Internet Electronic Journal of Molecular Design

March 2005, Volume 4, Number 3, Pages 194–209

Editor: Ovidiu Ivanciuc

Proceedings of the Internet Electronic Conference of Molecular Design 2004
IECMD 2004, November 29 – December 12, 2004

New Application Design for a 3D Hydropathic Map–Based Search for Potential Water Molecules Bridging between Protein and Ligand

Glen E. Kellogg, Micaela Fornabaio, Deliang L. Chen, and Donald J. Abraham

Department of Medicinal Chemistry and Institute for Structural Biology & Drug Discovery, School of Pharmacy, Virginia Commonwealth University, Richmond, VA 23298–0540 USA

Received: November 12, 2004; Accepted: January 7, 2005; Published: March 31, 2005

Citation of the article:

G. E. Kellogg, M. Fornabaio, D. L. Chen, and D. J. Abraham, New Application Design for a 3D Hydropathic Map–Based Search for Potential Water Molecules Bridging between Protein and Ligand, *Internet Electron. J. Mol. Des.* **2005**, *4*, 194–209, http://www.biochempress.com.

Inter*net* Electronic Journal of Molecular Design

BioChem Press
http://www.biochempress.com

New Application Design for a 3D Hydropathic Map–Based Search for Potential Water Molecules Bridging between Protein and Ligand [#]

Glen E. Kellogg,* Micaela Fornabaio, Deliang L. Chen, and Donald J. Abraham

Department of Medicinal Chemistry and Institute for Structural Biology & Drug Discovery, School of Pharmacy, Virginia Commonwealth University, Richmond, VA 23298–0540 USA

Received: November 12, 2004; Accepted: January 7, 2005; Published: March 31, 2005

Internet Electron. J. Mol. Des. **2005**, *4* (3), 194–209

Abstract

The design of an extension to the HINT (Hydropathic INTeractions) program for locating and automatically placing in relevant orientations bridging water molecules is described in detail. This application is used to analyze the structures of HIV–1 protease–inhibitor complexes for five key bridging water molecules. The tool locates the water molecules with an overall accuracy of 1.28 ± 0.55 Å, and orients them (relative to potentially erroneous molecular mechanics optimized water molecules) to 41 ± 23 degrees. This automated placement is in contrast to other programs that calculate 3D contour maps of energetically–likely binding locations for water molecules, which must be followed by manual placement and energy minimization of these water positions and orientations to create the model. Also, the new object–oriented toolkit design for the HINT program (http://www.edusoft–lc.com/toolkits/) is described. The toolkit includes molecule, atom and monomer objects to represent chemical structure. A second class of objects describes 3D maps and includes tools for their manipulation. The hint objects are designed around the primary goals of HINT: partitioning (calculating $\mathrm{LogP_{o/w}}$ for) molecules, calculating interaction scores between molecules and calculating 3D maps of molecular and intermolecular hydropathy.

Keywords. HINT; hydropathy; GRID; HIV–1 protease; object oriented program.

Abbreviations and notations

HINT, Hydropathic INTeractions	$\mathrm{LogP_{o/w}}$, $\log_{10}$ of partition coefficient for 1–octanol/water
HIV–1, human immunodeficiency virus 1	3D, three dimensional

1 INTRODUCTION

We have been the proponents for an alternative molecular modeling forcefield based on the experimental information from the $\mathrm{LogP_{o/w}}$ (partitioning coefficient for water/1–octanol). Because this forcefield, which we refer to as HINT (for Hydropathic INTeractions) [1], is uniquely derived from a free energy measurement of interactions between small molecules and the two solvents, it implicitly includes solvation, desolvation and entropic effects. In particular, the HINT forcefield

[#] Presented in part at the Internet Electronic Conference of Molecular Design 2004, IECMD 2004.

* Correspondence author; phone: +01–804–828–6452; fax: +01–804–827–3664; E–mail: glen.kellogg@vcu.edu.

rewards hydrophobic–hydrophobic interactions as part of a free energy score [2]. We, and others, have described HINT in a number of publications [1–8], and have shown it to be useful in a fairly wide variety of biomacromolecular simulations.

Recently, we have re–coded the HINT algorithms in an object–oriented software toolkit. This has facilitated a number of recent enhancements and extensions of the HINT methodology, and has opened up possibilities for applying selected HINT algorithms to an even wider array of applications. In particular, the HINT free energy binding score is attractive as a tool in virtual screening.

In this paper we illustrate an application of this new toolkit in that we have written a program that searches for and places water molecules in positions optimum for bridging protein–ligand interactions. We also describe in an appendix the basic structure of the HINT toolkit, and document the relationships between the various molecule, atom, etc. objects. The HINT toolkit is written in very basic C and has successfully compiled on a number of platforms including IRIX, Windows, Mac OS–X and LINUX.

2 APPLICATION DESIGN: HINT MAP–BASED WATER SEARCH

We have become quite interested in the role of water and/or other cofactors as they contribute to ligand binding efficacy [8] and protein–protein interactions [3]. As the water molecules present in crystal structures have often not been exhaustively determined and, even when present, can be of variable reliability, we have been investigating computational methods to locate and verify these water molecules. In this section we describe a new computational algorithm we have designed for this purpose. The underlying principle is interaction scoring based on the HINT algorithm. In the Appendix to this paper the computational infrastructure for the application, the HINT toolkit, will be described.

2.1 Algorithm

The six panels of Figure 1 illustrate the basic algorithm in two dimensions. First the region surrounding the ligand is placed in a grid box, with spacing of around 0.5 Å or less. The box must extend beyond the extents of the ligand by at least 5 Å. In Figure 1a grid points that are within the value “range” from atoms in *both* the ligand and protein are marked in green. These are potential locations for bridging waters. Range has useful values of around 4.0–6.0 Å. Of these grid points, many are next excluded because they are too close to existing atoms. The algorithm removes grid points that are within $(R_{VdW} + R_{solvent}) \times S_{bump}$ from any atom, where R_{VdW} is the Van der Waals radius of the atom, S_{bump} is a bump weight (0.7 to 1.0) and $R_{solvent}$ is the solvent (water) radius that is usually 1.4 Å. The remaining grid points are indicated in purple (Figure 1b).

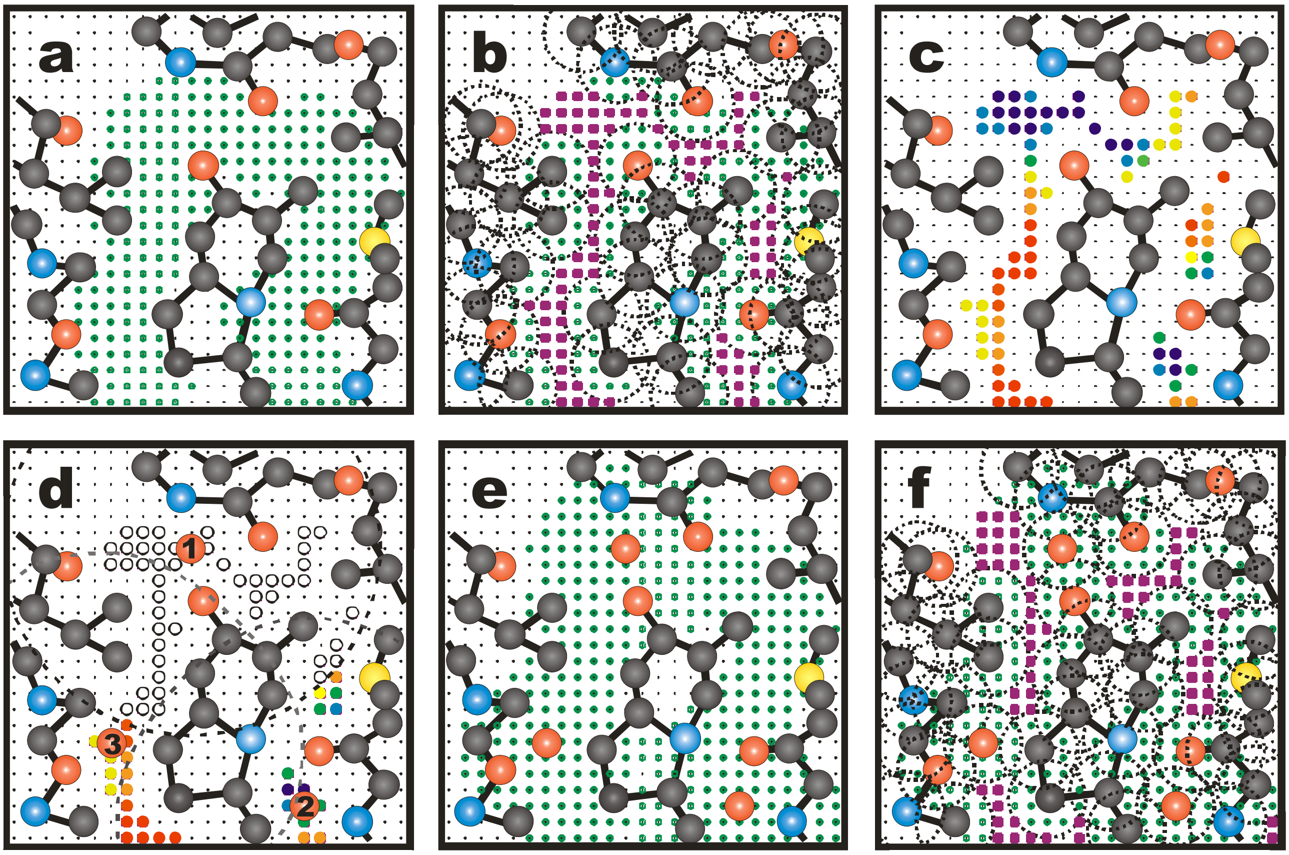


Figure 1. Algorithm for locating and placing bridging water molecules: a) grid points within range distance from atoms in both the ligand and protein are marked in green; b) grid points remaining after those too close to existing atoms are marked in purple; c) HINT scores for putative waters at each grid points are indicated by the color spectrum (blue–most favorable to red–least favorable); d) highest scoring water (1) is placed and all grid points within knockout distance are disabled and additional waters are placed similarly; e) next cycle uses new water molecules to define potential bridging grid points; and f) grid points too close to existing atoms (including those from new waters) are eliminated.

A putative water is placed at each of these grid points and exhaustively optimized using an algorithm we previously described [9]. The resulting “binding” interaction scores for these waters are qualitatively indicated by the colors in Figure 1c where blue indicates a more favorable position (while red is unfavorable). In this algorithm, not just the resulting scores, but the molecule objects (and atom structure) for each of the optimized water molecules are retained for future use. Next, the highest scoring water molecule (1) is placed on its grid point (Figure 1d) and all grid points within a “knockout” distance (typically 4.0–6.0 Å) are disabled. Of the remaining grid points, the highest scoring water (2) is placed and the knockout radius is again applied. This process is repeated until a minimum favorable score threshold is reached or the grid point pool is depleted. The next cycle commences (Figures 1e and 1f) where the water molecules from the previous cycle are used to redefine the grid points accessible for bridging and to exclude grid points too close to existing atoms.

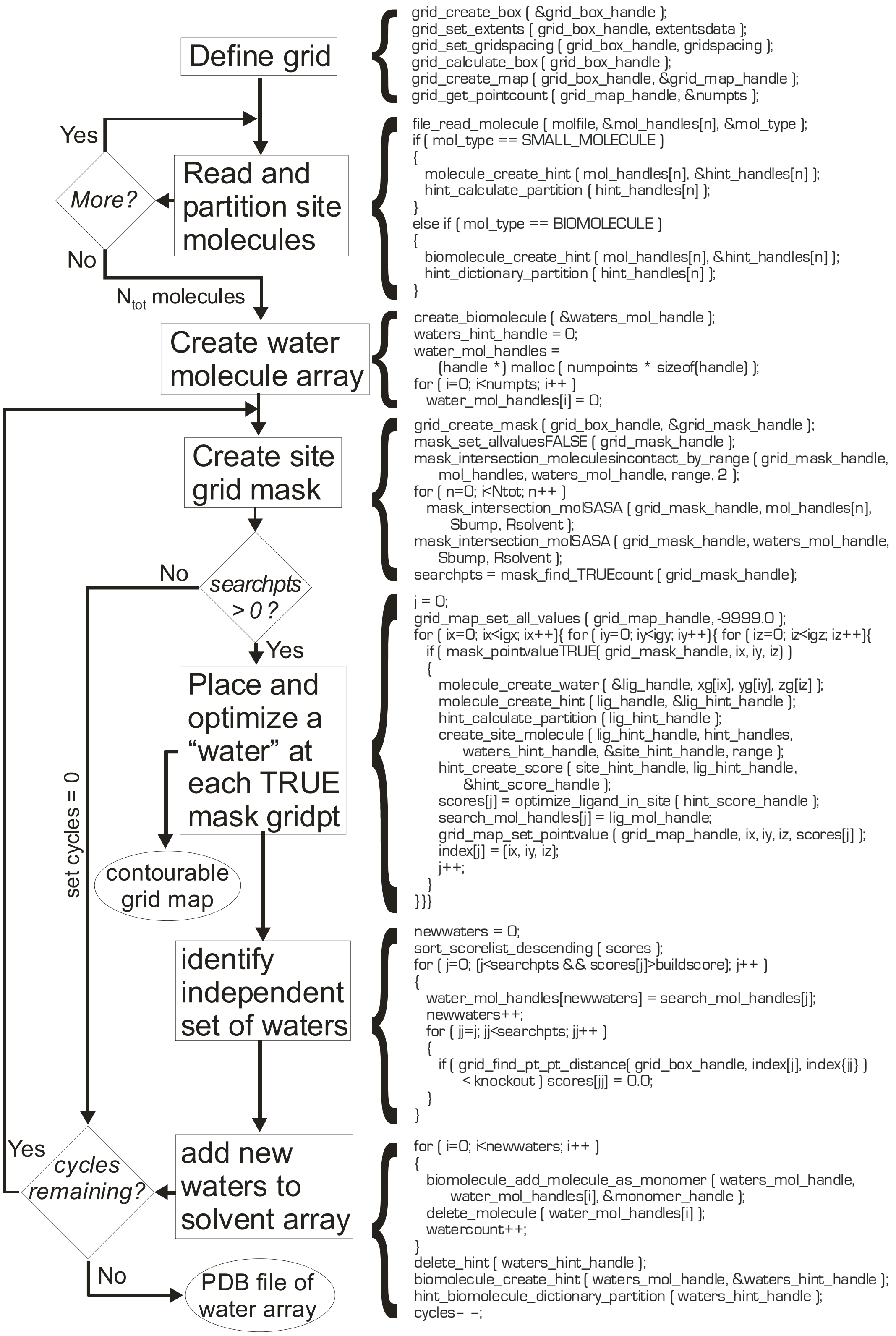
2.2 Pseudo–code

Figure 2 illustrates the algorithm in terms of a flowchart and blocks of pseudo–code. First, the extents of the region and grid spacing are used to create a **gridbox** object. From this a **gridmap** object is created. Each molecule occupying the region is then read creating **molecule** and/or **biomolecule** objects, an associated **hint** object is created for each molecule, and a **partition** is calculated for each **hint** object. The method for partitioning the molecule is dependent of the nature (*moltype*) of each molecule (see Appendix). Next, molecule objects are created for the water array: a biomolecule (*waters_mol_handle*) that will eventually incorporate the water molecules as a set of monomers, and a (temporary) array of small molecules (*water_mol_handles*) that is dimensioned to the number of points in **gridmap**. The **mask** object is created and calculated: 1) the set of molecules is tested to determine grid points that are within *range* distance of atoms of at least two molecules, and 2) each molecule occupying the region is tested to eliminate grid points within $(R_{VDW} + R_{solvent})\times S_{bump}$ of any atom. These remaining grid points comprise the *searchpts* set. If *searchpts* = 0, however, an exit condition is met, and *cycles* is set to zero and a PDB–format file of the water array is written.

The **mask** object is surveyed in three dimensions and a water **molecule** object is created at each of the TRUE grid points in the *searchpts* set. This “ligand” is used to define a unique “site” of atoms, culled from all molecules in the region, within *range* of the ligand. A hint **score** object is created to link the ligand and site, and the ligand orientation is optimized within the site yielding an optimum score. These scores are recorded in the **gridmap** object as field values, while the coordinate *index* and **molecule** handle for each of these ligands is retained. When this survey is complete the score values in the **gridmap** object can be written as a contourable map file, roughly similar to the output from the GRID [10] program (vide infra).

Next, the score list is sorted from highest to lowest; the water at the index of the highest score is added to the *water_mol_handles* list; and all grid points within *knockout* distance of that water are disabled in order to be certain that each new water molecule is independent. The next water is added at the index of the highest (remaining) score, *knockout* is again applied, etc., until no active grid points with viable scores remain. This set of new water molecules is added to the water array biomolecule and the small molecule waters are deleted.

If there are *cycles* remaining in the water search the water array biomolecule is prepared as an additional molecule occupying the site region, and the process is cycled back to the point where a new (updated) site **mask** object is created and calculated. Otherwise, if the cycles are depleted, the resulting water array is written as a PDB–format file.


Define grid
Yes
Read and partition site molecules
More?
No
N_{tot} molecules
Create water molecule array
Create site grid mask
No
searchpts > 0 ?
Yes
Place and optimize a “water” at each TRUE mask gridpt
set cycles = 0
contourable grid map
identify independent set of waters
Yes
cycles remaining?
add new waters to solvent array
No
PDB file of water array

```
grid_create_box ( &grid_box_handle );
grid_set_extents ( grid_box_handle, extentsdata );
grid_set_gridspacing ( grid_box_handle, gridspacing );
grid_calculate_box ( grid_box_handle );
grid_create_map ( grid_box_handle, &grid_map_handle );
grid_get_pointcount ( grid_map_handle, &numpts );
```

```
file_read_molecule ( molfile, &mol_handles[n], &mol_type );
if ( mol_type == SMALL_MOLECULE )
{
  molecule_create_hint ( mol_handles[n], &hint_handles[n] );
  hint_calculate_partition ( hint_handles[n] );
}
else if ( mol_type == BIOMOLECULE )
{
  biomolecule_create_hint ( mol_handles[n], &hint_handles[n] );
  hint_dictionary_partition ( hint_handles[n] );
}
```

```
create_biomolecule ( &waters_mol_handle );
waters_hint_handle = 0;
water_mol_handles =
     (handle *) malloc ( numpoints * sizeof(handle) );
for ( i=0; i<numpts; i++ )
  water_mol_handles[i] = 0;
```

```
grid_create_mask ( grid_box_handle, &grid_mask_handle );
mask_set_allvaluesFALSE ( grid_mask_handle );
mask_intersection_moleculesincontact_by_range ( grid_mask_handle,
     mol_handles, waters_mol_handle, range, 2 );
for ( n=0; i<Ntot; n++ )
  mask_intersection_molSASA ( grid_mask_handle, mol_handles[n],
       Sbump, Rsolvent );
mask_intersection_molSASA ( grid_mask_handle, waters_mol_handle,
     Sbump, Rsolvent );
searchpts = mask_find_TRUEcount ( grid_mask_handle);
```

```
j = 0;
grid_map_set_all_values ( grid_map_handle, -9999.0 );
for ( ix=0; ix<igx; ix++){ for ( iy=0; iy<igy; iy++){ for ( iz=0; iz<igz; iz++){
  if ( mask_pointvalueTRUE( grid_mask_handle, ix, iy, iz) )
  {
    molecule_create_water ( &lig_handle, xg[ix], yg[iy], zg[iz] );
    molecule_create_hint ( lig_handle, &lig_hint_handle );
    hint_calculate_partition ( lig_hint_handle );
    create_site_molecule ( lig_hint_handle, hint_handles,
        waters_hint_handle, &site_hint_handle, range );
    hint_create_score ( site_hint_handle, lig_hint_handle,
        &hint_score_handle );
    scores[j] = optimize_ligand_in_site ( hint_score_handle );
    search_mol_handles[j] = lig_mol_handle;
    grid_map_set_pointvalue ( grid_map_handle, ix, iy, iz, scores[j] );
    index[j] = (ix, iy, iz);
    j++;
  }
}}}
```

```
newwaters = 0;
sort_scorelist_descending ( scores );
for ( j=0; (j<searchpts && scores[j]>buildscore); j++ )
{
  water_mol_handles[newwaters] = search_mol_handles[j];
  newwaters++;
  for ( jj=j; jj<searchpts; jj++ )
  {
    if ( grid_find_pt_pt_distance( grid_box_handle, index[j], index{jj} )
        < knockout ) scores[jj] = 0.0;
  }
}
```

```
for ( i=0; i<newwaters; i++ )
{
  biomolecule_add_molecule_as_monomer ( waters_mol_handle,
      water_mol_handles[i], &monomer_handle );
  delete_molecule ( water_mol_handles[i] );
  watercount++;
}
delete_hint ( waters_hint_handle );
biomolecule_create_hint ( waters_mol_handle, &waters_hint_handle );
hint_biomolecule_dictionary_partition ( waters_hint_handle );
cycles– –;
```

Figure 2. Flowchart and blocks of pseudo–code for water search algorithm.

3 RESULTS AND DISCUSSION

As mentioned above, the issue of water presence as it relates to drug binding is of primary interest to our research program. In particular we have recently examined the role of water in ligand–bound complexes of HIV–1 protease [8]. In this section we illustrate how the new algorithm described above performs with respect to predicting and placing the water molecules from that earlier study.

3.1 Fragment and Molecule Complementarity Searches

The prototype program for using property complementarity on a three–dimensional grid to identify likely sites of interaction was Peter Goodford's GRID [10]. GRID presents an arsenal of potential atom, multi–atom fragment and molecular probes that can be placed on each of a set of grid points describing the region of interest. At each grid point the interaction energy for that probe interacting with the existing molecule, *e.g.*, a protein, is calculated and retained as the field value for a contour map of the region. This map can be contoured at specific energy levels and by visual inspection determine the energetically likely locations where that fragment (or molecule) may be found. In the case of water molecules it is only a matter of then placing them within these contours [8,10–12]. This placement usually should be followed by a energy minimization because the orientation of the water molecule is generally not revealed by the energy contour. An alternative application of GRID is to use several varied probes such as amine nitrogen, aromatic carbon, carboxylate, etc. over the region of interest to help define a three–dimensional (inverse pharmacophore) pattern for docking [7,10,13] or designing [10,14,15] ligands bound at the protein.

The MCSS (Multiple Copy Simultaneous Search) [16,17] method of Karplus randomly places a large number of small molecules that are representative of functional groups, *e.g.*, methanol for hydroxyl (–OH), in a protein active site. The interactions between these and the proteins are simultaneously optimized and as copies of the same molecule coalesce, they are pruned to reduce the computational expense. The resulting set of molecules, optimized for placement and orientation can be used to map potential binding locations for the functional group of interest within the protein active site. This method has also been used for the prediction of water locations within active sites [11,18].

3.2 Tuning of Adjustable Parameters in HINT Method

The user–adjustable parameters for the HINT algorithm were described above. As we built and optimized this application we discovered that the values of these parameters are very critical for the successful location of crystallographically–known water molecules. First, we need to locate the "holes" where the waters are potentially located: (*a*) we are using a *gridspacing* of 0.5 Å on all axes because larger values, *i.e.*, 1.0 Å may miss some of the smaller holes, while smaller values rapidly increase the CPU expense; (*b*) with *range* the grid points that are potentially bridging, *i.e.*, accessible to more than one molecule are identified. Although it masks a set larger than really necessary, we have found a value of around 6.0 Å works best. This means that a water would be considered bridging if it is within 6.0 Å of at least one atom in the ligand and at least one atom in the protein; (*c*) grid points from this set that are already occupied by other atoms are excluded. Since we are using the accepted value of $R_{solvent}$ = 1.4 Å, the key adjustable parameter here appears to be S_{bump}. If a hole is small, as it is for water 301 of HIV–1 protease–ligand complexes (next section), a too large S_{bump} will fail to locate the hole; in contrast a too small S_{bump} may allow water molecules to be built unrealistically close to other atoms in the ensemble. After numerous computational experiments we have settled on S_{bump} = 0.85 as the best compromise value. However, it should be noted that this is contingent on the Van der Waals radii set used in the model [19].

The second phase involves selecting and adding to the **biomolecule** object the appropriate set of water molecules from the gridpoint scores. Three variables are in play here: *buildscore* – the minimum acceptable score for a water molecule, *knockout* – the minimum distance between two water centers built in the same cycle, and *cycles* – the number of iterations over the entire algorithm allowed before the search process ends. While we have observed crystallographic water molecules with HINT scores < 0 in some studies [3,8], we have set the *buildscore* threshold at a positive number, ca. 500, for this work. The *knockout* radius turns out to be a very sensitive instrument for building the final water set. In this work we have used a value of 4.0 Å, which we have found to most accurately match the experimental positions of waters in the HIV–1 protease data set. However, values of 5.0 or 6.0 Å give somewhat different sets of water, but it should be noted that these bridging water sets serve essentially the same energetic role in terms of stabilizing the ligand binding by bridging between the protein and ligand. Finally, only a small handful of cycles are necessary to locate and build the "true" bridging water set. Later cycles, however, are interesting in themselves as successive layers of water molecules are added to the regions outside of the binding site. We are terming this effect "saturation" and are exploring it further. In order to accelerate completion we are incrementing S_{bump} each round by 0.075.

3.3 Key Waters in the Active Site of HIV–1 Protease

In a series of recent papers we have been evaluating HINT as a tool for free energy scoring in protein–ligand complexes [4,5,8]. The most recent paper [8] focused on the energetic contributions of bridging water in ligand binding calculations for an extensive series of 23 HIV–1 protease–inhibitor complexes. While 71/109 of the water molecules of interest (see Table 1) had been located crystallographically, to identify the others, sites where they would be expected to bind were examined with GRID. An additional 31 water molecules were thus placed, leaving 7 sites that GRID indicated were too sterically constrained for water. Although GRID functions with impressive accuracy, this is an extremely tedious process, suggesting to us that a more automated procedure could perhaps be created that would combine the energy search and placement steps in one computational procedure.

For the present study, we have reexamined the HIV–1 protease–inhibitor molecular models [8] using the algorithm described above. The key features of HIV–1 protease with respect to this study are a conserved water molecule, water 301, located on the HIV–1 protease symmetry axis (bridging the two subunits), and two pairs, 313/313' and 313bis/313bis' [8], of largely conserved water molecules located in more peripheral areas of the active site. Water 301 is hydrogen bonded to the Ile50 and Ile150 protease residues and to the inhibitors. It has been observed in all HIV–1 protease–ligand complexes, except where it has been, by design, displaced. Water 313 was named by Jhoti and colleagues [20] and can be found near the salt bridge between Asp29 and Arg108 interacting with both protein and ligand(s). Water 313' is in the pseudo–symmetric site near Asp129 and Arg8. Waters 313bis and 313bis' interact strongly with residues Arg87/Thr26 (313bis) and Arg187/Thr126 (313bis') and rather weakly with the ligands. Waters 313bis and 313bis' also interact strongly with waters 313 and 313', respectively.

Sybyl "mol2" files for each of the complexes, with the protein, ligand and water sets separated as distinct objects were used for this study. Two comparisons of water location and orientations were made, first with the x–ray crystallography–positioned water molecules, followed by proton–only minimization of the entire complex structure with the Tripos forcefield, using Gasteiger–Hückel charges, to a gradient of 0.005 kcal (mol Å)$^{-1}$. Second, the "final" water sets in these models [8] result from crystallographic data supplemented by GRID analyses, followed by final positioning with the HINT water optimization tool [9]. Two metrics for describing the results are reported: a) the distance between the predicted and actual water oxygen atoms; and b) the angle between the dipole moments of the predicted and actual water molecules. These data are summarized in Table 1 for the 23 protein–ligand complexes in the study. Specific data for each water molecule is provided in Table 2 (supplementary material).

Table 1. Average position[a] and orientation[b] error data for key water molecules in HIV–1 protease complexes

Water	X–ray/Molecular Mechanics			Final (X–ray or GRID/HINT Optimization)		
	$d_{O–O}$, Å	$\Theta_{d–d}$, deg	Water count	$d_{O–O}$, Å	$\Theta_{d–d}$, deg	Water count
Water 301	1.31 ± 0.47	56 ± 27	17	1.28 ± 0.46	49 ± 20	17
Water 313	1.60 ± 0.38	69 ± 40	12	1.42 ± 0.53	39 ± 25	20
Water 313'	1.41 ± 0.57	57 ± 41	11	1.37 ± 0.63	36 ± 37	19
Water 313bis	1.07 ± 0.53	51 ± 17	16	1.14 ± 0.53	39 ± 17	23
Water 313bis'	1.14 ± 0.55	59 ± 15	15	1.21 ± 0.51	45 ± 14	23
All	1.28 ± 0.55	58 ± 29	71	1.28 ± 0.56	41 ± 23	102

[a] The distance between the oxygen atoms of the water molecules generated by the algorithms in this work and the oxygen atoms of the crystallographic waters or final modeled waters from Ref. [8]. [b] The angle between the dipole moments of the molecules generated by the algorithms in this work and the dipole moments of the molecular mechanics minimized waters or the final modeled waters from Ref. [8].

The algorithm does a fair–to–good job of locating the key water molecules, with an average positional error of 1.28 ± 0.55 Å. The errors do vary with the specific water, with lower errors associated with the more peripheral, but apparently conserved, waters 313bis and 313bis'. The largest errors are associated with waters 313 and 313', where asymmetry of the ligands has more effect on the water positions, and the crystallography located only 23 out of the 39 waters that were used in the final models. The positional error of water 301 is surprisingly large, 1.31 ± 0.47 Å. It was our initial thesis that this water, considering its highly conserved nature, would be the simplest to locate and place. However, because water 301 makes such tight hydrogen bonds with both the protein and ligand, we found that it was quite difficult to tune the steric search parameters to locate it. The problem is that opening the "hole" for this water, while simultaneously disallowing other water positions that are too close to protein or ligand atoms in other sites, can not be realistically achieved. Thus, in this work, where we specifically required water 301 to be found, the side effect was that our algorithm reported additional, possibly spurious, water molecules in and near the active site. It should be noted that none of the HIV–1 protease/inhibitor crystal structures examined in this study were reported to atomic or near–atomic resolution, the best being 1.8 Å. As the number of water molecules found by crystallography is proportional to the resolution [21,22], it is likely that some of the "questionable" waters found by our algorithm may actually be "real". We believe that further refinement of the method and parameters may improve the utility of this tool, especially with respect to these last issues.

The orientation angle errors of the predicted waters are, on the other hand, quite encouraging; angle errors of 40 – 60 degrees are acceptable as the same key polar interactions can be made with either water molecule. Here, however, there is a noticeable difference in prediction errors computed against the x–ray/molecular mechanics waters compared to prediction errors computed against the final/HINT optimized waters. One reason is that a number of the energy minimized water molecules were trapped in local minima and did not orient with the best hydrogen bonding patterns.

The last question to be answered is how does this algorithm perform when compared to GRID? Somewhat less than half (43/101) of the waters found by our algorithm are within the corresponding (–9 kcal mol^{-1}) GRID contours, while about two–thirds (52/74) of the waters reported crystallographically are. Most of the other waters are less than ≈ 0.5 Å outside the contours. However, the relevant point is that, instead of this retrospective analysis, actual placement of water molecules within the contours is more subjective. For example, in the region of waters 313/313bis (left hand side, Figure 3) in PDB structure 1HIV, [23] it can be easily argued that the –9 kcal mol^{-1}

GRID contours are indicating both of these crystallographic water molecules since they are known to be there. In the symmetry–related water 313'/313bis' region only water 313bis' is clearly in the density, but there is a *very* small GRID contour near water 313'. Would four water molecules have been placed by a user in these contour envelopes in the absence of the crystal data?

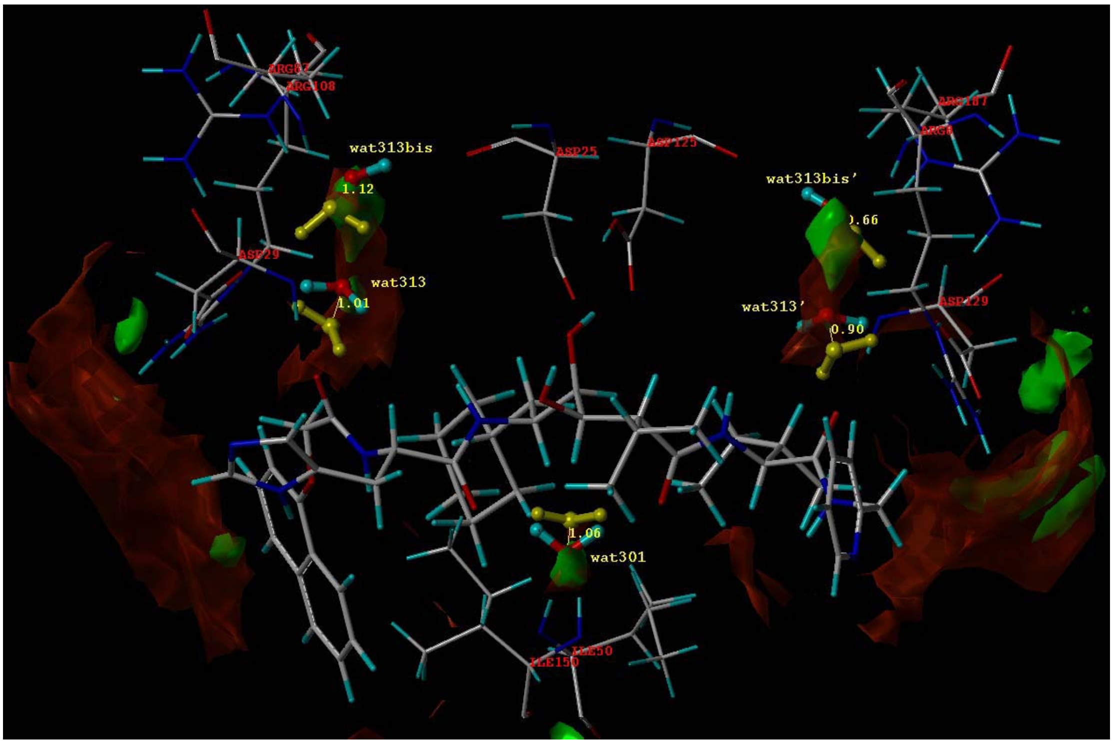


Figure 3. The ligand binding site in HIV–1 protease–inhibitor complex 1HIV. GRID contours are indicated with opaque green contour surfaces, HINT water search contours are indicated by translucent orange surfaces. The waters placed by the HINT algorithm are colored yellow and the distances between these and the optimized crystallographic waters are shown.

4 CONCLUSIONS

The HINT toolkit provides an object–oriented entry point to the HINT forcefield model and algorithms. The application presented here, a tool for locating and placing bridging water molecules in protein–ligand complexes, can be coded with a relatively simple program using calls to the HINT toolkit. While further development and tuning of the application is currently underway, it does perform the task intended – inventing an energetically reasonable set of bridging water molecules that would impact the ligand binding process. On the other hand, we have to say that, while it is awkward to apply, the accuracy of the GRID program and its forcefield is truly impressive. Only a relatively small number of actual water molecules reported by crystallography in this series of complexes were not confirmed by GRID. Nevertheless, this current HINT–based algorithm is an important piece of our overall goal of building an *integrated* virtual screening platform that incorporates water searches of a similar nature for each docked putative ligand.

Acknowledgment

We gratefully acknowledge useful discussions with J. Neel Scarsdale, Jason P. Rife and H. Tonie Wright (VCU) and Pietro Cozzini and Andrea Mozzarelli (University of Parma) in developing the water search algorithm. Francesca Spyrakis (University of Parma) helped with building the HIV–1 protease complex models. The hint toolkit concept was inspired by the pioneering work of David Weininger (Daylight Chemical Information Systems) and Anthony Nicholls (OpenEye Scientific Software). This work has been partially supported by Virginia Commonwealth University. GRID has been made available to VCU by Molecular Discovery, Ltd. (www.moldiscovery.com). The hint model was developed in the VCU laboratory of G.E.K.; the hint toolkit is available from eduSoft LC (www.edusoft–lc.com).

Supplementary Material

Table 2, with specific geometric parameters for all water molecules analyzed in this study is available as supplementary material. In addition, mol2 files for the 23 HIV–1 protease/ligand complexes with final water positions are available as supplementary material.

Table 2. Position[a] and orientation[b] error data for key water molecules in HIV-1 protease complexes[c]

PDB [d]	Water 301				Water 313				Water 313'			
	$d_{\text{O-O}}$, Å *xtal*	$d_{\text{O-O}}$, Å *final*	$\Theta_{\text{d-d}}$, deg *MM*	$\Theta_{\text{d-d}}$, deg *final*	$d_{\text{O-O}}$, Å *xtal*	$d_{\text{O-O}}$, Å *final*	$\Theta_{\text{d-d}}$, deg *MM*	$\Theta_{\text{d-d}}$, deg *final*	$d_{\text{O-O}}$, Å *xtal*	$d_{\text{O-O}}$, Å *final*	$\Theta_{\text{d-d}}$, deg *MM*	$\Theta_{\text{d-d}}$, deg *final*
1HXW	0.821	0.803	43.1	33.0	2.480	2.480	37.5	45.0	2.226	2.248	65.2	28.9
1HVJ	1.563	1.484	65.1	71.2	-	0.442	-	14.0	-	-	-	-
1HXB	1.191	1.144	44.5	46.2	-	1.390	-	7.1	-	2.652	-	169.0
1HTG	2.320	2.252	87.0	73.4	1.663	1.643	31.5	28.6	2.043	2.041	41.4	25.8
7HVP	1.633	1.604	71.6	68.2	2.435	2.366	49.9	19.0	2.226	1.883	64.6	21.0
1HPV	1.333	1.293	22.6	24.7	-	2.161	-	90.8	1.806	1.766	169.1	91.3
1HPS	0.908	0.843	73.7	29.3	-	1.432	-	19.7	-	2.133	-	25.2
4PHV	0.775	0.803	24.3	30.0	1.326	1.280	27.5	23.3	1.182	1.117	14.7	6.6
1AAQ	0.537	0.526	68.5	56.8	-	0.627	-	26.9	-	0.710	-	27.8
1HTF	1.553	1.525	88.3	53.0	1.168	1.080	39.9	34.8	1.651	1.590	73.4	29.9
1HIH	0.607	0.571	5.7	18.6	1.683	1.539	42.9	16.7	1.479	1.454	43.7	26.0
1SBG	1.597	1.575	107.2	84.4	1.579	1.510	96.3	65.6	-	0.903	-	43.4
1HVK	1.097	1.115	33.5	39.1	-	-	-	-	-	-	-	-
1HVI	1.620	1.589	66.8	60.2	-	0.954	-	16.6	-	-	-	-
1HVL	1.645	1.620	52.7	52.3	-	-	-	-	-	-	-	-
1HIV	1.105	1.058	74.1	62.8	1.041	1.011	33.3	20.7	0.945	0.904	32.0	21.6
1HBV	1.957	1.886	30.1	24.1	-	-	-	-	-	1.775	-	13.4
1QBT	-	-	-	-	-	2.380	-	43.6	-	1.751	-	37.1
1DMP	-	-	-	-	-	0.602	-	29.7	-	0.453	-	6.9
1AJX	-	-	-	-	1.878	1.828	135.1	80.9	0.774	0.785	45.8	10.1
1G35	-	-	-	-	1.463	1.372	89.0	55.1	0.747	0.739	37.6	49.7
1G2K	-	-	-	-	1.193	1.188	127.9	74.1	-	0.694	-	26.3
1AJV	-	-	-	-	1.249	1.213	113.6	59.6	0.465	0.514	35.4	19.4

[a] The distance between the oxygen atoms of the water molecules generated by the algorithms in this work and the oxygen atoms of the crystallographic waters (xtal) or (final) modeled waters from Ref. [8]. [b] The angle between the dipole moments of the molecules generated by the algorithms in this work and the dipole moments of the molecular mechanics minimized waters (MM) or the (final) modeled waters from Ref. [8]. [c] When no data is indicated for crystallographic or molecular mechanics waters, the crystal structure did not identify that particular water molecule. When no data is indicated for the final optimized waters, GRID was unable to place a water molecule in that particular region due to steric constraints. [d] PDB code for complex. See Ref. [8] for references and molecular structures.

Table 2. (Continued)

PDB ^{d}	Water 313bis				Water 313bis'			
	$d_{O\text{-}O}$, Å *xtal*	$d_{O\text{-}O}$, Å *final*	$\Theta_{d\text{-}d}$, deg *MM*	$\Theta_{d\text{-}d}$, deg *final*	$d_{O\text{-}O}$, Å *xtal*	$d_{O\text{-}O}$, Å *final*	$\Theta_{d\text{-}d}$, deg *MM*	$\Theta_{d\text{-}d}$, deg *final*
1HXW	1.069	1.036	40.9	33.3	0.971	0.986	66.5	51.5
1HVJ	-	0.859	-	31.1	-	1.492	-	53.8
1HXB	1.134	1.173	56.3	59.4	2.010	2.015	51.7	46.0
1HTG	1.482	1.456	65.8	51.9	1.123	1.139	69.0	59.4
7HVP	1.579	1.554	75.9	58.9	0.653	0.462	47.9	34.6
1HPV	0.866	0.878	50.7	44.6	0.984	1.015	68.2	48.6
1HPS	1.157	1.104	48.1	31.4	-	1.394	-	51.8
4PHV	0.650	0.654	81.5	38.2	1.334	1.315	81.3	80.7
1AAQ	-	0.672	-	19.3	-	2.011	-	34.6
1HTF	0.518	0.440	42.5	26.8	0.892	0.916	60.3	45.5
1HIH	0.510	0.505	49.6	10.6	1.412	1.359	83.2	62.6
1SBG	2.555	2.571	59.7	70.1	0.640	0.621	45.7	38.0
1HVK	-	2.005	-	46.7	-	1.519	-	44.8
1HVI	-	1.551	-	54.2	-	1.319	-	37.9
1HVL	-	1.609	-	55.7	-	1.201	-	13.5
1HIV	1.166	1.119	52.1	51.7	0.631	0.662	67.9	51.5
1HBV	1.041	1.064	13.9	9.0	1.590	1.607	35.7	40.3
1QBT	-	1.666	-	49.0	-	0.707	-	36.9
1DMP	-	0.787	-	31.7	-	0.897	-	42.8
1AJX	0.208	0.230	26.8	11.3	0.676	0.716	31.6	23.4
1G35	1.122	1.122	40.5	27.4	1.118	1.148	65.1	49.5
1G2K	1.129	1.120	62.1	43.9	2.505	2.568	66.2	57.3
1AJV	0.891	0.939	42.8	33.8	0.626	0.681	46.6	39.7

Appendix 1: Toolkit Design

1.1 Molecules, biomolecules, atoms and monomers

The first task is to define chemistry in terms of objects. We have chosen to define two types of molecule objects, the first for a small **molecule** where there are no predefined monomers, and the second for a **biomolecule** where monomers are defined. Table 3 lists typical items included in these two object types. In both the **molecule** and **biomolecule** structures, the first item is a “handle” or pointer to the **atom** structure. In the **biomolecule** structure there is an array of handles for the **monomer** structures. Other items in both the **molecule** and **biomolecule** structures refer to properties of the molecules such as name, number of atoms, etc.

The **atom** structure (Table 3) is created after the **molecule** (or **biomolecule**) structure and is allocated memory consistent with the number of atoms in the molecule. Thus, each item in the **atom** structure, which represents properties of individual atoms, is addressed by atom number. The *type_id* parameter is a code to a specific, forcefield–dependent atom (potential) type. We have coded both TAFF (Sybyl Tripos) and CVFF (insightII Constant Valence) FF types. The toolkit is structured such that calls to the routines expect atom 1 to be the first atom. The *icon* and *bnd* arrays define the bonds to each atom. *icon* is a connection matrix where each value is the atom number of an bonded atom. A value of zero is interpreted as no bond. The *bnd* array is constructed in parallel and indicates the bond order as a floating point value for the bond, e.g., an order of 1.5 indicates an aromatic bond. Each atom also has a pointer (*monomer*) that will, if non–NULL, indicate the **monomer** structure of which it is a member.

There is a **monomer** structure (Table 3) for each defined monomer in a biomolecule. (Note that in the **biomolecule** structure, *monomers* is an array of handles.) The **monomer** structure straddles, in a sense, both the **biomolecule** and **atom** structures and has linkages to both. The item *biomolecule* is a handle to the parent **biomolecule** structure that in turn references the **atom** structure. Also, however, the **monomer** item *atoms* is an array of atom numbers in that structure that are members of the monomer. A number of items in the **monomer** structure, e.g., *motif*, *acidbase* and *terminationtype*, are codes to specific properties of the monomer that define its secondary structure, ionization state and related properties. The *type_id* parameter in **monomer** is a code to a specific residue type, again defined in terms of TAFF and CVFF.

Table 3. Molecule, biomolecule, atom and monomer data structures

Structure	Item	Data type	Description
molecule	atoms	handle	Pointer to atoms structure
	atomcount	int	Number of atoms
	forcefield	int	Code for forcefield type with respect to atom types
	molname	char[64]	String with molecule name
	formula	int[elements]	Count of each element, e.g., H[6]C[2]O[1]
biomolecule	atoms	handle	Pointer to atoms structure
	atomcount	int	Number of atoms
	forcefield	int	Code for forcefield type with respect to atom types
	molname	char[64]	String with molecule name
	monomercount	int	Count of monomers in biomolecule
atom	atomicnumber	int	Atomic number
	hydrogens	int	Count of implicit hydrogens
	type_id	int	Code for atom types
	x, y, z	floats	Cartesian coordinates of atom
	formalcharge	float	Formal charge of atom
	icon	int[8]	Connection matrix for atom
	bnd	float[8]	Bond order matrix for atom
	monomer	handle	Pointer to monomer structure containing atom (or NULL)
	atomname	char[8]	String with atom name
monomer	biomolecule	handle	Pointer to biomolecule structure
	motif	int	Code for structural motif of monomer, e.g., α–helix, β–sheet, etc.
	type_id	int	Code for monomer type, e.g., Ala, Lys,, A, C, etc.
	acidbase	int	Code for acid/base condition of monomer, i.e., ionization state
	terminationtype	int	Code for termination of monomer, i.e., none, N–terminal, C–terminal, O3’, O5’, etc.
	atomcount	int	Count of atoms in monomer
	atoms	int[128]	List of specific atoms (in atom structure)
	monomer_name	char[8]	Name of monomer
	chain_name	char[3]	Name of chain membership

1.2 Three–dimensional grid maps.

There are two primary structures associated with the creation and manipulation of 3D grid maps: **gridbox** and **gridmap**. The **gridbox** structure (Table 4) encodes the information describing the placement, orientation and extents of the grid. The **gridmap** structure (Table 4) has a handle, *gridbox*, to its associated **gridbox** structure. (The inverse is not the case because one grid box may orient more than one map.) There is also a convenience handle, *mapsource*, to the source of the data in the map, e.g., a HINT object. The main data item in the **gridmap** structure is *values*, a floating point array of the map data indexed to grid point ip such that ip = ix + (igx*iy) + (igx*igy*iz), for (ix, iy, iz), where 0≤ix<igx, 0≤iy<igy and 0≤iz<igz. The **mask** structure directly parallels the **gridmap** structure but is constructed with integer (int) data rather than floating point (float) data.

Table 4. Grid box and map data structures

Structure	Item	Data type	Description
gridbox	xcen, ycen, zcen	floats	Cartesian coordinates of grid box center
	xwid, ywid, zwid	floats	Widths of grid box on x, y and z axes
	xg, yg, zg	float[igx], float[igy], float[igz]	Coordinate arrays for grid points on x, y and z axes
	gsx, gsy, gsz	floats	Grid spacings on x, y and z axes
	igx, igy, igz	ints	Number of grid points on x, y and z axes
gridmap	maptype	int	Code for map type, i.e., calculation type, etc.
	mapsource	handle	Pointer to data source for map, e.g., HINT object structure, etc.
	gridbox	handle	Pointer to grid box structure
	pointcount	long int	Count of data points in map
	minvalue	float	Minimum map field value
	maxvalue	float	Maximum map field value
	sumvalues	float	Sum of all field values in map
	sumsquarevalues	float	Sum of squares of all field values in map
	values	float[pointcount]	Array of field values, to find value at (ix, iy, iz): ip=ix+(igx*iy)+(igx*igy*iz)
	description	char[256]	String with optional comments, etc. concerning map
mask	maptype	int	Code for map type, i.e., calculation type, etc.
	values	int[pointcount]	Array of field values, to find value at (ix, iy, iz): ip=ix+(igx*iy)+(igx*igy*iz)

1.3 HINT objects

The primary structure is **hint** (Table 5), which is derived from a **molecule** or **biomolecule** structure. This structure largely holds pointers, i.e., to the parent structure, *molecule*, or to the derivative structures *partition* and *atomdata*. The distinction between small molecule and biomolecule, indicated by *moltype*, is important for the later partitioning of the molecule.

The HINT **partition** and the HINT **atomdata** (Table 5) structures are created when the molecule is partitioned, which is the step where HINT parameters (a_i – hydrophobic atom constant and S_i – solvent–accessible–surface–area) are assigned to each atom. The sum Σ a_i is the $\mathrm{LogP}_{o/w}$ for the molecule. Like the **atom** structure (above), the **atomdata** structure is dimensioned for the number of atoms in the molecule. There are two principal means (*logpmethod*) of partitioning a molecule in HINT. Small molecules are partitioned using an adaptation of the CLOGP method of Leo,[24] which is similar to the method or Rekker.[25] This method (calculate) takes into account the atom types and connections of the molecule. The dictionary method[26] invokes lookup tables based on atom and residue names and types. For the calculate method there are two methods of polar proximity correction. The standard (CLOGP–like) method is based on the connection distances via–bonds. The alternative is a through–space method for which the user can create specific mathematical functions. *solventcondition* is applied to biomolecules partitioned with the dictionary method: the pH can be considered as acid, base, neutral or inferred. In the latter case the specific protonation status of each monomer is examined and the *solventcondition* is assigned to each accordingly. *hydrogentreatment* refers to how hydrogens in the structures will be partitioned: one of all (partition all hydrogens), polar–only (partition only polar hydrogens and incorporate non–polar hydrogens into non–polar united atoms) or united (all hydrogens are incorporated in united atoms).

Two of the functions of HINT are calculating interaction scores which can be related to free energy [2,3–5,7,8] and calculating 3D contour maps that display a number of hydropathic properties.[1,27] These two functions are embodied in the HINT objects **score** and **map** (Table 5). Both the HINT **score** structure and the HINT **map** structure can be created for either unimolecular (molecular, inverse or intramolecular) or bimolecular (intermolecular) cases. Handles referencing the parent **hint** object(s), *hint* and *hint2*, are included in these structures. For the HINT **map** structure there is also a handle (*gridmap*) to the resulting **gridmap**. The *hintdistancefunction* has been described previously.[1–4,27] It includes both terms for hydropathic interactions (usually a simple exponential) and Van der Waals (Lennard–Jones potential function). Codes and values for a number of calculational parameters (*tabletype*, *maptype*, *dataselect*, *volumeaverage* and *gridsizescale*) are briefly described in Table 5. The key results from a score calculation are the values recorded in the *scores* array that can be parsed by interaction type if desired.

Table 5. HINT structures

Structure	Item	Data type	Description
hint	molecule	handle	Pointer to parent molecule structure
	partition	handle	Pointer to partition structure
	atomdata	handle	Pointer to HINT atom data structure
	moltype	int	Code for molecule type, i.e., small, biomolecule, solvent array
	atomcount	int	Count of atoms in molecule
partition	logpmethod	int	Code for logP calculation method, i.e., calculate or dictionary
	polarproximitytype	int	Code for polar proximity method, i.e., via–bond or through–space
	solventcondition	int	Code for solvent condition, i.e., acid, base, neutral or inferred
	hydrogentreatment	int	Code for disposition of hydrogens during partitioning, i.e., united atoms, polar only, or include all
	LogP	float	The resulting $\mathrm{LogP}_{o/w}$ for molecule
atomdata	A	float	Hydrophobic atom constant for atom
	S	float	Solvent–accessible–surface–area for atom
	nhydpol	int	Code (parent atom number) for polar fragment of which atom is a member
	hydpol	float	Fragment constant for polar group when atom is parent of fragment, else NULL
score	hint	handle	Pointer to first HINT object structure
	hint2	handle	Pointer to second HINT object structure (if necessary, else NULL)
	hintdistancefunction	structure	Data defining functional form and parameters for HINT distance function
	tabletype	int	Code for type of score calculation, i.e., intermolecular, intramolecular, etc.
	dataselect	int	Code for type of interactions to be included in score, i.e., all, polar–only or hydrophobic–only
	scores	float[8]	Interaction scores indexed by category, i.e., total, H–bond, hydrophobic, etc.
map	hint	handle	Pointer to first HINT object structure
	hint2	handle	Pointer to second HINT object structure (if necessary, else NULL)
	gridmap	handle	Pointer to grid map object structure
	hintdistancefunction	structure	Data defining functional form and parameters for HINT distance function
	maptype	int	Code for type of HINT map calculation, i.e., intermolecular, intramolecular, etc.
	volumeaverage	int	Average (TRUE) or not (FALSE) each grid point from 8 surrounding pseudopoints
	dataselect	int	Code for type of interactions to be included in map, i.e., all, polar–only or hydrophobic–only

5 REFERENCES

[1] F. C. Wireko, G. E. Kellogg, and D. J. Abraham, Allosteric Modifiers of Hemoglobin: 2. Crystallographically Determined Binding Sites and Hydrophobic Binding / Interaction Analysis of Novel Hemoglobin Oxygen Effectors, *J. Med. Chem.* **1991**, *34*, 758–767.

[2] G. E. Kellogg and D. J. Abraham, Hydrophobicity: Is $\mathrm{LogP}_{o/w}$ More than the Sum of its Parts?, *Eur. J. Med. Chem.* **2000**, *35*, 651–661.

[3] J. C. Burnett, P. Botti, D. J. Abraham, and G. E. Kellogg, Computationally Accessible Method for Estimating Free Energy Changes Resulting from Site Specific Mutations of Biomolecules: Systematic Model Building and Structural/Hydropathic Analysis of Deoxy and Oxy Hemoglobins, *Proteins: Struct. Funct. Genet.* **2001**, *42*, 355–

377.

[4] P. Cozzini, M. Fornabaio, A. Marabotti, D. J. Abraham, G. E. Kellogg, and A. Mozzarelli, Simple Intuitive Calculation of Free Energy of Binding of Protein–Ligand Complexes. 1. Models without Explicit Constrained Water, *J. Med. Chem.* **2002**, *45*, 2469–2483.

[5] M. Fornabaio, P. Cozzini, A. Mozzarelli, D. J. Abraham, and G. E. Kellogg, Simple, Intuitive Calculations of Free Energy of Binding for Protein–Ligand Complexes. 2. Computational Titration and pH Effects in Molecular Models of Neuraminidase–Inhibitor Complexes, *J. Med. Chem.* **2003**, *46*, 4487–4500.

[6] R. Gussio, D. W. Zaharevitz, C. F. McGrath, N. Pattabiraman, G. E. Kellogg, C. Schultz, A. Linke, C. Kunick, M. Loest, L. Meijer, and E. A. Sausville, Structure–Based Design Modifications of the Paullone Molecular Scaffold for Cyclin–Dependent Kinase Inhibition, *Anti–Cancer Drug Des.* **2000**, *15*, 53–66.

[7] D. J. Cashman and G. E. Kellogg, A Computational Model for Anthracycline Binding to DNA: Tuning Groove–Binding Intercalators for Specific Sequences, *J. Med. Chem.* **2004**, *47*, 1360–1374.

[8] M. Fornabaio, F. Spyrakis, A. Mozzarelli, P. Cozzini, D. J. Abraham, and G. E. Kellogg, Simple, Intuitive Calculations of Free Energy of Binding for Protein–Ligand Complexes. 3. The Free Energy Contribution of Structural Water Molecules in HIV–1 Protease Complexes, *J. Med. Chem.* **2004**, *47*, 4507–4516.

[9] G. E. Kellogg and D. L. Chen, The Importance of Being Exhaustive. Optimization of Bridging Structural Water Molecules and Water Networks in Models of Biological Systems, *Chem. Biodiv.* **2004**, *1*, 98–105.

[10] P. J. Goodford, A Computational Procedure for Determining Energetically Favorable Binding sites on Biologically Important Macromolecules, *J. Med. Chem.* **1985**, *28*, 857–864.

[11] R. A. Powers and B. K. Shoichet, Structure–Based Approach for Binding Site Identification on AmpC β–Lactamase, *J. Med. Chem.* **2002**, *45*, 3222–3234.

[12] W. E. Minke, D. J. Diller, W. G. J. Hol, and C. L. M. J. Verlinde, The Role of Waters in Docking Strategies with Incremental Flexibility for Carbohydrate Derivatives: Heat–Labile Enterotoxin, a Multivalent Test Case, *J. Med. Chem.* **1999**, *42*, 1778–1788.

[13] W. Bitomsky and R. C. Wade, Docking of Glycosaminoglycans to Heparin–Binding Proteins: Validation for aFGF, bFGF, and Antithrombin and Application to IL–8, *J. Am. Chem. Soc.* **1999**, *121*, 3004–3013.

[14] W. Sippl, Development of Biologically Active Compounds by Combining 3D QSAR and Structure Based Design Methods, *J. Comput.–Aided Mol. Design* **2002**, *16*, 825–830.

[15] M. A. Kastenholz, M. Pastor, G. Cruciani, E. E. J. Haaksma, and T. Fox, GRID/CPCA: A New Computational Tool to Design Selective Ligands, *J. Med. Chem.* **2000**, *43*, 3033–3044.

[16] A. Miranker and M. Karplus, Functionality Maps of Binding Sites: A Multiple Copy Simultaneous Search Method, *Proteins Struct. Funct. Genet.* **1991**, *11*, 29–34.

[17] A. Caflish, A Miranker, and M. Karplus, Multiple Copy Simultaneous Search and Construction of Ligands in Binding Sites: Application to Inhibitors of HIV–1 Aspartic Proteinase, *J. Med. Chem.* **1993**, *36*, 2142–2167.

[18] D. Joseph–McCarthy, A. A. Federov, and S. C. Almo, Comparison of Experimental Functional Group Mapping of an RNase A Structure: Implications for Computer–Aided Drug Design, *Protein Engineer.* **1996**, *9*, 773–780.

[19] J. E. Huheey, Inorganic Chemistry: Principles of Structure and Reactivity, 3rd Edition, Harper & Row, New York, NY, 1983, pp 256–260.

[20] H. Jhoti, O. M. P. Singh, M. P. Weir, R. Cooke, P. Murray–Rust, and A. Wonacott, X–ray Crystallographic Studies of a Series of Penicillin–Derived Asymmetric Inhibitors of HIV–1 Protease. *Biochemistry* **1994**, *33*, 8417–8427.

[21] M. Levitt, and B. H. Park, Water: Now You See It, Now You Don't, *Structure* **1993**, *1*, 223–226.

[22] O. Carugo and D. Bordo, How Many Water Molecules Can be Detected by Protein Crystallography?, *Acta Crystallogr D. Biol Crystallogr.* **1999**, *55*, 479–483.

[23] P. Ashorn, T. J. McQuade, S. Thaisrivongs, A. G. Tomasselli, W. G. Tarpley, and B. Moss, An Inhibitor of the Protease Blocks Maturation of Human and Simian Immunodeficiency Viruses and Spread of Infection, *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 7472–7476.

[24] C. Hansch and A. J. Leo, Substituent Constants for Correlation Analysis in Chemistry and Biology, John Wiley and Sons, Inc. New York, NY, 1979.

[25] R. F. Rekker, The Hydrophobic Fragmental Constant. Its Derivation and Applications. A Means of Characterizing Membrane Systems, Elsevier, New York, NY, 1977.

[26] G. E. Kellogg, G. S. Joshi, and D. J. Abraham, New Tools for Modeling and Understanding Hydrophobicity and Hydropathic Interactions, *Med. Chem. Res.* **1992**, *5*, 444–453.

[27] G. E. Kellogg and D. J. Abraham, KEY, LOCK, and LOCKSMITH. Complementary Hydrophobicity Map Predictions of Drug Structure from a Known Receptor/Receptor Structure from Known Drugs, *J. Mol. Graph.***2000**, *10*, 212–217.