Manual for SEED

a program for docking molecular fragments into a rigid protein

SEED = Solvation Energy for Exhaustive Docking

SEED developers

1998 - 2003: Nicolas Majeux, Marco Scarsi, Fabian Dey,
Claus Ehrhardt and Amedeo Caflisch

2011 - 2013: Fabian Dey, Tim Knehans, Emilie Frugier, and Amedeo Caflisch

Manual for Version 3.3.5

November 2013

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, our numerical continuum electrostatic methodology [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 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 29-31) files are indicated by **i** and **p**, respectively.

The path for the parameter file is in the first line of the input file (referred to as **i1**).

- A Sybyl mol2 format file (**i9**) for the receptor with partial charges on the 9th column in the field `@<TRIPOS>ATOM`.

- A Sybyl mol2 format file for each fragment to dock, with partial charges as for the receptor and with a four-letter fragment name `xxxx` (`# RESN xxxx`) at the beginning of the file. If the keyword `RESN` is not found a fragment name is automatically assigned. A fragment file can contain several fragment conformations. The integer in the first line of **i13** is the number of fragment files and the first term of the following lines is the path of the mol2 file.

A.4 Most frequently modified parameters

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

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

i2 Dielectric constant of the solute (receptor and fragment), usually between 1.0 and 4.0.

i3 Percentage of polar and apolar vectors which passed the angle reduction criterion that should be used for docking. The CPU time required by SEED is proportional to these numbers.

i4 Number of cluster members saved in output files and postprocessed with the more accurate solvation model (see B.3.1).

i9 Receptor coordinate file (Sybyl mol2 format).

i10 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.

i14 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 (**p26**, see B.1.3) such that vectors pointing outside of the binding site are discarded. The points can be for example the ligand heavy atoms of a known ligand-receptor complex structure.

i12 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`.

i13 First line: the integer is the number of fragment files. The second term can be set to `y` (energy evaluation mode; see B.5) or to `n` (docking mode; see B.3). Following lines: the first column contains the path of the mol2 file and the second column allows the selection of apolar, polar docking or both. The fragment position is accepted if the total energy is smaller than a cutoff given in the third column. The second clustering (see page 22) 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 output file, whose filename is specified in **i5**, contains the energy values and results of clustering.

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

`FragmentName_clus_pproc.mol2` contains the fragment positions with best energy after the postprocessing step. In the `@<TRIPOS>SUBSTRUCTURE` section the cluster numbering and binding energy are stored in the third and second last columns.

A.6 Starting a new project

When a new project is started, it is useful to first generate the vectors without docking any fragment (**i13**: `0 n`). 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 (**i6-i8**: `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 **p5**.

- `polar_rec_reduc_angle.mol2` contains vectors of `polar_rec.mol2` which are selected according to an angle criterion (**i14**, **p26**). 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 (**i14**).

- `polar_rec_reduc.mol2` contains vectors of `polar_rec.mol2` (or of `polar_rec_reduc_angle.mol2` if the angle criterion has been activated (**i14**)) 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 **p7**. The number of selected vectors is controlled with **i3**.

- `apolar_rec.mol2` contains points distributed uniformly on the solvent-accessible surface of the receptor. The density of surface points is set in **p18**.

- `apolar_rec_reduc_angle.mol2` contains vectors of `apolar_rec.mol2` which are selected according to an angle criterion (**i14**, **p26**). Vectors pointing out-

side 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 (**i14**).

- `apolar_rec_reduc.mol2` contains points of `apolar_rec.mol2` (or of `apolar_rec_reduc_angle.mol2` if the angle criterion has been activated (**i14**)) 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 **p18**. The number of selected apolar points is controlled with **i3**.

A.7 What to do if output empty

If after starting a SEED run the program exits unexpectedly, the keyword `WARNING` should be looked for in the main output file (**i5**) 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 (**i12**) might be misplaced (outside the binding site), the checking of clashes (**p3**, see B.3.3; **p4**, see B.3.2 and B.3.1) too strict, the van der Waals energy cutoff (**p11**) for apolar fragments too severe, the total energy cutoff (third column of **i13**) 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
```

Input file

```
#           Parameter filename
i1  ./inputs/seed.par
#           Dielectric constant of the solute (receptor and fragment)
i2  2.0
#           Percentage of vectors for docking : polar / apolar
i3  0.8 0.8
#           Number of cluster members saved in output files
i4  10
#           The docked fragments are saved in the dir ./outputs
#           Filename for output log file
i5  ./outputs/seed.out
#           write (w) or read (r) Coulombic grid / grid filename
i6  w  ./scratch/coulombic_20residues.grid
#           write (w) or read (r) van der Waals grid / grid filename
i7  w  ./scratch/vdwaals_20residues.grid
#           write (w) or read (r) receptor desolvation grid / grid filename
i8  w  ./scratch/receptor_desolv_20residues.grid
#           Receptor coordinates (in mol2 format) filename
i9  ./inputs/thrombin.mol2
#           Binding site residue list
#           First line:  number of residues
i10 20
    27
    28
    43
    44
    47
    95
    96
    173
    193
    194
    195
    199
    219
    220
    221
    222
    223
    224
    225
    232
#           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
#           An example with 3 available points is commented out below
i14 0
#    -4.553   -2.308   -22.326
#     0.427   -1.012   -25.607
#    -1.388   -4.127   -28.99
#           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
i11 0
#           Spherical cutoff for docking (y,n / sphere center / sphere radius)
i12 n   -2.133 -1.359 -25.539   10.0
#           Fragment library specifications
#           First line: Number of fragments / dock+energy (n), only energy (y)
#           Following lines: Fragment filename /
#           apolar docking, polar docking, or both (a,p,b) /
#           energy cutoff in kcal/mol / 2nd clustering cutoff in kcal/mol
i13 3 n
./inputs/benzene.mol2           a     5.0   5.0
./inputs/phenole.mol2           p     5.0   5.0
./inputs/dhthfurn.mol2          p     5.0   5.0
end
```

B Carrying on

B.1 Vectors for docking

The binding site where the fragments have to be docked is defined by giving the list of the residues. The first line of **i10** 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 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 **p1**: 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 **i11** 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 **p5** is the maximal angular deviation from ideal hydrogen bond geometry and the second term is the number

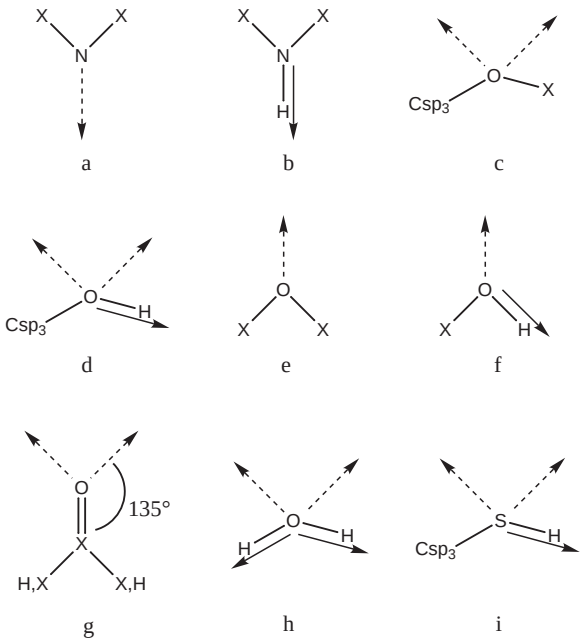

Figure 1: *Description of polar vectors for the fragment and for the receptor. X is a heavy atom. The broken arrow represents a vector of H-bond acceptor in the lone pair direction and the full arrow a vector of H-bond donor. The geometry of c, d, h and i is tetrahedral (angle of* 109°*). Examples: (a) 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.*

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 **p7** and those of the atom types are specified in the 4th column of **p1**. 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 **i3**. 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 **p2** (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 **p6**.

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 **p18**. 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 **p18**). 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 **p18**: 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 **p18**. The number of selected apolar points can be modified with the second term of **i3**. 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 **p6**.

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 (**i14**, **p26**).

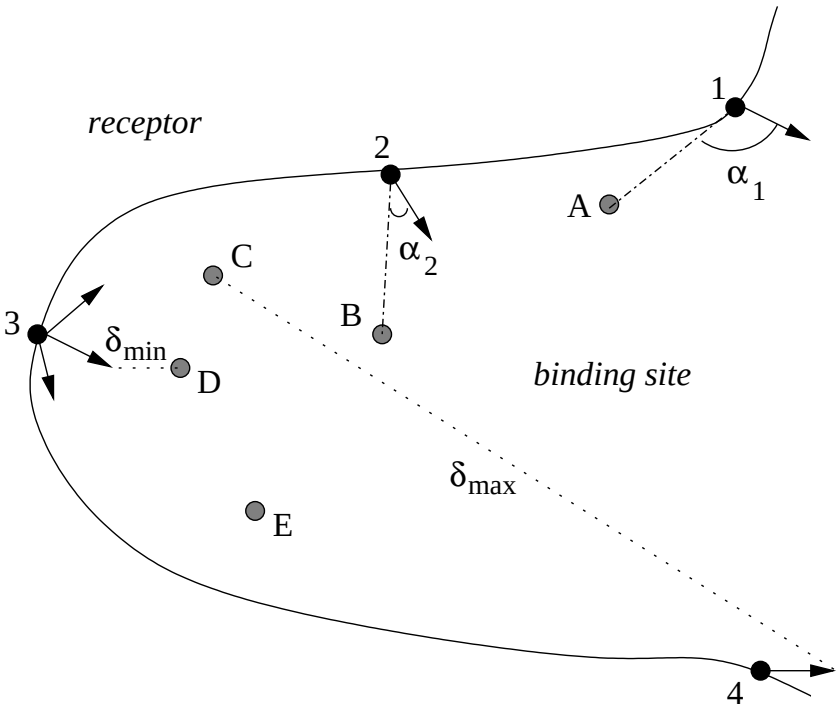

Figure 2: *1-4: receptor atoms and vectors. A-E: user-defined anchor points in the binding site (e.g., ligand heavy atoms). The angle between a vector and its closest anchor point in the binding site is shown for two vectors (α_1, α_2). Reasonable parameters should allow to remove the vector of atom 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 **i14** 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 ligand heavy atoms obtained from a known ligand-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 $\mathbf{p26}_1$ (first parameter in **p26**) if the distance between the vector and the closest anchor point is smaller or equal to $\delta_{min} \times \mathbf{p26}_3$; the angle cutoff is $\mathbf{p26}_2$ if the distance is larger or equal to $\delta_{max} \times \mathbf{p26}_4$. For other distances the angle cutoff value falls between $\mathbf{p26}_1$ and $\mathbf{p26}_2$ (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 **i13** allows the user to select apolar docking, polar docking or both.

B.2 Energy in SEED

Two energy models are implemented in SEED: they can be evaluated independently or combined in a two-step procedure (see B.3). In both cases the binding energy is 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 **i2** and dielectric constant of the solvent as the 3rd term of **p15**). In this approximation only the intra-solute electrostatic interactions need to be evaluated. This strongly reduces the number of interactions with respect to an explicit treatment of the solvent. The difference in electrostatic energy in solution upon binding of a fragment to a receptor can be calculated as the sum of the following three terms:

- Partial desolvation of the receptor: electrostatic energy difference upon binding an uncharged fragment to a charged receptor in solution.
- Screened fragment-receptor interaction: intermolecular electrostatic energy in solution.
- Partial desolvation of the fragment: electrostatic energy difference upon binding a charged fragment to an uncharged receptor in solution.

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 **p10**). 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 **p10**) 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

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

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 **p1**.

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:

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

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

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

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 **p15** 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:

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

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

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

Two approaches (**p16**) are possible to calculate the receptor electric displacement over a 3D grid (grid margin and spacing set in **p14**). They both fulfill the condition of validity of equations 4 and 5 and have been implemented in SEED. In the first the electric displacement of every charge of the receptor can be represented 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. The finite difference calculation of $\vec{D}$ for the isolated receptor (the program UHBD is employed here; **p17**) is performed only once at the beginning of the program run.

The receptor desolvation is computed from equation 5 together with equation 6 (Coulomb field approximation) or 7 (finite difference approximation).

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 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 %; **p10**) 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 can be activated (**p27**) 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 **p1**). 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 **p9**). 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 3a) is evaluated according to the Coulomb approximation of the electric displacement:

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

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 $\epsilon_{\text{int}} < \epsilon_{\text{w}}$.

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):

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

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 $\frac{1}{r^4}$.

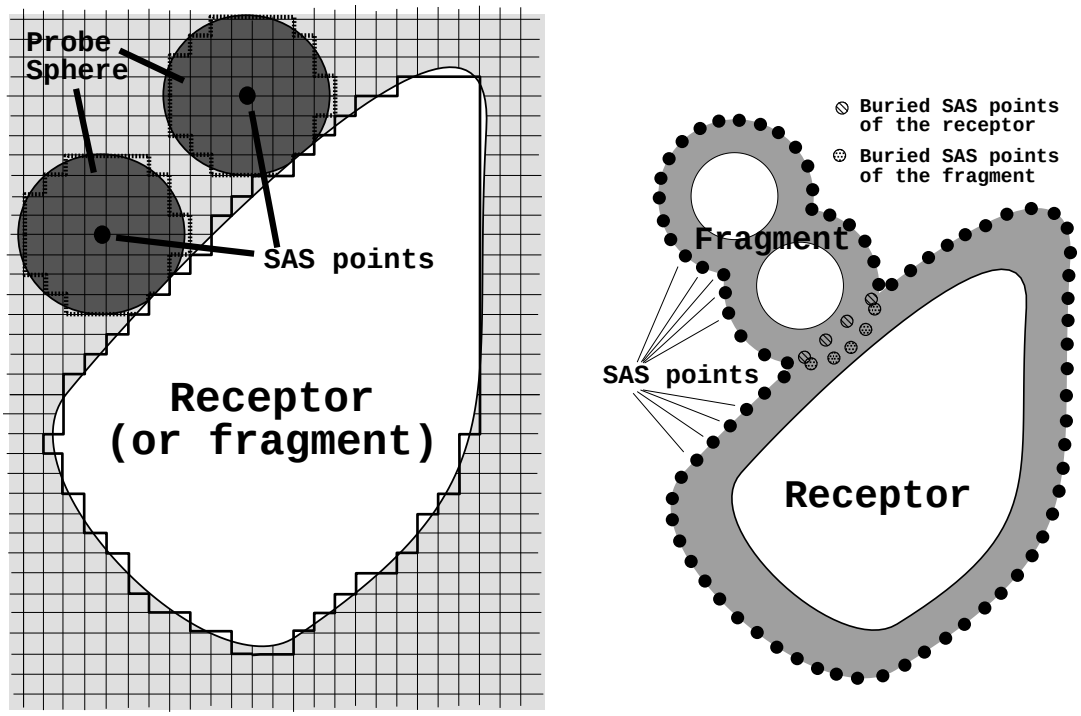

Figure 3: *(a) Evaluation of the receptor (fragment) desolvation due to a probe sphere rolling over the van der Waals surface of the receptor (fragment). A grid box of 0.5 Å spacing (light gray) is built around the molecule and the desolvation resulting from the occupation of every grid element is evaluated 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.*

Screened fragment-receptor interaction. The screened interaction between fragment and receptor ($E_{\text{elect}}^{\text{interm}}$ in kcal/mol) is calculated according to a linear distance-dependent dielectric model (first term of **p8**):

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

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 (**i2**). The calculation is done on a grid with trilinear interpolation. The grid margin and spacing are specified in **p8**.

Partial desolvation of the fragment. The desolvation of the fragment ($\Delta G_{\text{desolv}}^{\text{fragment}}$) 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 schemes

There are three docking schemes that the user can select in **p13**: a relatively slow scheme using the accurate energy, a faster scheme using the fast energy model and a two-step procedure called postprocess scheme. It is most convenient and appropriate to use the postprocess scheme in most projects. If the hardware resources allow for it, one can use directly the accurate docking scheme (i.e., skip the evaluation of the fast energy). The fast docking scheme alone should not be used because its energy evaluation is very approximative.

B.3.1 Postprocess scheme

An efficient way of speeding up the fragment screening by SEED is to divide the selection of favorable binding modes into two different steps (see Figure 4). In

the first step, the fast docking scheme is used (see B.3.2 below). The positions are then sorted according to binding energy and clustered (see below **Clustering procedure**). In the second step, the n best binding modes (n is set in **i4**) within each cluster are evaluated with the more accurate solvation model (see B.2.1).

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 **i12**) 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 **p4**.

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 **p11** 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 **i13**. 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 **p19**.

B.3.3 Accurate docking scheme

The accurate 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 **i12**) in which the geometry center of the fragment position must be to be accepted.

2. *Bad contacts detection.* A grid is defined over the whole receptor. This is followed by the generation of a list which contains for every cubic grid element the atoms which are enclosed by the surface of the cubic element. This list is calculated only once at the beginning of the program and is used to check bad contacts. A bad contact is defined as $r_{ij} < \alpha(R_i^{\mathrm{vdW}} + R_j^{\mathrm{vdW}})$ and a severe overlap is defined as $r_{ij} < \alpha\beta(R_i^{\mathrm{vdW}} + R_j^{\mathrm{vdW}})$, where r_{ij} is the distance between two atoms ($i \in$ fragment, $j \in$ receptor), and R_i^{vdW} and R_j^{vdW} are the corresponding van der Waals radii. The value of α is usually set to $2^{-1/6} = 0.89$ (second term of **p3**), the distance at which repulsion balances attraction, and a β of 0.6 (third term of **p3**) is employed normally. The distances between an atom of the fragment and the receptor atoms in the 27 grid cubes surrounding the fragment atom are calculated. This is repeated for every atom in the fragment. A grid spacing of twice the largest van der Waals radius is used to avoid missing clashes. For polar fragments the hydrogen atom involved in the H-bond is not checked for bad contacts. Fragments with one severe overlap or a number of bad contacts larger than a cutoff are discarded. This cutoff is the number of fragment atoms multiplied by the number given in the first term of **p3**.

3. *Van der Waals interaction for apolar docking.* To increase efficiency apolar fragments are discarded without evaluation of the electrostatic contribution if the van der Waals interaction (accurate model) is less favorable than a threshold value. This value is calculated by multiplying **p11** by the number of fragment atoms.

4. *Total energy.* The fragment position is finally accepted if the total energy (accurate model) is smaller than a cutoff given in the third column of **i13**. 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 **p19**.

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 are then clustered (a maximal number of positions can be set in **p23**) using as similarity criterion between two fragment positions A and B:

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

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 **p20** 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 **p21**) and $\delta = 0.4$ (second term of **p21**) yields large clusters which contain almost overlapping as well as more distant fragments; (2) a second clustering with $\gamma = 0.9$ (first term of **p22**) and $\delta = 0.9$ (second term of **p22**) 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 **i13**. 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.

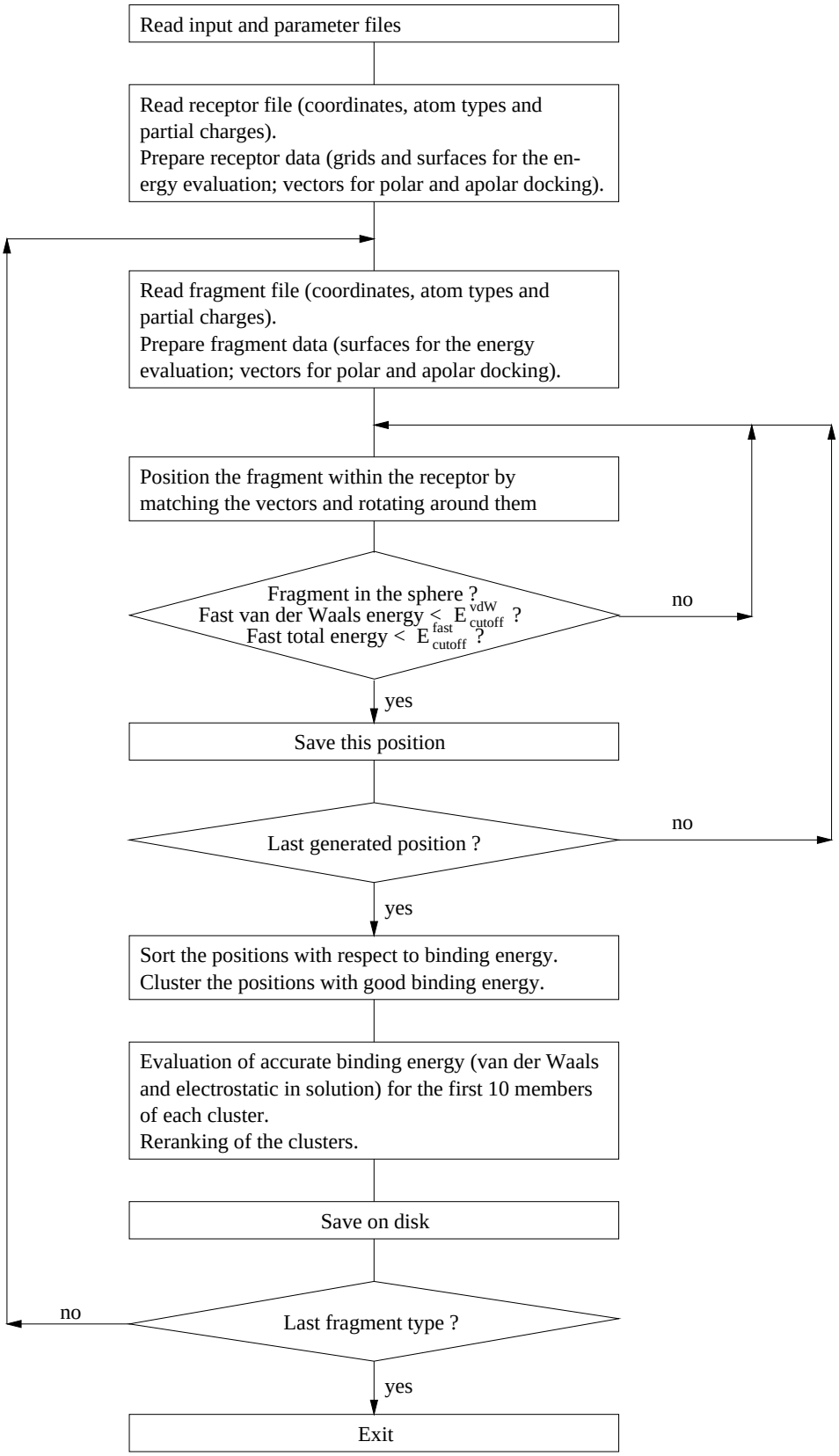

Figure 4: *Schematic representation of the SEED program (postprocess scheme).*

B.4 Geometrical centers for FFLD

The geometrical centers that are used in FFLD [7] to guide the docking of flexible ligands are built as follows [8]. Cluster representatives resulting from the postprocess scheme are subsequently grouped according to the coordinates of their geometrical centers using a threshold of **p28** (fourth term) Å. The geometrical centers of the first **p28** (third term) cluster representatives are removed and directly used for docking. The remaining geometrical centers are again grouped according to the same threshold. For each cluster an average geometrical center ($\mathbf{r}_{AGC}$) is calculated with the following procedure. First, all the positions with a binding energy greater than **p28** (fifth term) kcal/mol with respect to the cluster representative are discarded. Then, the $\mathbf{r}_{AGC}$ of a given cluster is evaluated as an energy weighted mean

$$\mathbf{r}_{AGC} = \sum_{i=1}^{N} \omega_i \mathbf{r}_i \tag{20}$$

$$\omega_i = \begin{cases} E_i \;/\; \sum_i E_i & \text{if} \quad E_{max} < 0 \\ (E_i - (E_{min} + E_{max})\,) \;/\; \sum_i (E_i - (E_{min} + E_{max})\,) & \text{if} \quad E_{min} > 0 \end{cases} \tag{21}$$

where E_{min} and E_{max} are the minimum and maximum energy within a cluster, respectively, and the sum runs over the N geometrical centers $\mathbf{r}_i$ of the fragments in the cluster. In the sporadic case of $E_{min} > 0$, by subtracting $(E_{min} + E_{max})$ from the fragment binding energies E_i, it is possible to give more weight to positions with small absolute energy values. For $E_{min} \leq 0 \leq E_{max}$, average centers for the subsets of favorable ($E_i < 0$) and unfavorable ($E_i > 0$) binding modes in a cluster ($\mathbf{r}_-$ and $\mathbf{r}_+$, respectively) are computed by equation 20 and the $\mathbf{r}_{AGC}$ is evaluated as

$$\mathbf{r}_{AGC} = \alpha_- \, \mathbf{r}_- + \alpha_+ \, \mathbf{r}_+ \tag{22}$$

where α_- and α_+ (sixth and seventh terms of **p28**, respectively) are multiplicative factors. The final list of geometrical centers used for docking is a compromise between the accuracy of the SEED binding energy (geometrical centers of **p28** (third term) cluster representatives optimally docked by SEED) and the diversity derived from the clustering procedure (average geometrical centers). The post-processing

of the optimal binding modes of the fragments leads to more heterogeneous binding site maps than using only the most favorable ones.

This post-processing is activated or deactivated by the first term of **p28**.

B.5 Energy evaluation mode

SEED allows the energy of a particular fragment position to be evaluated without using the docking mode. The second term of **i13** has to be set to `y` (energy evaluation mode) instead of `n` (docking mode). The fragment mol2 file must contain the coordinates of the relevant position.

B.6 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 (**i6**,**i7**,**i8**). The path of the main output file of SEED is set in **i5**. The first term of **p24** 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 **p24** 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 (**i3**), respectively. `apolar_rec_reduc_angle.mol2` contains the vectors of `apolar_rec.mol2` which are selected if the angle criterion (**i14**, **p26**) is activated. The polar vectors are in `polar_rec.mol2`, `polar_rec_reduc_angle.mol2`, and `polar_rec_reduc.mol2`.

B.6.1 Postprocess scheme

`FragmentName_clus_pproc.mol2` contains the top poses ranked by accurate energy. It is a mol2 file and includes the hydrogen atoms. In the `@<TRIPOS>SUBSTRUCTURE`

section the cluster numbering and total accurate energies are stored in the third and second last columns (for the computer programs FFLD [7] and SLD, unpublished).

`FragmentName_clus_pprocr.chm` contains the representatives of the clusters, i.e., the top pose (ranked by accurate energy) within each cluster. (It is a CHARMM-format file for visualization with the molecular modeling program WITNOTP; A. Widmer, unpublished).

`FragmentName_clus_pproc.chm` is the same as `FragmentName_clus_pproc.mol2` but in CHARMM format. (Only the coordinates of the heavy atoms are included for CCLD [9]).

`FragmentName_geomcent.mol2` contains the geometrical centers for FFLD (see B.4). The right-hand column contains their corresponding energies. The first **p28** (third term) geometrical centers are characterized by zinc atoms, the remaining geometrical centers by chlorine atoms. The second term of **p28** is the maximal number of geometrical centers to write in this file.

The writing of the above `*_pproc*` files is activated or deactivated by **p25**.

B.6.2 Accurate docking scheme

`FragmentName.chm` contains the top 999 poses ranked by accurate energy.

`FragmentName_clus.chm` contains the top 999 positions before clustering. Only the coordinates of the heavy atoms are included (for the computer program CCLD).

`FragmentName_clus_reduc.chm` contains the fragment positions of the top clusters up to 999 positions with a limited number (**i4**) of positions per cluster. Only the coordinates of the heavy atoms are included (for CCLD).

`FragmentName_xxx.mol2` are xxx files, each with a single pose where xxx=1 is the top pose according to accurate energy, xxx=2 is the second best pose, etc. The total number of this type of file is limited to **p12** (which is usually set to 1).

B.6.3 Fast docking scheme

It is used in the postprocessing scheme. Note that the fast docking scheme should not be run alone except for debugging purposes. `FragmentName.chm`, `FragmentName_clus.chm`, `FragmentName_clus_reduc.chm` and `FragmentName_xxx.mol2` are the same as for the accurate docking scheme (see B.6.2) but with the fast energy.

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.

[7] N. Budin, N. Majeux, and A. Caflisch. Fragment-based flexible ligand docking by evolutionary optimization. *Biol. Chem.*, 382:1365–1372, 2001.

[8] M. Cecchini, P. Kolb, N. Majeux, and A. Caflisch. Automated docking of highly flexible ligands by genetic algorithms: A critical assessment. *J. Comput. Chem.*, 25:412–422, 2004.

[9] A. Caflisch. Computational combinatorial ligand design: Application to human α-thrombin. *J. Comput.-Aided Mol. Design*, 10:372–396, 1996.

Parameter file (1/3)

```
# parameters are cgenff and param36
#
#     Atom  element  van der Waals
#     type  number  radius energy_min
#   ------------------------------------
p1  240
1      H      1     0.224   0.0460
2      HC     1     0.224   0.0460
3      HA     1     1.320   0.0220
4      HT     1     0.224   0.0460
5      HP     1     1.358   0.0300
6      HB1    1     1.320   0.0220
7      HR1    1     0.900   0.0460
8      HR2    1     0.700   0.0460
9      HR3    1     1.468   0.0078
10     HS     1     0.450   0.1000
...
236    HB2    1     1.340   0.0280
237    HA1    1     1.340   0.0450
238    HA2    1     1.340   0.0340
239    HA3    1     1.340   0.0240
240    CAI    6     1.990   0.0730
#
# 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
# ------------------
#
#
#
p2  2.9
9
7    7    3.10
7    8    2.90
7   16    2.90
8    8    2.70
8   16    2.90
16  16    2.90
30   7    2.10
30   8    2.05
30  16    2.36
279
49 195    2.60
53 195    2.60
55 195    2.60
52 195    2.60
90 195    2.60
49 199    2.60
53 199    2.60
55 199    2.60
52 199    2.60
90 199    2.60
49 208    2.60
53 208    2.60
55 208    2.60
52 208    2.60
90 208    2.60
49 216    2.60
53 216    2.60
55 216    2.60
52 216    2.60
90 216    2.60
49 200    2.60
53 200    2.60
55 200    2.60
52 200    2.60
90 200    2.60
49 168    2.75
...
```

Parameter file (2/3)

```
...
177 68    2.75
211 60    2.75
211 68    2.75
180 60    2.75
180 68    2.75
187 60    2.75
187 68    2.75
188 60    2.75
188 68    2.75
189 60    2.75
189 68    2.75
185 60    2.75
185 68    2.75
174 60    2.75
174 68    2.75
207 60    2.75
207 68    2.75
217 60    2.75
217 68    2.75
#
#      ------------------------------------------------------------------------
#           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)
p3  2.0 0.89 0.6
#           van der Waals energy cutoff (kcal/mol):
#           used if fast energy is switched on
p4  1.0
#           Angle (deg) and number of points on the sphere around the HB vectors
p5  50.0 100
#           Number of fragment rotations around each axis
p6  36
#           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
p26 70.0 10.0 1.2 0.8
#           Removal of the rec. polar vect.: vdW radius probe
p7  1.83
#           Coulombic fast energy: 1=distance dept diel / grid margin / grid spacing
p8  1 10.0 0.5
#           van der Waals fast energy: grid margin / grid spacing
p9  10.0 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
p10 10.0 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
p11 -0.333333
#           Maximal number of mol2 files to be written for each fragment type
p12 1
#           Energy evaluation (f_ast,s_low,p_ostprocess=f+s)
p13 p
#           Modification of March 2003:
#           Empirical correction term (y,n) to the Coulomb field approximation
#           [eq. 8 in Lee et al, Journal of Chemical Physics 116 (24), 10606, 2002]
#           for the accurate screened interaction and fragment desolvation energies
p27 y
#           Solvation grid: grid margin / grid spacing
p14 10.0 0.5
```

Parameter file (3/3)

```
#           Solvation: water radius / # points per sphere to generate SAS /
#                      solvent dielectric constant
p15 1.4 500 78.5
#           Solvation: Algorithm used to calculate the receptor desolvation
#                      (Coul approx. = co; UHBD finite diff. = fd) /
p16 co
#           Solvation: UHBD executable /
#                      directory where to write UHBD output
p17 /prog/uhbd-5.1MSI/bin.sgi/uhbd_180   ./UHBD_output/
#           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
p18 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
p19 1.0 1.0 1.0 1.0
#           GSEAL : sim. weight factors (150 atom el.) 0 or # non-default + list
p20 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
p21 0.9 0.4
#
#           Second clustering: to discard redundant positions
#           GSEAL similarity exponential factor / cutoff factor
p22 0.9 0.9
#           Maximal number of positions to be clustered
p23 15000
#           Geometrical centers for FFLD: (added in June 2004)
#           activation or deactivation of this option (y/n) /
#           maximal number of geometrical centers in *_geomcent.mol2 files /
#           number of cluster representatives to select without averaging /
#           distance threshold (in A) for grouping the cluster representatives /
#           maximal energy difference (in kcal/mol) allowed in a cluster with
#           respect to its most favorable energy /
#           multiplicative factors for weighing cluster positions with
#           negative and positive energies, respectively
p28 y 20 5 1.0 3.0 0.8 0.2
#           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)
p24 1000 1
#           Write *_pproc* files in postprocess mode (y/n)
p25 y
end
```

S

```
                          Change-log file
------------------------------------------------------------------------

SEED 1.0   Release                  1999

Version of SEED corresponding to

N Majeux, M Scarsi, J Apostolakis, C Ehrhardt, A Caflisch,
Proteins: Structure, Function and Genetics, 37, pp.  88-105, 1999

------------------------------------------------------------------------

SEED 2.0   Release     September 25, 2000

Version of SEED corresponding to

N Majeux, M Scarsi, A Caflisch,
Proteins: Structure, Function and Genetics, 42, pp. 256-268, 2001

------------------------------------------------------------------------

SEED 2.1   Release     February  25, 2002

The following is introduced after the SEED 2.0 release dated
September 25, 2000.

1. Enhancement of the receptor vector selection
   Name          : Nicolas Majeux
   Email address : majeux@bioc.unizh.ch
   Institution   : Department of Biochemistry
                   University of Zurich
   Date          : February 25, 2002

To discard polar and apolar receptor vectors that point outside of
the binding site a selection using an angle criterion can be
activated (page 10 of the corresponding SEED documentation).

Files modified:
source/funct.h
source/main.c
source/reduc_polvectre.c
source/reinfi.c
source/solv_lookup.c
inputs/seed.inp
inputs/seed.par

New documentation:
doc.ps

------------------------------------------------------------------------

SEED 3.0   Release     March      5, 2003

The following is introduced after the SEED 2.1 release dated
February 25, 2002.

1. Enhancement of the generalized Born model
   Name          : Nicolas Majeux
   Email address : majeux@bioc.unizh.ch
   Institution   : Department of Biochemistry
                   University of Zurich
   Date          : March 5, 2003

Empirical correction term to the Coulomb field approximation
  [MS Lee, FR Salsbury Jr, CL Brooks III,
   Journal of Chemical Physics, 116 (24), 10606, 2002]
```

```
for the accurate screened interaction and fragment desolvation
energies.

Files modified:
source/main.c
source/solv_frag.c
source/solv_lookup.c
source/funct.h
source/reinfi.c
inputs/seed.par

New documentation:
doc.ps

_____________________________________________________________________

SEED 3.1   Release     May       26, 2004

The following is introduced after the SEED 3.0 release dated
March 5, 2003.

1. Reduction of the number of output files
   Name          : Nicolas Majeux
   Email address : majeux@bioc.unizh.ch
   Institution   : Department of Biochemistry
                   University of Zurich
   Date          : May 26, 2004

The following files are not written out any more if the
"postprocess scheme" is selected: FragmentName_xxx.mol2,
FragmentName.chm, FragmentName_clus.chm, FragmentName_clus_reduc.chm .

File modified:
source/main.c

New documentation:
doc.ps

_____________________________________________________________________

SEED 3.2   Release     June      24, 2004

The following is introduced after the SEED 3.1 release dated
May 26, 2004.

1. Calculation of geometrical centers for FFLD
   Name          : Nicolas Majeux
   Email address : majeux@bioc.unizh.ch
   Institution   : Department of Biochemistry
                   University of Zurich
   Date          : June 24, 2004

The geometrical centers for FFLD are written out in
FragmentName_geomcent.mol2 files. The method is described
on page 413 of
M Cecchini, P Kolb, N Majeux, A Caflisch, Journal of Computational
Chemistry, 25, pp. 412−422, 2004.
Also on page 24 of the corresponding SEED documentation.

Files modified:
source/main.c
source/reinfi.c
source/funct.h
inputs/seed.par

File added:
```

```
source/geomcent.c

New documentation:
doc.ps

_____________________________________________________________________

SEED 3.3   Release     August     9, 2004

The following is introduced after the SEED 3.2 release dated
June 24, 2004.

1. Enhancement of input/output and energy evaluation mode
   Name          : Nicolas Majeux
   Email address : majeux@bioc.unizh.ch
   Institution   : Department of Biochemistry
                   University of Zurich
   Date          : August 9, 2004

The keyword RESN in fragment files is not mandatory any more. If RESN
is not found a fragment name is automatically assigned.
The maximal number of geometrical centers to write in *_geomcent.mol2
files is set in the parameter file.
In the postprocess mode the writing of *_pproc* files can be
activated or deactivated in the parameter file.
In the energy evaluation mode energies are not evaluated if the
ligand does not lie completely in the grids. Moreover a warning
message is written in the output log file.

Files modified:
source/refrfi_mol2.c
source/funct.h
source/reinfi.c
source/checkresn.c
source/geomcent.c
source/main.c
inputs/seed.par

New documentation:
doc.ps

_____________________________________________________________________
```

```
_____________________________________________________________________

SEED 3.3.5   Release     November     22, 2013

The following is introduced after the SEED 3.3 release dated
August 8, 2004.

1. Various modifications to output and file handling

   Name          : Fabian Dey, Tim Knehans
   Email address : fabian.dey@hotmail.com; t.knehans@bioc.uzh.ch
   Institution   : Department of Biochemistry
                   University of Zurich
   Date          : November 22, 2013

Bugfix: enabled docking of water (line 112 ff)
Fixed: dynamic memory allocation to store fragment energies
Fixed: removal ofwrite/read cycles of temporary files from harddisk
Fixed: small speedup of 3D similarity calculation for clustering routine
Fixed: exception handling
Bugfix: reduction of the polar points routine when only one point was
created to begin with
Fixed: speed up of fast receptor desolvation routine
Fixed: removed output of fast algorithm-calculated energy values from seed.out

Files modified:
source/reduc_polvectre.c
source/solv_frag_fast.c
source/simila.c
source/main.c

2. Adaptions of .par files for CHARMM36, CGenFF and the MATCH atom typer

   Name          : Emilie Frugier
   Email address : e.frugier@bioc.uzh.ch
   Institution   : Department of Biochemistry
                   University of Zurich
   Date          : November 22, 2013

Added: ligand atom types/parameters for CHARMM General Force Field (CGENFF)
Added: protein atom types/parameters for CHARMM36
Added: ligand atom types/parameters for MATCH atom types (CGENFF-derived)
Changed: seed.inp default parameter file now contains CHARMM36+CGENFF
Changed: MSI-CHARMM atom types now in seed_charmm-msi.par

Files created:
inputs/seed_charmm36-c37b2_cgenff-v2b8.par
inputs/seed_charmm36-c37b2_cgenff-v2b6_MATCH.par
inputs/seed_charmm-msi.par

Files changed:
inputs/seed.inp

New documentation:
doc_seed_3_3_5.pdf

_____________________________________________________________________
```