Manual for SEED

a program for fragment docking with force field-based evaluation of binding energy

SEED = Solvation Energy for Exhaustive Docking

SEED developers
1998 - 2008: Nicolas Majeux, Marco Scarsi, Joannis Apostolakis, Fabian Dey, Claus Ehrhardt and Amedeo Caflisch
2011 - 2013: Fabian Dey, Tim Knehans, Emilie Frugier and Amedeo Caflisch
2016 - 2017: Cassiano Langini, Jean-Rémy Marchand and Amedeo Caflisch

Manual for Version 4.0

June 2017

Kindly reference the original paper if you use SEED:

N. Majeux, M. Scarsi, J. Apostolakis, C. Ehrhardt, and A. Caflisch. Exhaustive docking of molecular fragments on protein binding sites with electrostatic solvation.
Proteins: Structure, Function and Genetics, **37**:88-105, 1999. [click here for pdf]

The description of the fast energy evaluation is in the second SEED paper:
N. Majeux, M. Scarsi, and A. Caflisch. Efficient electrostatic solvation model for protein-fragment docking.
Proteins: Structure, Function and Genetics, **42**:256-268, 2001. [click here for pdf]

To improve this documentation, please send comments and feedback to:
Amedeo Caflisch, caflisch@bioc.uzh.ch

Contents

A Getting started

A.1 Aim of SEED

The docking approach implemented in the program SEED [1, 2] determines optimal positions and orientations of small-to-medium-sized molecular fragments in the binding site of a rigid protein (hereafter also referred to as receptor) and ranks them according to their binding energy. Polar fragments are positioned such that at least one hydrogen bond with optimal distance to a protein polar group is made (polar docking). For the docking of apolar fragments a novel procedure has been developed to select in an accurate and efficient way the hydrophobic regions of the protein, i.e., those with low electrostatic desolvation and favorable van der Waals interactions with an uncharged probe sphere. Furthermore, the numerical Generalized Born methodology developed in the Caflisch group [3, 4] and *ad hoc* look-up tables are employed to efficiently evaluate the protein and fragment desolvation upon binding and the screened electrostatic interaction.

A.2 Running SEED

The command to run SEED in a Unix shell is

```
seed.exe seed.inp >& seed.log
```

where `seed.exe` and `seed.inp` are the executable and the input file, respectively.

A.3 Files required for a SEED run

The files required for a SEED run are:

- The file `seed.inp` contains the most frequently modified input values.
- The parameter file `seed.par` contains less frequently modified input/output options, parameters for docking, energy and clustering.

  The two files `seed.inp` and `seed.par` have comment lines which start with a # and both files terminate with the word `end`. All the other lines are information read by the program. In the following, lines referring to the input (see page 6) and parameter (see pages 23-26) files are indicated by **i** and **p**, respectively. Please note that the values shown are only indicative. For a working example refer to the test cases provided with the distribution.

  The path for the parameter file is in the first line of the input file (referred to as **i1**).
- A standard SYBYL mol2 format file containing all the fragments to dock (**i7**). This file is simply the concatenation of all the fragments, expressed in mol2 format. Note that different conformations of the same fragment are treated as different fragments. Partial charges are written in the 9th column in the `@<TRIPOS>ATOM` record as for the receptor.
  In order to assign the correct Van der Waals parameters from the file `seed.par`, the CHARMM atom types should be specified in the mol2 file. This is done using the

alternative atom type specified by the record `@<TRIPOS>ALT_TYPE`, which takes the following form:

```
@<TRIPOS>ALT_TYPE

CGenFF_4.0_ALT_TYPE_SET

CGenFF_4.0 1 CG331 2 CG301 3 CG331 4 CG324 ...
```

Where `CGenFF_4.0_ALT_TYPE_SET` sets a user-defined name (for example `CGenFF_4.0`) for the alternative atom type set. This name is repeated on the next line, followed by the list of atom number-atom type pairs for each atom in the molecule. This list should span a single line, but can be broken by using `\\`.
The first line of the SYBYL record `@<TRIPOS>MOLECULE` specifies the fragment name. It is convenient (but not necessary) to have unique names for each fragment. In case fragments with duplicate names are found in the input, they will be renamed in all the output files appending to their name the dollar sign $ and an incremental index. As the fragment mol2 input file is read sequentially, the number of fragments in it does not have to be specified a priori.

- A standard SYBYL mol2 file (**i2**) for the receptor with partial charges on the 9th column in the `@<TRIPOS>ATOM` record and CHARMM atom types specified by the `@<TRIPOS>ALT_TYPE` record (refer to the fragment file description for details).

A.4 Input file parameters

In the following, parameters of the input file `seed.inp` (page 6) are explained.

i1 Path for the parameter file `seed.par`.

i2 Receptor coordinate file (SYBYL mol2 format).

i3 The first line is the number of residues in the binding site and the following lines are the residue sequential numbers (e.g. if Arg__38 is the first residue of the protein, its sequential number is 1 and not 38). Binding site metal ions have to be in the list.

i4 The first line is the number of user-selected points in the binding site and the following lines are their coordinates. These points are used to select polar and apolar receptor vectors that meet an angle criterion (**p14**, see B.1.3) such that vectors pointing outside of the binding site are discarded. The points can be for example the fragment heavy atoms of a known fragment-receptor complex structure.

i5 The first line is the total number of vectors (refer to B.1 for their meaning) for the metal ions and each of the following lines contains the atom index of the metal as it is in the receptor mol2 file and the coordinates of the vector extremity.

i6 Coordinates of the center and radius of the sphere in which the geometry center of the fragment position must be to be accepted. This filter can be discarded by selecting `n` instead of `y`.

i7 First line: a character specifying the computation mode: `e` if energy evaluation only has to be carried out (see B.4) or `d` if docking is also requested (see B.3). Following line: the first column contains the path of the fragment mol2 file and the second column allows the selection of apolar, polar docking or both. The fragment position is accepted if

the total energy (according to the approximate model of section B.2.2) is smaller than a cutoff given in the third column. The second clustering (see **??**) is applied on the positions for which the binding energy of the cluster representative is smaller than a cutoff value specified in the 4th column.

A.5 Most important output files

The main output file, whose filename is specified in **p6**, contains the energy values and results of clustering.

A directory `outputs` in which all the output files are written is automatically created by the program. Note that if a directory named `outputs` is already present, it will be overwritten by the SEED run.

<`FragmentMol2FileName`>`_clus.mol2` contains the fragment positions with best energy after the postprocessing step. This file is the concatenation of a mol2 file for each saved pose. The maximum number of poses to be saved per cluster can be set in **p5** (first value). The comment line of the SYBYL mol2 record `@<TRIPOS>MOLECULE` (6th line after the record identifier) contains some useful information about the pose, *i.e.* increasing pose index, cluster number, total energy and fragment number (`Fr_nu`). The latter represents the program internal numbering of the pose and it is not interesting *per se*, but it can be used to match the pose to docking information written in the main output file.

`seed_clus.dat` is a summary table containing the separate energy terms for each fragment position saved to <`FragmentMol2FileName`>`_clus.mol2`. This information can be also retrieved from the main output file. Columns are organized as follows:

1. **Name:** Fragment name.
2. **Pose:** Incremental pose number. This index restarts at 1 for each new fragment.
3. **Cluster:** Cluster number.
4. **Fr_nu:** Fragment number. This is SEED internal pose number.
5. **Tot:** Total binding energy.
6. **ElinW:** Electrostatic interaction in water.
7. **rec_des:** Desolvation of the receptor upon complex formation.
8. **frg_des:** Desolvation of the fragment upon complex formation.
9. **vdW:** Van der Waals interaction energy.
10. **DElec:** Electrostatic difference upon fragment binding. It is given by $ElinW - DG\_hydr$. It roughly represents how good the fragment feels in the protein compared to how good it feels in water.
11. **DG_hydr:** Free energy of hydration of the fragment.
12. **Tot_eff:** Tot/HAC.
13. **vdW_eff:** vdW/HAC

14. **Elec_eff:** $ElinW/HAC$

15. **HAC:** Heavy atom count. It is the total number of non-hydrogen atoms in the fragment.

16. **MW:** Molecular weight of the fragment.

<FragmentMol2FileName>_best.mol2 contains the best fragment positions, according to the total energy, irrespective of the cluster they belong to (parameter **p5**, second value). The difference with respect to <FragmentMol2FileName>_clus.mol2 is that the user can set the total number of poses to be saved instead of the number of cluster members.

seed_best.dat is the same as seed_clus.dat but matching <FragmentMol2FileName>_best.mol2.

Note that the number of cluster members to be saved (first value of **p5**) determines the maximum number of poses for which to evaluate the slow energy during postprocessing. Thus in general it is advisable to set this number to a value higher than one, in order to be sure to consider a meaningful number of poses, and to suppress the corresponding mol2 file output (first value of **p3** set to no) as it may quickly become big.

A.6 Starting a new project

When a new project is started, it is useful to first generate the vectors without docking any fragment (**i7** set to d). Of the six files listed below one should visualize the two files polar_rec_reduc.mol2 and apolar_rec_reduc.mol2. It is useful to modify the appropriate parameters if the vector distributions do not meet the user's expectation, since fragments are docked using the vectors present in the two aforementioned files. After this test one has just to read the maps (**p7-p8**: r) instead of generating them again.

- polar_rec.mol2 contains vectors distributed uniformly on a spherical region around each ideal H-bond direction. The deviation from ideal hydrogen bond geometry and the number of additional vectors to distribute uniformly on the spherical region are set in **p12**.

- polar_rec_reduc_angle.mol2 contains vectors of polar_rec.mol2 which are selected according to an angle criterion (**i4**, **p14**). Vectors pointing outside of the binding site are discarded. The file polar_rec_reduc_angle.mol2 exists only if the angle criterion has been activated by the user (**i4**).

- polar_rec_reduc.mol2 contains vectors of polar_rec.mol2 (or of polar_rec_reduc_angle.mol2 if the angle criterion has been activated (**i4**)) which are selected according to favorable van der Waals interaction between all the receptor atoms and a spherical probe on the vector extremity. The aim is to discard receptor vectors that point into region of space occupied by other atoms of the receptor and select preferentially vectors in the concave regions of the receptor. The van der Waals radius of the probe is specified in **p15**. The number of selected vectors is controlled with **p2**.

- apolar_rec.mol2 contains points distributed uniformly on the solvent-accessible surface of the receptor. The density of surface points is set in **p22**.

- `apolar_rec_reduc_angle.mol2` contains vectors of `apolar_rec.mol2` which are selected according to an angle criterion (**i4**, **p14**). Vectors pointing outside of the binding site are discarded. The file `apolar_rec_reduc_angle.mol2` exists only if the angle criterion has been activated by the user (**i4**).

- `apolar_rec_reduc.mol2` contains points of `apolar_rec.mol2` (or of `apolar_rec_reduc_angle.mol2` if the angle criterion has been activated (**i4**)) which are selected according to their hydrophobicity. For this purpose a low dielectric sphere is placed on each of these points. The hydrophobicity is defined as the weighted sum of the receptor desolvation energy due to the presence of the probe and the probe/receptor van der Waals interaction. The weighting factors and the probe radius are set in **p22**. The number of selected apolar points is controlled with **p2**.

A.7 Troubleshooting in case of empty output

If after starting a SEED run the program exits unexpectedly, the keyword `WARNING` should be looked for in the main output file (**p6**) to find an hint on possible problems (wrong path for filenames, unknown value for some parameters ...).

If the main output file does not contain any fragment position for a given fragment type, it can be due to several reasons: the center of the sphere (**i6**) might be misplaced (outside the binding site), the checking of clashes (**p10**, see **??**; **p11**, see B.3.2 and **??**) too strict, the van der Waals energy cutoff (**p23**) for apolar fragments too severe, the total energy cutoff (third column of **i7**) too stringent. To find out what the reason could be, the following part of the main output file should be investigated:

```
Total number of generated fragments of type 1 (BENZ) : 118800
Fragments that passed the sphere checking : 102894
Fragments that passed the bump checking : 49007
Fragments that passed the vdW energy cutoff : 22100
Fragments that passed the total energy cutoff : 17794
```

```
#           Parameter filename
i1  ./seed4_cgenff4.par
#           Receptor coordinates (in mol2 format) filename
i2  receptor.mol2
#           Binding site residue list
#           First line:  number of residues
#          Following lines: residue numbers (one each line)
i3  1
      101
#           Modification of February 2002:
#           List of points (e.g. ligand heavy atoms of a known ligand-receptor
#           complex structure) in the binding site used to select polar and apolar
#           rec. vectors which satisfy the angle criterion (see parameters file)
#           First line:  number of points (0: no removal of vectors using the angle
#           criterion)
#           Following lines:  coordinates of the points
i4  6
      62.881  39.578  -4.449
      61.755  39.106  -4.248
      61.465  37.747  -3.723
      60.663  39.820  -4.501
      60.749  41.185  -5.002
      59.457  41.937  -4.647
#           Metals in the binding site
#           Make sure that the residue number of the metal is in the
#           binding site residue list.
#           First line:  total number of coordination points
#           Following lines:  atom number of metal / x y z of coordination point
i5  0
#           Spherical cutoff for docking (y,n / sphere center / sphere radius)
i6  n  2.133  1.359  25.539   10.00
#           Fragment library specifications
#           First line: Number of fragments / dock+energy (d), only energy (e)
#           Second line: Fragment library filename /
#           apolar docking (a), polar docking (p), or both (b) /
#           energy cutoff in kcal/mol / 2nd clustering cutoff in kcal/mol
i7  d
    /home/ligands/library.mol2          b     2.0   2.0
    end
```

Figure 1: *Seed input file (**seed.inp**).*

B Carrying on

B.1 Vectors for docking

The binding site where the fragments are to be docked is defined by a list of receptor residues (possibly including explicit water molecules, treated as part of the receptor molecule). The first line of **i3** is the number of residues in the binding site and the following lines are the residue sequential numbers (e.g. if Arg_38 is the first residue of the protein, its sequential number is 1 and not 38). If a metal ion belongs to the binding site, its sequential number also has to be in the list.

B.1.1 Vectors for polar docking

Fragments are considered polar if they have at least one H-bond donor or acceptor. SEED docks polar fragments where at least one hydrogen bond with good geometry is made. First, predefined rules (Figure 1) allow the distribution of vectors of unitary length on all

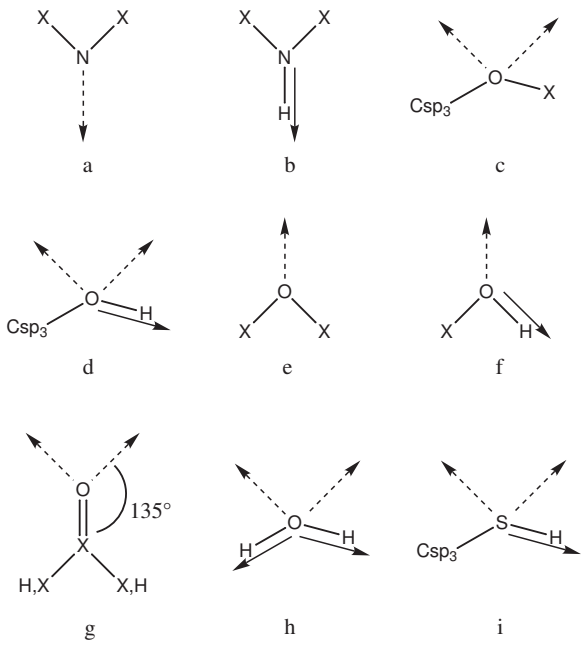

Figure 2: *Description of polar vectors for the fragment and for the receptor. X is a heavy atom. The broken arrow represents a vector of H-bond acceptor in the lone pair direction and the full arrow a vector of H-bond donor. The geometry of c, d, h and i is tetrahedral (angle of* 109°*). Examples: (a) imidazole, pyridine, (b) protein backbone, imidazole, indole, (c) ethers, (d) Ser and Thr side chains, sugars, (e) methoxybenzene, (f) Tyr side chain, phenol, (g) Asn, Gln, Asp, and Glu side chains, protein backbone, acetamide, (h) water, (i) Cys side chain.*

H-bond groups of the fragment in a direction for an ideal H-bond geometry. For example, if a nitrogen atom is bound to two heavy atoms, one H-bond vector is generated in the direction of either the lone pair (Figure 1a) or the NH bond (Figure 1b). The same procedure is then used for the polar groups in the receptor binding site (backbone and

side chains). These rules are based on the atomic element number. A correspondence between atom types and atomic element numbers has to be given in **p29**: the first line is the total number of correspondences and the first three terms of the following lines are respectively a sequential number, the atom type and the atomic element number. Vectors for metal ions have to be provided by the user. The first line of **i5** is the total number of vectors for the metal ions and each of the following lines contains the atom number of the metal as it is in the receptor mol2 file and the coordinates of the vector extremity. The vector is then built by joining the vector extremity to the metal ion center.

For the receptor polar groups and metal ions an additional set of vectors is distributed uniformly on a spherical region around each of the ideal directions to increase the spatial sampling. The first term of **p12** is the maximal angular deviation from ideal hydrogen bond geometry and the second term is the number of additional vectors to distribute uniformly on the spherical region.

To discard receptor vectors that point into a region of space occupied by other atoms of the protein and select preferentially vectors in the concave regions of the receptor a spherical probe is set on the vector extremity at a distance corresponding to the sum of the van der Waals radii of the acceptor or donor atom and the probe. The van der Waals radius of the probe in Å is specified in **p15** and those of the atom types are specified in the 4th column of **p29**. The van der Waals interaction (see below) between the probe and all the receptor atoms is then evaluated except for the receptor hydrogen atom involved in the H-bond. The vectors which show less favorable van der Waals energies are discarded. The number of selected polar vectors is modified through the first term of **p2**. Finally, the docking itself is achieved by matching a H-bond vector of the receptor with a H-bond vector of the fragment at a distance that depends on the atom types of donor and acceptor involved in the hydrogen bond. These bond lengths are specified in **p30** (a default length on the first line and two blocks where lengths are set between element types and atom types respectively; each block starts with the number of following lines in the block). The fragment is then rotated around the H-bond axis to increase sampling. The number of rotations is set in **p13**.

B.1.2 Vectors for apolar docking

SEED docks apolar fragments into hydrophobic regions of the receptor. First, a number of points are distributed uniformly on the solvent-accessible surface (SAS) of the fragment. The density of surface points for the fragment is set in the second term of **p22**. Second, an automatic procedure defines the hydrophobic regions on the receptor. For this purpose a number of points are uniformly distributed on the SAS of the binding site (density of surface points for the receptor in the first term of **p22**). A low dielectric sphere is placed on each of these points, and the receptor desolvation energy (see below) and the probe/receptor van der Waals interaction are evaluated. The radius of the sphere is the third term of **p22**: a value of 1.4 Å allows a finer description of the narrow pockets than with a value of 1.8 Å. The points on the receptor SAS are then ranked according to the sum of the two energy terms weighted by scaling factors that are the last two terms of **p22**. The number of selected apolar points can be modified with the second term of **p2**. For both the fragment and the receptor, vectors are defined by joining each point on the SAS with the corresponding atom center. Finally, apolar fragments are docked by matching a vector of the fragment with a vector of the receptor at the optimal van der Waals distance. To improve sampling additional rotations of the fragment are performed

around the axis joining the receptor atom and fragment atom. The number of rotations is set in **p13**.

B.1.3 Selection of receptor vectors using an angle criterion

(Modification of February 2002)

To discard polar and apolar receptor vectors that point outside of the binding site a selection using an angle criterion (Figure 2) can be activated (**i4**, **p14**).

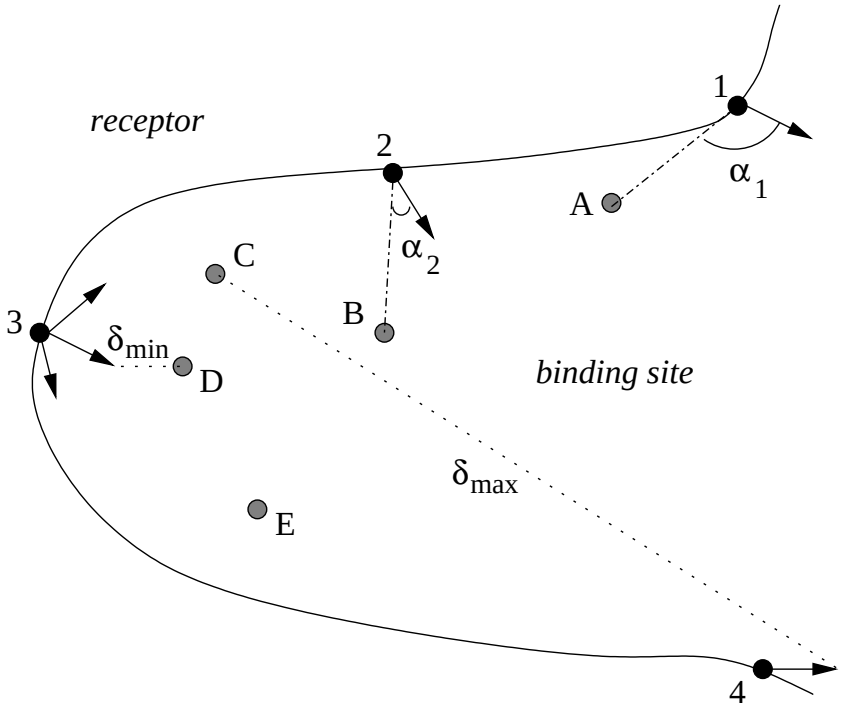

Figure 3: *1-4: receptor atoms and vectors. A-E: user-defined anchor points in the binding site (e.g., fragment heavy atoms). The angle between a vector and its closest anchor point in the binding site is shown for two vectors (α_1, α_2). Reasonable parameters should allow to remove the vector of atom 1 from the list of receptor vectors and keep the vector of atom 2. δ_{min} and δ_{max} are defined in the text.*

It is applied directly after vectors have been distributed on the binding site, i.e., before the selection by means of a spherical probe for polar vectors and before the selection by means of a low dielectric sphere for apolar vectors. The first line of **i4** is the number of user-defined anchor points in the binding site and the following lines are their coordinates. The anchor points can be for example the fragment heavy atoms obtained from a known fragment-receptor complex structure. The minimal and maximal distances (δ_{min} and δ_{max}) between the extremity of the vectors and the anchor points in the binding site are first evaluated. A vector is then discarded if the angle between the vector and the closest anchor point in the binding site (angle anchor_point–vector_origin–vector_extremity) is larger than an angle cutoff. The angle cutoff is $\mathbf{p14}_1$ (first parameter in **p14**) if the distance between the vector and the closest anchor point is smaller or equal to $\delta_{min} \times \mathbf{p14}_3$; the angle cutoff is $\mathbf{p14}_2$ if the distance is larger or equal to $\delta_{max} \times \mathbf{p14}_4$. For other distances the angle cutoff value falls between $\mathbf{p14}_1$ and $\mathbf{p14}_2$ (linear dependence). Reasonable

parameters provide permissive angle cutoffs for vectors close to an anchor point and stricter angle cutoffs for distant vectors.

B.1.4 Polar and apolar docking

Some "polar" fragments can have considerable hydrophobic character (e.g., diphenylether). Therefore, they can also be docked by the procedure for apolar fragments. The second column of **i7** allows the user to select apolar docking, polar docking or both.

B.2 Energy in SEED

The binding energy is evaluated in SEED as the sum of the Van der Waals interaction and the electrostatic energy. The main assumption underlying the evaluation of the electrostatic energy in solution of a fragment-receptor complex is the description of the solvent effects by continuum electrostatics. The system is partitioned into solvent and solute regions and different dielectric constants are assigned to each region (dielectric constant of the solute, i.e. receptor and fragment, in **p1** (usually between 1.0 and 4.0) and dielectric constant of the solvent, normally 80 for water, as the 3rd term of **p21**). In this approximation only the intra-solute electrostatic interactions need to be evaluated explicitly, strongly reducing the number of energy evaluations with respect to an explicit treatment of the solvent.
The procedure to calculate the difference in electrostatic energy ΔG_{electr} upon binding of a fragment to a receptor is depicted in Fig. 4. The binding process (first row of Fig. 4) is decomposed into a cycle by introducing an uncharged copy of the solute (white-filled receptor and fragment). The binding free energy can then be decomposed into three terms:

- Partial desolvation of the receptor: electrostatic energy difference upon binding an uncharged fragment to a charged receptor in solution.
- Partial desolvation of the fragment: electrostatic energy difference upon binding a charged fragment to an uncharged receptor in solution. This term makes use of the GB treatment in the accurate energy model.
- Screened fragment-receptor interaction: intermolecular electrostatic energy in solution. This is represented as the swapping of the charged and uncharged fragment in the bottom of Fig. 4. The term is written as the pairwise sum $\sum_{i \in fragment, j \in protein} q_i \phi_j$ where the values of ϕ_j (i.e., the electrostatic potential of the protein atom j) are calculated by the GB model in the accurate energy description.

Note that the addition of the uncharged solute (highlighted by the brown box) does not modify the electrostatic energy as this solute does not interact with water.
The single terms can be evaluated in SEED according to two different energy models: the accurate model (section B.2.1) and the fast model (section B.2.2). The two models are combined together in a two-step procedure (see B.3)

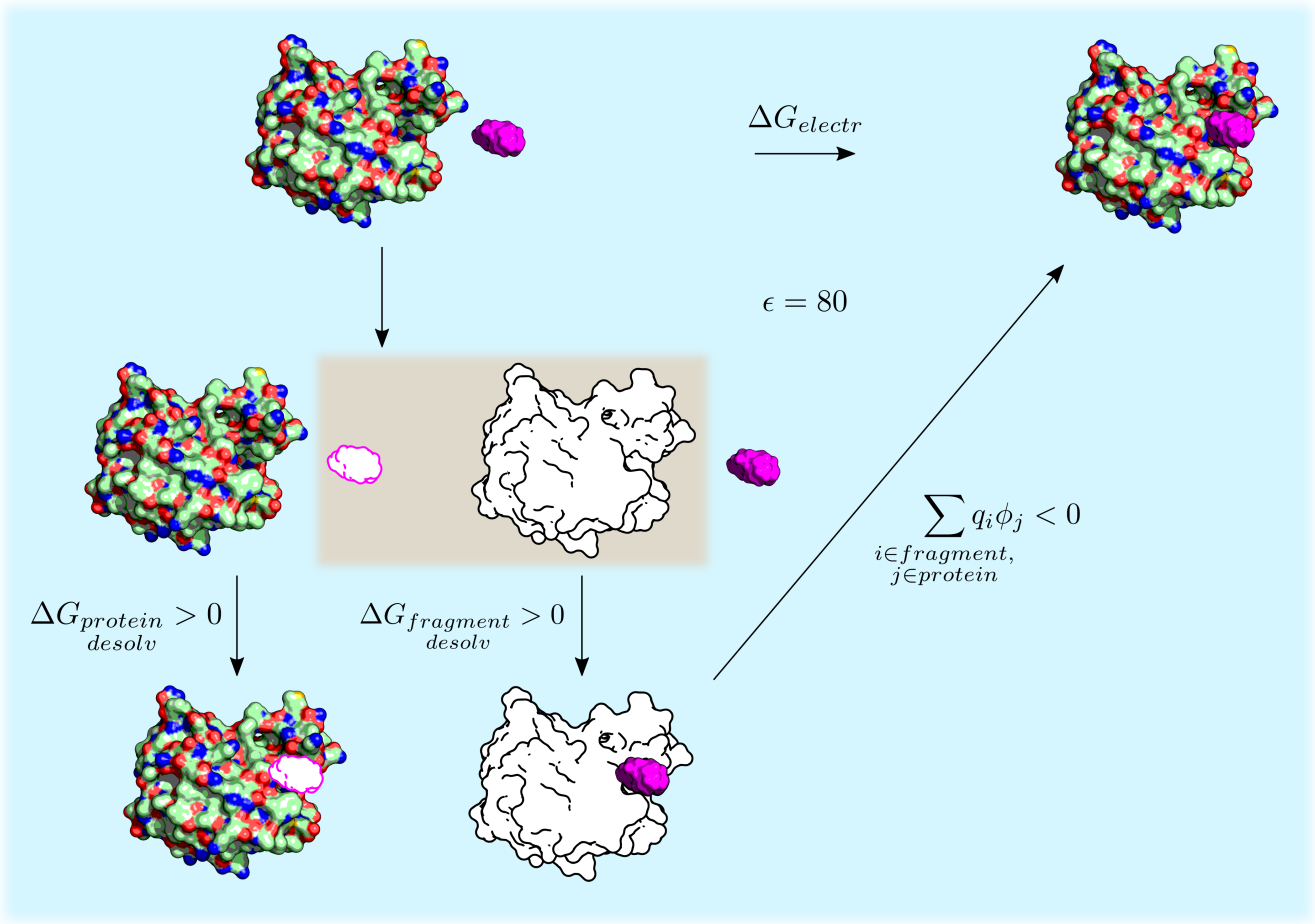

Figure 4: *Decomposition of the electrostatic energy difference upon receptor-fragment binding* ΔG_{electr}*. Charged solute is colored and uncharged solute is white. The fragment is in purple. The whole cycle takes place in water (blue background). The brown background box highlights the introduction of an uncharged copy of the solute.*

B.2.1 Accurate model

Van der Waals interaction. A list of residue centroids is generated during the initial phase of the program and is used for an efficient estimation of van der Waals and screened electrostatic interactions. The atom closest to the geometrical center of the residue is selected as centroid for residues with zero formal charge while the atom closest to the charge center is chosen for charged residues. The latter choice is more appropriate for the electrostatic interaction (see below). A 3D grid is built over the receptor with a distance between neighbor grid points of usually 1 Å (set in the second term of **p18**). Each centroid is assigned to the closest cubic element of the grid. Given a grid point m, all the grid points falling at a distance from m smaller than a given cutoff (first term of **p18**) define a pseudo-sphere associated to m. The neighbor list of a given fragment atom contains the atoms belonging to the receptor residues whose centroid is included in the pseudo-sphere centered on the grid point closest to the fragment atom. This increases the efficiency because it avoids to calculate the distances between each fragment atom and all the receptor atoms when evaluating the interaction energy of a new fragment position. The van der Waals interaction energy is then computed between each atom of the fragment and the receptor atoms in the neighbor list according to

$$E_{ij}^{\mathrm{vdW}} = \sqrt{\varepsilon_i \varepsilon_j} \left\{ \left(\frac{R_i^{\mathrm{vdW}} + R_j^{\mathrm{vdW}}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_i^{\mathrm{vdW}} + R_j^{\mathrm{vdW}}}{r_{ij}} \right)^{6} \right\} \quad (1)$$

where ε_i is the minimum of the van der Waals potential between two atoms of type i at optimal distance of $2 \cdot R_i^{\mathrm{vdW}}$. R_i^{vdW} and ε_i are specified in the 4th and 5th columns of **p29**.

Partial desolvation of the receptor. The electrostatic energy in solution of the receptor can be expressed in terms of the electric displacement vector $\vec{D}(\vec{x})$ and of a location dependent dielectric constant $\epsilon(\vec{x})$ as an integral over three-dimensional space R^3:

$$E = \frac{1}{8\pi} \int_{R^3} \frac{\vec{D}^2(\vec{x})}{\epsilon(\vec{x})} \, d^3x \quad (2)$$

Since $\vec{D}(\vec{x})$ is additive, for point charges it can be rewritten as a sum over all charges i of the receptor:

$$\vec{D}(\vec{x}) = \sum_i \vec{D}_i(\vec{x}) \quad (3)$$

Docking an uncharged molecular fragment in the receptor binding site has the only effect of modifying the dielectric properties of part of the binding site. Over the volume occupied by the fragment the dielectric constant changes from the solvent value (ϵ_{w}) to the interior value (ϵ_{int}). The volume occupied by the fragment consists of the actual volume of the fragment and the interstitial volume enclosed by the reentrant surface between fragment and receptor (the first two terms of **p18** are used for the construction of the SAS, i.e. solvent accessible surface, employed in this scheme). **In the limit in which the receptor electric displacement vector $\vec{D}$ does not change significantly upon fragment docking (i.e., for small fragments and not close to a cluster of charges on the protein surface),** the variation of the electrostatic energy of the receptor can be written

according to equation 2 as:

$$\Delta E = \frac{\tau}{8\pi} \int_{V_{frag}} \vec{D}^2 \left(\vec{x}\right) d^3x \tag{4}$$

where $\tau = \frac{1}{\epsilon_{\text{int}}} - \frac{1}{\epsilon_{\text{w}}}$ and V_{frag} is the volume occupied by the fragment. On a 3D grid equation 4 becomes:

$$\Delta E = \frac{\tau}{8\pi} \sum_{k \in V_{frag}} \vec{D}^2 \left(\vec{x_k}\right) \Delta V_k \tag{5}$$

The receptor electric displacement is calculated over a 3D grid (grid margin and spacing set in **p20**) and it is approximated by the Coulomb field

$$\vec{D}\left(\vec{x}\right) = \sum_i q_i \, \frac{\left(\vec{x} - \vec{x_i}\right)}{\left|\vec{x} - \vec{x_i}\right|^3} \tag{6}$$

This is an analytical approximation of the total electric displacement Alternatively, $\vec{D}$ can be calculated exactly for the isolated receptor by a finite difference solution of the Poisson equation and assumed not to change significantly upon fragment docking:

$$\vec{D}\left(\vec{x}\right) = -\epsilon\left(\vec{x}\right) \nabla \phi\left(\vec{x}\right) \tag{7}$$

where ϕ is the electrostatic potential solution of the Poisson equation. This calculation is not available any longer within SEED.

Screened fragment-receptor interaction. The fragment-receptor interaction in solution is calculated via the GB approximation. The interaction energy in solution between two charges embedded in a solute is

$$E_{ij}^{int} = \frac{q_i q_j}{\epsilon_{\text{int}} r_{ij}} - \frac{q_i q_j \tau}{R_{ij}^{GB}} \tag{8}$$

where

$$R_{ij}^{GB} = \sqrt{r_{ij}^2 + R_i^{eff} R_j^{eff} exp\left(\frac{-r_{ij}^2}{4 R_i^{eff} R_j^{eff}}\right)} \tag{9}$$

q_i is the value of the partial charge i, while r_{ij} is the distance between charge i and j. R_i^{eff} is the effective radius of charge i and it is evaluated numerically on a 3D grid covering the solute as described in [3]. It is a quantity depending only on the solute geometry and represents an estimate of the average distance of a charge from the solvent.

The intermolecular interaction energy is calculated as:

$$E^{int} = \sum_{\substack{i \in fragment \\ j \in list_i}} E_{ij}^{int} \tag{10}$$

where $list_i$ contains the receptor atoms belonging to the neighbor list of fragment atom i. The electrostatic neighbor list includes all the receptor atoms of the van der Waals neighbor list (see above) and one atom for every charged residue whose centroid falls within a given cutoff (radius of the pseudo-sphere increased by 30%; **p18**) of the binding site residues. The atom selected is the one closest to the center of charge. Supplementing the van der Waals neighbor list with a monopole approximation of distant charged residues dramatically reduces the error originating from the long range effects of electrostatics.

Partial desolvation of the fragment. The fragment intramolecular energy in solution is calculated with the GB formula as described in [3]:

$$E = \sum_{i \in fragment} E_i^{self} + \sum_{\substack{i>j \\ i,j \in fragment}} \left(\frac{q_i q_j}{\epsilon_{\text{int}} r_{ij}} - \frac{q_i q_j \tau}{R_{ij}^{GB}} \right) \tag{11}$$

where the two sums run over the partial charges of the fragment. Equation 11 differs from equation 10 due to the presence of the *self-energy* term $\sum_i E_i^{self}$. This term is not zero only in the case of intramolecular energies. E_i^{self} is the *self-energy* of charge i and represents the interaction between the charge itself and the solvent. It is calculated as

$$E_i^{self} = \frac{q_i^2}{2R_i^{vdW}\epsilon_{\text{int}}} - \frac{q_i^2 \tau}{2R_i^{eff}} \tag{12}$$

where R_i^{vdW} is the van der Waals radius of charge i.

The difference in the intramolecular fragment energy upon binding to an uncharged receptor in solution is:

$$\Delta E = E^{docked} - E^{free} \tag{13}$$

where E^{docked} and E^{free} are the energies of the fragment bound and unbound to the receptor in solution, respectively. They are evaluated according to equation 11. For the unbound fragment (E^{free}) the effective radii are calculated considering as solute the volume enclosed by the molecular surface of the fragment. For the bound fragment (E^{docked}) the solute is the volume enclosed by the molecular surface of the receptor-fragment complex. E^{free} is evaluated only once per fragment type, while E^{docked} is recalculated for every fragment position in the binding site.

Empirical correction term. (Modification of March 2003) An empirical correction term (equation 8 in [5]) to the Coulomb field approximation in the generalized Born model is used for the accurate screened interaction and fragment desolvation energies. Protein desolvation does not use the GB model.

B.2.2 Fast model

Van der Waals interaction. The van der Waals interaction between a fragment and the receptor is described as the sum of a steep repulsion and an attractive dispersion term

with the 6-12 Lennard-Jones model:

$$E_{\text{vdW}} = \sum_{i \in \text{fragment}} \sum_{j \in \text{receptor}} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) \tag{14}$$

where r_{ij} is the distance between atoms i and j, A_{ij} and B_{ij} are van der Waals repulsion and attraction parameters. The assumption that the receptor is rigid favors the use of a grid-based evaluation of the interaction. To make the fragment and receptor terms in equation (14) factorizable, the geometric mean approximation is used: $A_{ij} = \sqrt{A_i A_j}$ and $B_{ij} = \sqrt{B_i B_j}$, with $A_i = \varepsilon_i (2R_i^{\text{vdW}})^{12}$ and $B_i = 2\varepsilon_i (2R_i^{\text{vdW}})^{6}$. R_i^{vdW} is the van der Waals radius of atom i and ε_i is the minimum of the van der Waals potential between two atoms of type i at optimal distance of $2R_i^{\text{vdW}}$ (R_i^{vdW} and ε_i are specified in the 4th and 5th columns of **p29**). A grid is spanned over the binding site of the receptor and the grid spacing is usually 0.2 Å or 0.3 Å (grid margin and spacing set in **p17**). When the program starts, for every grid point p the two following "receptor potentials" are calculated and stored in look-up tables:

$$\phi^A(p) = \sum_{j \in \text{receptor}} \frac{\sqrt{A_j}}{r_{pj}^{12}} \qquad \text{and} \qquad \phi^B(p) = \sum_{j \in \text{receptor}} \frac{\sqrt{B_j}}{r_{pj}^{6}} \tag{15}$$

where the sums run over the receptor atoms which are within a 10 Å cutoff distance of the grid point. The contribution of fragment atom i with coordinates $\vec{x}_i$ is evaluated by multiplying its van der Waals parameters ($\sqrt{A_i}$ and $\sqrt{B_i}$) with the "receptor potentials" (ϕ^A and ϕ^B, respectively). The value of the potential is derived from the eight points of the grid surrounding $\vec{x}_i$ by the trilinear interpolation method [6].

Partial desolvation of the receptor. A preliminary step consists of the evaluation of the receptor desolvation due to a low dielectric probe sphere of 1.4 Å radius rolling over the van der Waals surface of the receptor. The center of the sphere spans the solvent accessible surface (SAS). A number of points are distributed uniformly on the SAS of the receptor with a given surface density (usually 0.5 points per Å^2, see below) to describe the different positions of the center of the probe sphere. Furthermore, a cubic grid of 0.5 Å spacing is used to discretize the volume surrounding the receptor. The volume occupied by the probe sphere is then approximated on the cubic grid. The receptor desolvation resulting from the probe sphere at a point p on the SAS of the receptor (see Figure 5a) is evaluated according to the Coulomb approximation of the electric displacement:

$$\Delta G^p_{\text{desolv}} = \frac{1}{8\pi} \left(\frac{1}{\epsilon_{\text{int}}} - \frac{1}{\epsilon_{\text{w}}} \right) \sum_{k_p \in V_{probe}} \left(\sum_{j \in receptor} q_j \frac{\left(\vec{x}_{k_p} - \vec{x_j} \right)}{\left| \vec{x}_{k_p} - \vec{x_j} \right|^3} \right)^2 \Delta V \tag{16}$$

where the index k_p runs over the cubic grid elements occupied by the probe sphere and ΔV is the volume of a cube. Further, $\vec{x_j}$ is the coordinate vector of the receptor atom j, $\vec{x}_{k_p}$ the position of the cube included in the probe sphere, and ϵ_{int} and ϵ_{w} are the solute and solvent dielectric constants, respectively. The receptor desolvation due to the probe sphere is calculated only once at the beginning of SEED for every point on the SAS. It is always positive, i.e., unfavorable, because $\epsilon_{\text{int}} < \epsilon_{\text{w}}$.

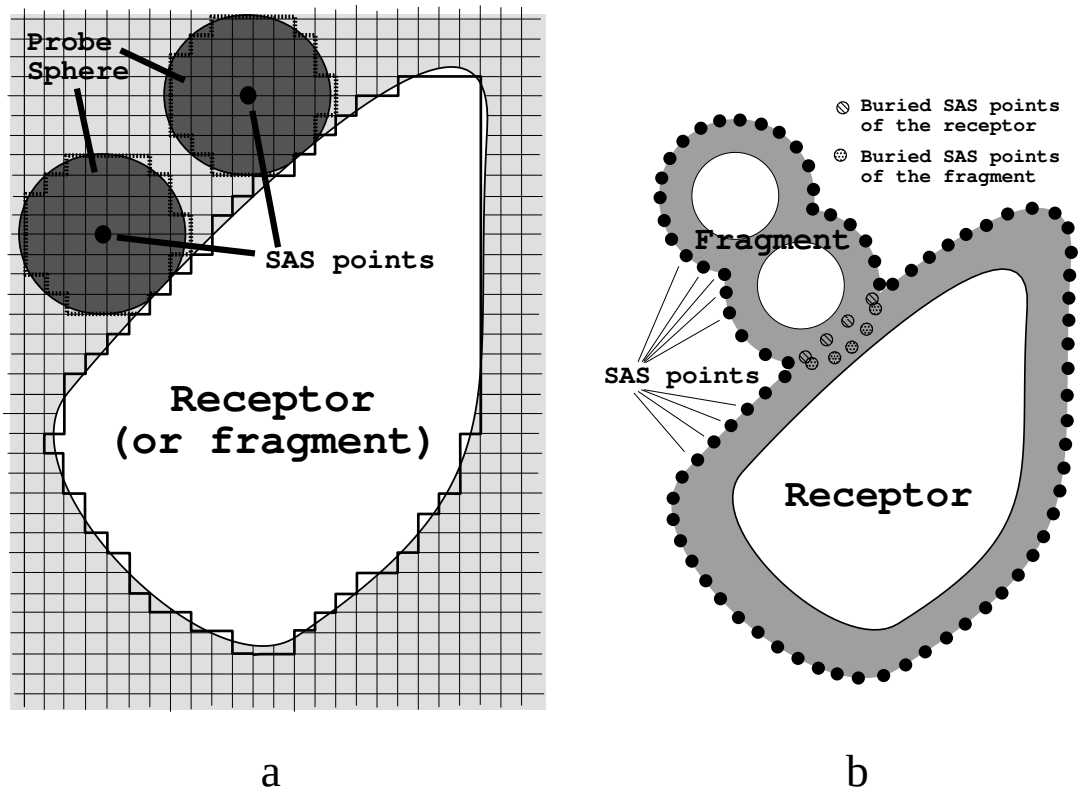

Figure 5: *(a) Evaluation of the receptor (fragment) desolvation due to a probe sphere rolling over the van der Waals surface of the receptor (fragment). A grid box of 0.5 Å spacing (light gray) is built around the molecule and the desolvation resulting from the occupation of every grid element is evaluated as described in [1]. The desolvation due to the probe sphere in a given position is approximated by the sum of the values of the desolvation due to the cubes within the sphere (dark gray). (b) Evaluation of the receptor and fragment desolvations upon binding. The points on the SAS of the receptor and the SAS of the fragment represent the positions of the rolling probe sphere. The receptor desolvation is approximated by the sum of the desolvation values associated with the SAS points of the receptor buried within the SAS of the fragment. The fragment desolvation is approximated by the sum of the desolvation values associated with the SAS points of the fragment buried within the SAS of the receptor.*

The receptor desolvation upon binding is approximated by the sum of the values of the desolvation operated by the probe sphere over the SAS receptor points that are included within the SAS of the fragment (see Figure 3b):

$$\Delta G^{\text{receptor}}_{\text{desolv}} = \sum_{p \in \text{SAS}^{buried}_{receptor}} \Delta G^{p}_{\text{desolv}} \tag{17}$$

Since the adjacent positions of the sphere can partially overlap, the total receptor desolvation is scaled by a multiplicative factor [2]. The assumption underlying this model is that the main contribution to the receptor desolvation results from the removal of the first shell of water. This approximation is justified by the fact that the desolvation of a spherical ion by a small low dielectric sphere at a distance r from the ion varies as $\frac{1}{r^4}$.

Screened fragment-receptor interaction. The screened interaction between fragment and receptor ($E^{\text{interm}}_{\text{elect}}$ in kcal/mol) is calculated according to a linear distance-dependent dielectric model (first term of **p16**):

$$E^{\text{interm}}_{\text{elect}} = 332 \sum_{\substack{i \in fragment \\ j \in receptor}} \frac{q_i q_j}{\epsilon_{\text{int}} r^2_{ij}} \tag{18}$$

where q_i and q_j are the partial charges (in electronic units) of atoms i and j, r_{ij} is the distance between them (in Å), and ϵ_{int} is the value of the interior, i.e., solute, dielectric constant (**p1**). The calculation is done on a grid with trilinear interpolation. The grid margin and spacing are specified in **p16**.

Partial desolvation of the fragment. The desolvation of the fragment ($\Delta G^{\text{fragment}}_{\text{desolv}}$) is evaluated in a way that is specular to the receptor desolvation (see Figures 3a and 3b). First, the fragment desolvation due to a probe sphere rolling over the fragment SAS is calculated once for every fragment type. Subsequently, the fragment desolvation upon binding is approximated by the sum of the desolvation values associated with the points on the SAS of the fragment that are included within the SAS of the receptor. The same scaling factor as for the receptor desolvation is employed [2].

B.3 Docking scheme

In order to efficiently speed up the screening of fragments, SEED relies on a two-step docking workflow we will refer to as postprocess scheme (see Figure 4); each part of it is detailed below in the corresponding subsections.

In the first step, after the initial alignment of the fragment to a reference frame (prealignment of the fragment B.3.1), the generated poses are screened and filtered according to the fast energy model (fast docking scheme B.3.2). Poses passing all the filters are sorted according to their binding energy and clustered in order to find the most representative ones and discard redundant positions (clustering procedure B.3.3).

In the second step the best binding modes within each cluster are rescored and ranked with the more accurate solvation model (accurate rescoring B.3.4).

B.3.1 Pre-alignment of the fragment

Before starting the docking procedure the fragment is pre-aligned to a reference frame. The first atom of the fragment is put in the origin, the second is put along the positive x axis, the third in the xy plane and the others are aligned accordingly. If the third atom is collinear with the first two, the fourth atom is considered and so on. As the positioning of points on the Solvent Accessible Surface has a dependency on the absolute orientation of the fragment, the pre-alignment step is necessary to ensure consistency of results when docking (first line of **i7** set to `d`) the same fragment provided in different spatial orientations or when carrying out an energy evaluation (first line of **i7** set to `e`) of a fragment previously docked by SEED. Residual differences may be due to the limited precision of coordinates in usual mol2 files.

B.3.2 Fast docking scheme

The fast docking scheme proceeds by applying sequentially the following filters:

1. *Sphere for acceptance of the fragment position.* This filter is optional. The user can specify a sphere (coordinates of the center and radius in **i6**) in which the geometry center of the fragment position must be to be accepted.

2. *Van der Waals interaction used as bad contacts detection.* The fast van der Waals energy is used to detect clashes: a fragment is discarded if it is less favorable than a threshold value set in **p11**.

3. *Van der Waals interaction for apolar docking.* To increase efficiency apolar fragments are discarded without evaluation of the electrostatic contribution if the fast van der Waals interaction is less favorable than a threshold value. This value is calculated by multiplying **p19** by the number of fragment atoms.

4. *Total energy.* The fragment position is finally accepted if the total energy (fast model) is smaller than a cutoff given in the third column of **i7**. The total energy is the sum of the van der Waals interaction energy plus the electrostatic contribution (screened intermolecular energy and protein and fragment desolvation terms). These four terms can be weighted by the scaling factors in **p23**.

B.3.3 Clustering procedure

The docking of a given fragment (with energy evaluation as described above) is followed by sorting and clustering. Within each fragment type, positions are first sorted according to binding energy. Positions whose binding energy is lower than a user-specified threshold value (see filter 4. in B.3.2) are then clustered (a maximal number of positions can be set in **p27**) using as similarity criterion between two fragment positions A and B:

$$S(A,B) = \frac{S_{AB}}{\max(S_{AA}, S_{BB})} \quad \text{where} \quad S_{XY} = \sum_{i \in X} \sum_{j \in Y} w_{t_i t_j} \exp(-\gamma r_{ij}^2) \tag{19}$$

where r_{ij} is the distance between two atoms ($i \in$ fragment position X, $j \in Y$), $w_{t_i t_j}$ is a user-controlled matrix whose coefficients reflect the similarity between element types (in

most cases a unit matrix is used; otherwise the non-default coefficients have to be set in **p24** by giving the number of non-default values on the first line and two element types with the non-default value on the following lines) and γ is a coefficient which acts on the broadness of the distribution of the positions. B is assigned to the cluster of A if $S(A, B)$ is larger than a cutoff value δ with $0 \leq \delta \leq 1$. The clustering proceeds in two steps : (1) a first clustering with $\gamma = 0.9$ (first term of **p25**) and $\delta = 0.4$ (second term of **p25**) yields large clusters which contain almost overlapping as well as more distant fragments; (2) a second clustering with $\gamma = 0.9$ (first term of **p26**) and $\delta = 0.9$ (second term of **p26**) is done on each cluster found in (1) to eliminate fragments which are very close in space. The second clustering is applied on the positions for which the binding energy of the representative is smaller than a cutoff value specified in the 4th column of **i7**. A single clustering step with $\gamma = 0.9$ and $\delta = 0.9$ would generate too many small clusters. Hence, the first step is a real clustering while the second step is done only to discard redundant positions.

B.3.4 Accurate rescoring

The n best binding modes within each cluster (n is set in **p5**) are rescored and ranked according to the accurate solvation model. Note that it is possible that poses passing the total energy filter during fast docking result in a binding energy above the cutoff (set in the 3rd column of **i7**) with the accurate model. These poses appears in the docking summary in the `seed.out` log, but are not written to any other output files of the postprocess scheme (see B.5.1).

B.4 Energy evaluation mode

SEED allows the energy of a particular fragment position to be evaluated without using the docking mode. **i7** has to be set to `e` (energy evaluation mode) instead of `d` (docking mode). The fragment mol2 file must contain the coordinates of the relevant position, for which the binding energy has to be evaluated.

B.5 SEED output files

The grids for fast van der Waals energy, fast screened interaction energy and receptor desolvation can be saved on disk and reused for a subsequent run (**p7**,**p8**,**p9**). The path of the main output file of SEED is set in **i6**. The first term of **p28** is the maximal number of lines that can be written in the main output file for each docking step of each fragment type. The second term of **p28** gives control on which information may be discarded in the output file.

A directory `outputs` in which most of the output files are written is automatically created by the program.

`apolar_rec.mol2` and `apolar_rec_reduc.mol2` contain the apolar vectors before and after reduction (**p2**), respectively. `apolar_rec_reduc_angle.mol2` contains the vectors of `apolar_rec.mol2` which are selected if the angle criterion (**i4**, **p14**) is activated. The polar vectors are in `polar_rec.mol2`, `polar_rec_reduc_angle.mol2`, and `polar_rec_reduc.mol2`.

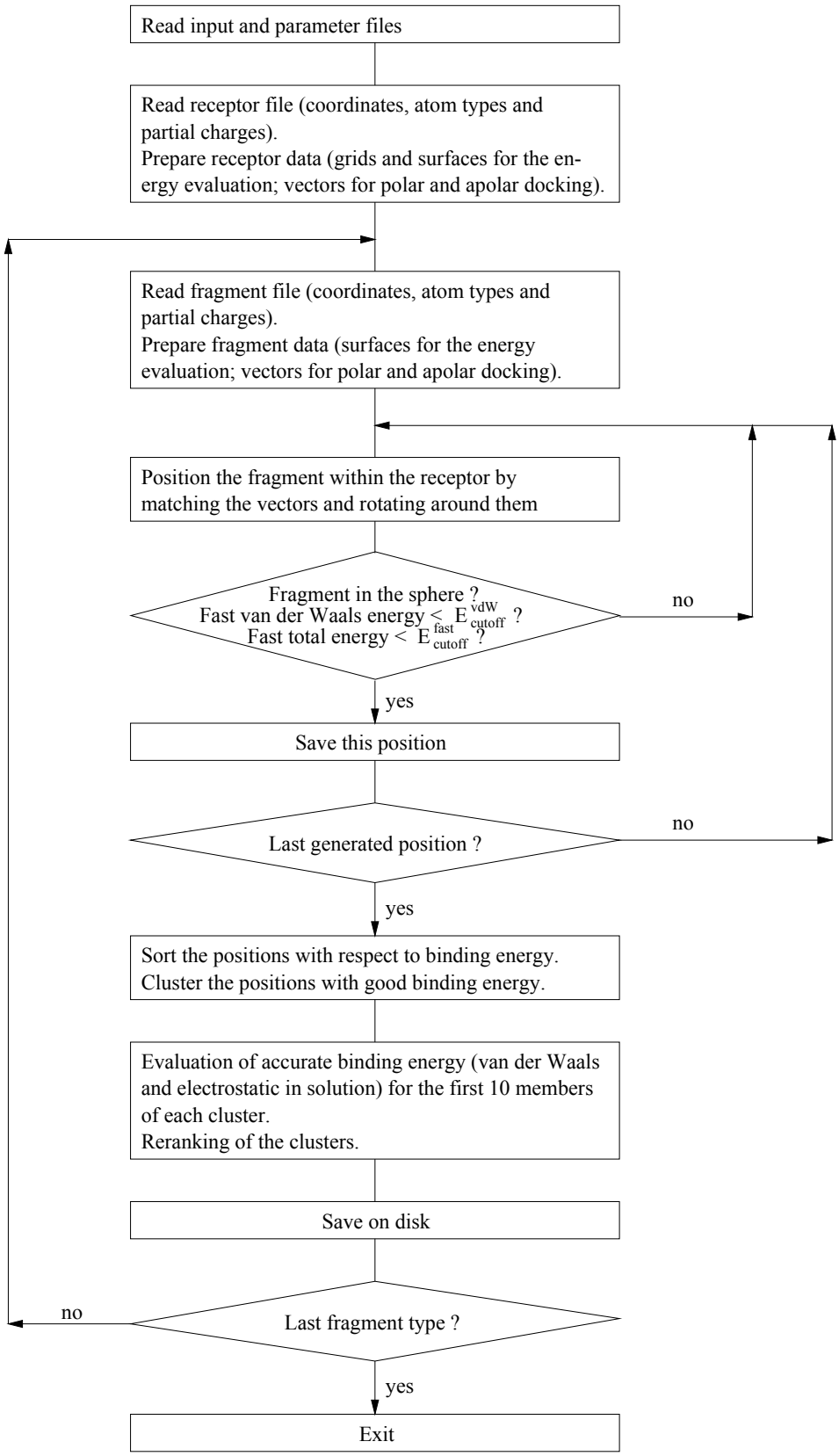

Figure 6: *Schematic representation of the SEED program (postprocess scheme).*

Out of the vectors that passed the angle reduction criterion, **p2** selects the percentage of them that should be used for docking (first value for polar and second for apolar vectors). The CPU time required by SEED is proportional to these numbers.

B.5.1 Postprocess scheme

<FragmentMol2FileName>_clus.mol2 contains the top poses per cluster ranked by accurate energy. It is a mol2 file and includes the hydrogen atoms. See A.5 for details.

<FragmentMol2FileName>_best.mol2 contains the top poses ranked by accurate energy. See A.5 for details.

seed_clus.dat is a summary table containing all the relevant energy terms corresponding to the poses saved in <FragmentMol2FileName>_clus.mol2. See A.5 for details.

seed_best.dat is a summary table containing all the relevant energy terms corresponding to the poses saved in <FragmentMol2FileName>_best.mol2. See A.5 for details.

The writing of the above *_clus.mol2 and *_best.mol2 file is activated or deactivated by **p3** (first and second value respectively). The writing of the *_clus.dat and *_best.dat summary table is activated or deactivated by **p4** (first and second value respectively). The maximum number of positions saved is specified in **p5**; the first value corresponds to the number of cluster members (*i. e.* number of poses per cluster) and the second to the total number of best poses, independently of the cluster they come from (see A.5 for details). Note that these are upper bounds as the number of generated poses may be smaller than the number of poses requested in output. All the four parameters for writing the output files (**p3** and **p4**) can be switched on/off independently.

References

[1] N. Majeux, M. Scarsi, J. Apostolakis, C. Ehrhardt, and A. Caflisch. Exhaustive docking of molecular fragments on protein binding sites with electrostatic solvation. *Proteins: Structure, Function and Genetics*, 37:88–105, 1999.

[2] N. Majeux, M. Scarsi, and A. Caflisch. Efficient electrostatic solvation model for protein-fragment docking. *Proteins: Structure, Function and Genetics*, 42:256–268, 2001.

[3] M. Scarsi, J. Apostolakis, and A. Caflisch. Continuum electrostatic energies of macromolecules in aqueous solutions. *J. Phys. Chem. A*, 101:8098–8106, 1997.

[4] M. Scarsi, J. Apostolakis, and A. Caflisch. Comparison of a GB solvation model with explicit solvent simulations: potentials of mean force and conformational preferences of alanine dipeptide and 1,2-dichloroethane. *J. Phys. Chem. B*, 102:3637–3641, 1998.

[5] M. S. Lee, F. R. Salsbury Jr., and C. L. Brooks III. Novel generalized Born methods. *J. Chem. Phys.*, 116(24):10606–10614, 2002.

[6] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in Fortran*. Cambridge University Press, 1992.

```
#seed.param v4.0
#
#      ------------------------------------------------------------------------
#          PARAMETERS FORMERLY IN SEED.INP
#          PARAMETERS FORMERLY IN SEED.INP
#          PARAMETERS FORMERLY IN SEED.INP
#      ------------------------------------------------------------------------
#           Dielectric constant of the solute (receptor and fragment)
p1   2.0
#           Ratio of kept vectors for docking : polar / apolar
p2   1.0 1.0
#           First value: write *_clus.mol2 file (y/n)
#           Second value: write *_best.mol2 file (y/n)
p3   n y
#          First value: write *_clus.dat summary table file (y/n)
#           Second value: write *_best.dat summary table file (y/n)
p4   y y
#           First value: maximum number of cluster members saved in *_clus*
#           output files. Note that this value determines the maximum number
#           of poses per cluster that go through slow energy evaluation.
#          Second value: maximum number of poses saved in *_best* output files.
p5   5 1
#           The docked fragments are saved in the dir ./outputs
#           Filename for output log file. This is the main SEED output file.
p6   ./outputs/seed.out
#           write (w) or read (r) Coulombic grid / grid filename
p7   w  ./scratch/coulomb.grid
#           write (w) or read (r) van der Waals grid / grid filename
p8   w  ./scratch/vanderwaals.grid
#           write (w) or read (r) receptor desolvation grid / grid filename
p9   w  ./scratch/desolvation.grid
#
#      ------------------------------------------------------------------------
#           PARAMETERS FOR DOCKING
#           DEFINITIONS:
#           fast energy: used only to preprocess, i.e., as a filter
#                        [Majeux et al, PROTEINS 42, 256-268, 2001]
#           accurate (and slower) energy: used to postprocess and should
#                                         NOT be turned off
#                        [Majeux et al, PROTEINS 37,  88-105, 1999]
#      ------------------------------------------------------------------------
#
#           Bump checking: used only if fast energy is switched off
#           n x atoms = maximum tolerated bumps /
#           scaling factor for interatomic distance /
#           severe overlap factor (beta factor in PROTEINS paper)
p10  2.0 0.89 0.6
#           van der Waals energy cutoff (kcal/mol):
#           used if fast energy is switched on
p11  1.0
#           Angle (deg) and number of points on the sphere around the HB vectors
p12  50.0 100
#           Number of fragment rotations around each axis
p13  72
#           Modification of February 2002:
#           Removal of rec. polar and apolar vectors using angle criterion
#           angle_rmin (deg)  angle_rmax   multipl_fact_rmin   multipl_fact_rmax
#           Method:
#           The minimal (minDist) and maximal (maxDist) distances
#           between the vectors and the points in the binding site
#           (as defined in the SEED input file) are evaluated.
#           A vector is discarded if the angle between the vector
#           and its closest point in the binding site is larger than
#           a cutoff angle value.
#           The cutoff angle value follows the following distribution:
#           -  angle_rmin  if distance <= (multipl_fact_rmin*minDist)
#           -  angle_rmax  if distance >= (multipl_fact_rmax*maxDist)
#           -  linear dependence (range between angle_rmin and angle_rmax)
#              for other distances
p14  70.0 10.0 1.2 0.8
```

*Seed input file (**seed.par**) 1/4.*

```
#           Removal of the rec. polar vect.: vdW radius probe
p15 1.83
#           Coulombic fast energy: 1=distance dept diel / grid margin / grid spacing
p16 1 20.0 0.5
#           van der Waals fast energy: grid margin / grid spacing
p17 20.0 0.3
#           accurate energy: Coulombic cutoff for formal charges is automatically
#                            set to 1.3 x van_der_Waals_cutoff
#           accurate energy (vdWaals): nonbonding cutoff / grid spacing
p18 12.0 1.0
#           Multiplicative factor (k) for apolar docking to skip evaluation of
#           electrostatics. The vdW energy cutoff is:
#           k x Number of fragment atoms, including hydrogen atoms
p19 -0.333333
#           Solvation grid: grid margin / grid spacing
p20 24.0 0.5
#           Solvation: water radius / # points per sphere to generate SAS /
#                      solvent dielectric constant
p21 1.4 500 78.5
#           Hydrophobicity maps: point densities (A^-2) on the SAS for apolar
#                    vectors on the receptor / on the fragment /
#                    probe radius to generate SAS for apolar vectors /
#                    scaling factor for desolvation and / vdW interactions
p22 1.0 1.0 1.4 1.0 1.0
#           Scaling factors for fast and also accurate energy evaluation:
#           van der Waals / electrostatic interaction / receptor desolvation /
#           fragment desolvation
p23 1.0 1.0 1.0 1.0
#      ------------------------------------------------------------------------
#           Clustering parameters
#           Clustering parameters
#           Clustering parameters
#      ------------------------------------------------------------------------
#           GSEAL : sim. weight factors (150 atom el.) 0 or # non-default + list
p24 4
6  6  2.0
7  7 10.0
8  8 10.0
16 16 10.0
#           The clustering with GSEAL proceeds in two steps: the
#           first clustering yields large clusters which contain almost
#           overlapping as well as more distant fragments; the second
#           clustering is done on each cluster found in the first clustering
#           to eliminate fragments which are very close in space.
#
#           First clustering:  overall clustering
#           GSEAL similarity exponential factor / cutoff factor
p25 0.9 0.4
#
#           Second clustering: to discard redundant positions
#           GSEAL similarity exponential factor / cutoff factor
p26 0.9 0.9
#           Maximal number of positions to be clustered
p27 20
#           Number of lines to be written in the output file for the sorted
#           energies and the two clustering procedures /
#           Printlevel (0=lean, 1=adds sorting before postprocessing, 2=adds 2nd
#           clustering)
p28 100 1
#      ------------------------------------------------------------------------
#           Forcefield parameters from now on
#           Forcefield parameters from now on
#           Forcefield parameters from now on
#      ------------------------------------------------------------------------
# Parameters are CGenFF v4 (2016) and param36 (2012) with water and ions as per
# toppar_water_ions.str
#
# NB: if the user includes metal ions in the binding site, appropriate hydrogen bond
# distances with these ions should be provided in the "Hydrogen bond distances
# between donor and acceptor" section below
```

*Seed input file (**seed.par**) 2/4.*

```
#
# Hydrogen bond distances were specified for charge-charge and charge-neutral
# hydrogen bonding pairs according to the following:
#
# (+) charge H-bond donor of protein[4]: NR3, NC2, NP, NH3
# neutral H-bond donor of protein [7]: NR1, NH1, NH2, NY, OH1, S, OT [water]
# neutral H-bond acceptor of protein [8]: OH1, S, O, OB, OS, NR2, OT [water], OX
# (-) charge H-bond acceptor of protein [1]: OC
# (+) charge H-bond donor of ligand [5]: NG2P1, NG2R52, NG3P1, NG3P2, NG3P3
# neutral H-bond donor of ligand [13]: NG2S1, NG2S2, NG2S3, NG2R43, NG2R51, NG2R53,
# NG2R61, NG311, NG321, NG331, NG3C51, OG311, SG311
# neutral H-bond acceptor of ligand [29]: NG1T1, NG2D1, NG2S0, NG2R50, NG2R57,
# NG2R60, NG2R62, NG2R67, NG2RC0,  NG301, NG3N1, OG2D1, OG2D3, OG2D4, OG2D5, OG2R50,
# OG3R60, OG301, OG302, OG303, OG304, OG311, OG3C31, OG3C51, OG3C61, SG2D1, SG2R50,
# SG311, SG301
# (-) charge H-bond acceptor of ligand [5]: OG2D2, OG2N1, OG2P1, OG312, SG302
# Polar atoms, none of above categories NG3P0, NG2O1, SG3O2, SG3O1, SG3O3
#
#     Atom  element  van der Waals
#     type  number  radius energy_min (absolute value)
#   -----------------------------------
p29 225
1      H      1     0.22450 0.046000  # polar H
2      HC     1     0.22450 0.046000  # N-ter H
3      HA     1     1.32000 0.022000  # nonpolar H
4      HP     1     1.35820 0.030000  # aromatic H
5      HB1    1     1.32000 0.022000  # backbone H
6      HR1    1     0.90000 0.046000  # his he1, (+) his HG,HD2
7      HR2    1     0.70000 0.046000  # (+) his HE1
8      HR3    1     1.46800 0.007800  # neutral his HG, HD2
9      HS     1     0.45000 0.100000  # thiol hydrogen
10     HE1    1     1.25000 0.031000  # for alkene; RHC=CR
...

...
220    ZN     30    1.0900  0.2500    # zinc (II) cation
221    CAD    48    1.3570  0.1200    # cadmium (II) cation
222    CLA    17    2.27    0.150     # Chloride Ion
223    CT     6     2.275   0.020     # TIP3P
224    OT     8     1.7682  0.1521    # TIP3P
225    HT     1     0.2245  0.0460    # TIP3P
#
# Hydrogen bond distances between donor and acceptor
#
# The table cannot be separated
#
# First  line: Default distance for all atom and element types
#
# First block:
# element  donor
# number   acceptor
# i  j     distance
# ------------------
#
# Second block:
# atom     donor
# type     acceptor
# i  j     distance
# ------------------
#
#
#
p30 2.9
9
7    7    3.1
7    8    2.9
7   16    2.9
8    8    2.7
8   16    2.9
16  16    2.9
```

*Seed input file (**seed.par**) 3/4.*

```
30   7    2.1
30   8    2.05
30  16    2.36
507
33  42  2.6  #  charged donor and charged acceptor
33  176  2.6  #  charged donor and charged acceptor
33  180  2.6  #  charged donor and charged acceptor
33  181  2.6  #  charged donor and charged acceptor
33  188  2.6  #  charged donor and charged acceptor
33  199  2.6  #  charged donor and charged acceptor
...

...
200  42  2.75  # non charged donor and charged acceptor
200  176  2.75  # non charged donor and charged acceptor
200  180  2.75  # non charged donor and charged acceptor
200  181  2.75  # non charged donor and charged acceptor
200  188  2.75  # non charged donor and charged acceptor
200  199  2.75  # non charged donor and charged acceptor
#
# List of relative atomic weights
#
# First line: number of elements (without element 0)
# element_name element_number atomic_weight
#
118
LPH  0       0.000   # Lone Pair Hole
H    1       1.008   # Hydrogen
He   2       4.002   # Helium
Li   3        6.94   # Lithium
Be   4       9.012   # Beryllium
B    5       10.81   # Boron
C    6      12.011   # Carbon
N    7      14.007   # Nitrogen
O    8      15.999   # Oxygen
F    9      18.998   # Fluorine
Ne  10    20.17976   # Neon
...

...
Fl 114         289   # Flerovium
Mc 115         289   # Moscovium
Lv 116         293   # Livermorium
Ts 117         293   # Tennessine
Og 118         294   # Oganesson
#
end
```

p31

*Seed parameter file (**seed.par**) 4/4.*