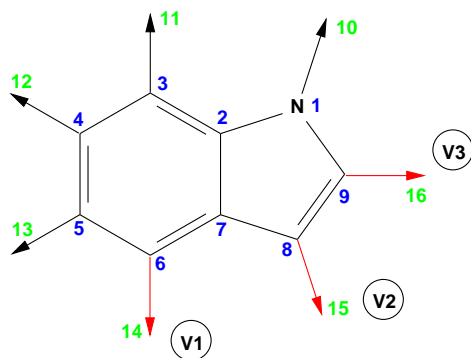


Figure 5: Scaffold definition

Filename	./indole.mol2
Min./max. number of vectors	y 2 3
Number of vectors	3
Vector definition	6 14 8 15 9 16 (= V1, V2 ,V3)
Substituent set specification	1 1 2 (for vectors V1, V2 ,V3)
Number of mandatory vectors	1
List of mandatory vectors	2 (= V2)

Table 4: Scaffold definition

Filename	./toluene.mol2
Substituent set ID	1
Number of vectors	3
Vector definition	2 1 2 3 2 4

Table 5: Substituent definition

## 4 Parameter File

### 4.1 Parameter File Structure

	# The following parameters are MSI CHARMM
	#
4.2.1	# atom element van der Waals
	# type number radius energy_min
	113
	1 B 5 1.17 0.01
	2 C 6 1.870 0.1410
	3 C3 6 1.870 0.1410
	4 C4 6 1.870 0.1410
	5 C5R 6 2.040 0.0500
	6 C5RE 6 2.100 0.0903
	7 C5RP 6 2.040 0.0500
	.
	.
	.
	.
	110 N02 7 1.83 0.09
	111 NR56 7 1.830 0.0900
	112 SK 16 1.89 0.043
	113 S04 16 1.89 0.043
	#
4.2.2	# atom atom bond
	# type type length
	1054
	1 B CT 1.58
	2 B N 1.42
	3 B OE 1.36
	4 B OT 1.325
	.
	.
	.
	.
	1052 MZR XCL 2.350
	1053 MZR XF 1.902
	1054 MZR XI 2.660
	#
4.2.3	# number of forbidden connections
	# list of forbidden connections
	57
	OC OW
	OC OT
	OC OS
	OC OH2
	OC OE
	OC OC
	OE OW
	OE OT
	OE OS
	OE OH2
	.
	.

Table 6: Excerpt of parameter file

## **4.2 Description Parameter File**

### **4.2.1 Atom Section**

Contains a list of atom properties extracted from the MSI CHARMM force field needed for calculating the similarity and the van der Waals energy between two molecules.

### **4.2.2 Bond Section**

Maintains a list of bond parameter extracted from the MSI CHARMM force field as well. The bond distances are needed in order to attach the substituents in an appropriate distance from the scaffold.

### **4.2.3 Forbidden Connections**

This section comprises a list of atom type pairs that should never be connected to one another.



## 5 Troubleshooting

### 5.1 What to do if ...

*VLG does not write any MOL2 files to disk:*

- Check if VLG exited normally (check for “WARNING” messages in output, section 5.2).
- Run VLG without any cutoff restriction (section 3.2.10).
- Increase output verbosity to at least “2” (section 3.2.4), rerun VLG and check if any molecules are being processed.
- Remove all forbidden connections in the parameter file (section 4.2.3) and rerun VLG.

*I checked the sections in the input file concerning the “WARNING”-messages but VLG still exits with the same error:*

- An incomplete or incorrect section, preceding the actual section where the error message was issued, might be responsible for the early termination of VLG.

### 5.2 Error Messages

- **WARNING : File read error occurred in ... section, Exiting !**  
Reason : End-of-file error when trying to read truncated or missing sections in files.  
Solution : Check structure of input file.
- **WARNING : Could not access output folder ... ; Exiting !**  
Reason : Folder does not exist or user does not have write permissions on output folder.  
Solution : Check existence and write permissions of respective folder.
- **WARNING : The file ... cannot be created/opened.**  
Reason : File does not exist or user does not have write permissions on output folder.  
Solution : Check existence or write permissions in the folder where the files should be stored.
- **WARNING : number of scaffolds/substituents <= 0 , Exiting !**  
Reason : Missing scaffold or substituent declaration.  
Solution : Update input file.
- **WARNING : number of vectors for ... is < 1, Exiting !**  
Reason : Scaffolds or substituents where read that do not contain any attachment points.  
Solution : Remove respective scaffold or substituent from input file.

- **WARNING** : The connection vector ... for ... is not correct.  
Reason : Definition of a vector for a specific scaffold or substituent is wrong.  
Solution : Check vector definition.
- **WARNING** : The ... grid you want to read has not been created with the same input file parameters.  
Reason : Reading of a potential energy grid which has been created with other parameters.  
Solution : Re-write grids with the actual input file in use.
- **WARNING** : There are no parameters for atom type ...  
Reason : Encountered unknown atom types while reading certain Sybyl MOL2 files.  
Solution : Check and update atom type definition in MOL2 file or expand VLG parameter file.
- **WARNING** : incorrect ... parameter settings in input file, Exiting !  
Reason : Reading of nonsense values.  
Solution : Check consistency of input values in respective section.
- **WARNING** : incorrect receptor section in input file, Exiting !  
Reason : Incorrect receptor data was read.  
Solution : Check receptor section in input file.
- **WARNING** : ... should be either 'y' or 'n', Exiting !  
Reason : Incorrect switch setting.  
Solution : Check switches in input file.
- **WARNING** : The file switch for ... is neither 'w' nor 'r'.  
Reason : Incorrect switch setting.  
Solution : Check switches in input file.
- **WARNING** : There is more than one substructure in ...  
Reason : Sybyl MOL2-file contained more than one molecule.  
Solution : Remove all but one structure from file.
- **WARNING** : An atom is not in the ... energy grid !  
Reason : When adding substituents certain atoms lie outside of the pre-calculated energy grid.  
Solution : Increase grid margin.
- **WARNING** : Numeric overflow of long int [int size] ;  
reduce number of possibilities for scaffold ...  
Reason : The number of possibilities is too large to be handled by VLG.  
Solution : Reduce the number of substituents.
- **WARNING** : Substituent list index is not sequential, Exiting !  
Reason : The substituent set ID numbering needs to start with "1" and have a consecutive ordering.  
Solution : Update substituent set IDs in the input file (see 3.2.12).

- WARNING : substituent list definition for scaffold ... vector ... is not correct, Exiting !  
Reason : A substituent set ID was selected for a scaffold vector which was not defined in the substituent section of the input file (see 3.2.12).  
Solution : Update substituent set ID of the respective scaffold vector.

## 6 Acknowledgments

I thank A. Caffisch, N. Majeux, P. Kolb, P. Schuetz, A. Cavalli and M. Convertino for useful discussions and comments.

## References

- [1] Peter Ertl. Cheminformatics analysis of organic substituents: identification of the most common substituents, calculation of substituent properties, and automatic identification of drug-like bioisosteric groups. *J. Chem. Inf. Comput. Sci.*, 43:374–380, 2003.
- [2] M. Vieth, M.G. Siegel, R.E. Higgs, I.A. Watson, D.H. Robertson, K.A. Savin, G.L. Durst, and P.A. Hipskind. Characteristic physical properties and structural fragments of marketed oral drugs. *J. Med. Chem.*, 47(1):224–232, 2004.
- [3] F. A. Momany and R. Rone. Validation of the general purpose quanta 3.2/charmm force field. *J. Comput. Chem.*, 13(7):888–900, 1992.
- [4] W. Kabsch. A solution for the best rotation to relate two sets of vectors. *Acta Cryst.*, A32:922–923, 1976.
- [5] S. K. Kearsley and G. M. Smith. An alternative method for the alignment of molecular structures: Maximizing electrostatic and steric overlap. *Tetrahedron Computer Methodology*, 3:615–633, 1990.
- [6] G. Klebe, T. Mietzner, and F. Weber. Different approaches toward an automatic structural alignment of drug molecules: applications to sterol mimics, thrombin and thermolysin inhibitors. *J. Comput.-Aided Mol. Design*, 8:751–778, 1994.
- [7] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in Fortran*. Cambridge University Press, 1992.
- [8] W. Wenzel and K. Hamacher. Stochastic tunneling approach for global minimization of complex potential energy landscapes. *Phys. Rev. E*, 82(15):3003–3007, Apr 1999.
- [9] S. Kirkpatrick, C. D. Gelatt Jr., and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220:671–680, 1983.
- [10] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21:1087–1092, 1953.