Two Complementary Methods for Relative Quantification of Ligand Binding Site Burial Depth in Proteins: The ‘Cutting Plane’ and ‘Tangent Sphere’ Methods

Vicente M. Reyes, Ph.D.*
E-mail: vmrsbi.RIT.biology@gmail.com

*work done at:
Dept. of Pharmacology, School of Medicine,
University of California, San Diego
9500 Gilman Drive, La Jolla, CA 92093-0636
&
Dept. of Biological Sciences, School of Life Sciences
Rochester Institute of Technology
One Lomb Memorial Drive, Rochester, NY 14623

Abstract. We describe two complementary methods to quantify the degree of burial of ligand and/or ligand binding site (LBS) in a protein-ligand complex, namely, the ‘cutting plane’ (CP) and the ‘tangent sphere’ (TS) methods. To construct the CP and TS, two centroids are required: the protein molecular centroid (global centroid, GC), and the LBS centroid (local centroid, LC). The CP is defined as the plane passing through the LBS centroid (LC) and normal to the line passing through the LC and the protein molecular centroid (GC). The “exterior side” of the CP is the side opposite GC. The TS is defined as the sphere with center at GC and tangent to the CP at LC. The percentage of protein atoms (a.) inside the TS, and (b.) on the exterior side of the CP, are two complementary measures of ligand or LBS burial depth since the latter is directly proportional to (b.) and inversely proportional to (a.). We tested the CP and TS methods using a test set of 67 well characterized protein-ligand structures (Laskowski et al., 1996), as well as the theoretical case of an artificial protein in the form of a cubic lattice grid of points in the overall shape of a sphere and in which LBS of any depth can be specified. Results from both the CP and TS methods agree very well with data reported by Laskowski et al., and results from the theoretical case further confirm that that both methods are suitable measures of ligand or LBS burial. Prior to this study, there were no such numerical measures of LBS burial available, and hence no way to directly and objectively compare LBS depths in different proteins. LBS burial depth is an important parameter as it is usually directly related to the amount of conformational change a protein undergoes upon ligand binding, and ability to quantify it could allow meaningful comparison of protein dynamics and flexibility.

Abbreviations:  CP, cutting plane;  TS, tangent sphere;  CPM, cutting plane method;  TSM, tangent sphere method;  CPi, CP or CPM index;  TSi, TS or TSM index;  LBS, ligand binding site;  GC, global or molecular centroid;  LC, local or LBS centroid;  lig, ligand;  res, residue;  sdc, sidechain;  PPI, protein-protein-interaction/s;  PLI, protein-ligand interaction/s;  PLC, protein-ligand complex

Keywords:  protein-ligand interactions;  ligand burial;  ligand binding site burial;  protein centroid;  cutting plane method;  tangent sphere method;  protein flexibility;  protein dynamics
1 Introduction

This work was largely inspired by that of A. Ben-Shimon and M. Eisenstein in 2005. Depth of ligand burial in its cognate receptor protein is an important property of the latter biomolecule because it provides an indication of protein dynamics: deeply buried ligand binding sites imply a high degree of ligand-induced protein conformational change and hence flexibility (Rauh et al., 2004; Andersen et al., 2000). Several studies have tried to catalogue the degree of burial of active sites and have offered some generalizations relating catalytic activity with degree of ligand burial (Ben-Shimon et al., 2005). Other works focus on detecting crevices and grooves on protein surfaces, as an indirect way of locating ligand binding sites or active sites (Mitchell et al., 2001; Kawabata et al., 2007). In protein-protein complexes, on the other hand, the constituent monomers oftentimes overlap physically in the bound form, resulting from induced fit mechanism facilitated by conformational changes due to protein flexibility (Goh et al., 2004; Bui et al., 2006). There is a limit to such overlap, however: generally, the hydrophobic cores of the individual monomers are very seldom breached (Kimura et al., 2001). To the best of our knowledge, prior to this study there have been no numerical measures of LBS burial available that do not depend in protein size (i.e., relative), and hence no way to directly and objectively compare LBS depths in different proteins, although various recent works have succeeded to some degree (Tan K.P., et al. 2011; Coleman R.G. & Sharp K.A., 2010; Kalidas Y. & Chandra N., 2008; Coleman R.G. & Sharp K.A., 2006; Varrazzo D. et al., 2005; Tan K.P. et al., 2013).

Here we describe two complementary computational procedures we call the “cutting plane” and “tangent sphere” methods. The two methods are generally applicable to protein-ligand interactions (PLI) when the coordinates of residues in the ligand binding site are known, or to protein-protein interactions (PPI) if the coordinates of the residues at the PPI interface are known. In PLI, the methods are used to quantitatively determine the degree of burial of ligand binding sites in proteins (present work; see also and Reyes, V.M., unpublished [b.]), while in PPI, they may be used to quantitatively determine the degree of physical overlap between the monomer partners in the protein complex (Reyes, V.M. (unpublished [c.])). The ‘cutting plane’ is then defined as that normal to the line formed by the global and local centroids, and containing the latter (Figure 1A). Another name for ‘cutting plane’ is ‘secant plane’ but we shall use the former here. The equations of the cutting plane and tangent spheres are derived from the coordinates of LC and GC. From the equation of the CP, the percentage of protein atoms on the ‘external side’ of the CP (side opposite the global centroid, purple area) may be calculated exactly and is directly proportional to the degree of burial of the ligand binding site. Likewise, from the equation of the TS, the percentage of protein atoms inside the TS (green area) may be calculated exactly and is inversely proportional to the degree of burial of the ligand binding site. The CP and TS methods find specific use in a procedure we recently developed regarding the prediction of protein-ligand interactions (Reyes, V.M. and Sheth, V.N., 2011; Reyes, V.M., unpublished [a.]) and protein-protein interaction partners (Reyes, V.M. (unpublished [c.]).

2 Datasets and Methods

2.1 The Test Sets. We have used two test sets in this work: (1.) an artificial protein made up of a 3D cubic lattice of points (whose edge = 1.5 units) in the overall shape of a sphere of radius 50 units, and (2.) a set of 67 representative protein 3D structures used as a dataset in the work of Laskowski et al., 1996. We decided to apply our methods to an artificial protein (above) so that we could study its behavior in as controlled a way as possible.

2.1.1 Creating the Artificial Protein. The artificial protein was a grid of points in 3D space in the shape of a sphere with center at the origin and radius 50.0 units (Angstroms). First a 3D cubic grid of points centered at the origin with edges 100 units long (-50.0 ≤ x ≤ 50.0, -50.0 ≤ y ≤ 50.0, -50.0 ≤ z ≤ 50.0) was generated by running a Fortran program that employs a set of three nested do loops, one each for generating the x-, the y-, and the z-coordinates. The grid of points generated are separated by equal distances of 1.5 Å along the x, y and z directions. Then points lying inside the sphere with equation \( x^2 + y^2 + z^2 ≤ 50.0^2 \) are collected. The LBS entrance is assumed to be at the ‘north pole’, i.e., at the point (0,0,50) and LBS centroids are assumed to all lie along the z-axis from (0,0,50) to (0,0,-50), and equally spaced at a constant separation of 1.0 Å, yielding a
total of 101 LBS’s in all (including the one at the origin). Thus, 101 artificial proteins are created, all in the form of a spherical grid of points centered at the origin with radius 50, but differing in the location of their LBS’s: the first with LBS at (0,0,50), the shallowest LBS; the second with LBS at (0,0,49), the second shallowest LBS; the third with LBS at (0,0,48), the third shallowest LBS; …; the 99th with LBS at (0,0,48), the third deepest LBS; the 100th with LBS at (0,0,49), the second deepest LBS; and the 101th with LBS at (0,0,50), the deepest LBS. These 101 artificial proteins were then analyzed using the CP and TS methods.

2.1.2 The Laskowski Data Set. This data set is composed of 67 single-chain monomeric enzymes with bound ligands, some of which contain single and others multiple bound ligands of various sizes. These proteins together with the ligands they contain are shown in Table 1 of Laskowski, et al., 1996. These proteins were selected from the 1995 PDB release based on their E.C. classifications, such that all E.C. numbers are represented once, the one with the highest resolution and best R-factor being selected. Although all are single-chain, 31 have a single domain, 30 have two domains, five have three and one has four. These 67 test structures contain varying numbers of bound ligands (from none to 11). Each was down into 1:1 protein-ligand complexes. For example, if the protein P in the structure has 4 different ligands W, X, Y and Z bound, it is broken down into four protein-ligand complexes, namely, P-W, P-X, P-Y and P-Z, corresponding to four LBSs. Thus, in all, a total of 184 LBS’s were derived from this set of 67 test protein structures. It is this set of 184 LBS’s that we test in this work.

2.2 The ‘Cutting Plane’ Method. The cutting plane method (together with the tangent sphere method) is illustrated in Figure 1A. From the coordinates of the amino acid residues making up the LBS, the LBS centroid, a local centroid (LC), may be calculated. On the other hand, the protein molecular centroid, the global centroid (GC), may be calculated from the PDB coordinate file. From these two points, the equation of the cutting plane in the form Ax + By + Cz + D = 0 may be determined by application of simple vector algebra as follows:

Let the local and the global centroids be the points L(p,q,r) and G(s,t,u). The CP is normal to line segment LG and contains L. Thus vector \( \overrightarrow{LG} = \langle s-p, t-q, u-r \rangle \) is normal to the CP. For any point Q(x,y,z) on the CP, vector \( \overrightarrow{LQ} = \langle x-p, y-q, z-r \rangle \) must be normal to vector \( \overrightarrow{LQ} \). For any point Q(x,y,z) on the CP, \( \overrightarrow{LQ} \). The equation of the cutting plane is then \( A \times B + C 	imes D = 0 \), where A = s-p, B = t-q, C = u-r and D = p(p-s) + q(q-t) + r(r-u). It can be readily shown that the global centroid is always on the positive side of the CP; hence, the percentage of protein atoms on the exterior side of the CP is readily calculated by setting Ax + By + Cz + D ≥ 0. In Figure 1B, the behavior of the CP index (CPI; the % of protein atoms on the external side of the CP) is monitored as the LBS goes from very shallow to very deep (purple areas); the CPI increases from 0% to 100%.

2.3 The Tangent Sphere Method. The tangent sphere method (together with the cutting plane method) is illustrated in Figure 1A. Knowing the coordinates of the LBS centroid (the local centroid, LC) and the protein molecular centroid (the global centroid), the equation of the tangent sphere can be readily determined since its center is at GC and its radius is the distance between LC and GC. The percentage of protein atoms inside the TS may then be readily calculated by setting \( x^2 + y^2 + z^2 ≤ r^2 \), where r = distance between GC and LC. As with the CP method, the TS method was done in three different ways, namely the ‘sidechain submethod’; the ‘residue submethod’; and the ‘ligand submethod.’ And similarly, results from these three submethods were very similar to each other, implying again that the submethod used does not affect the results. In Figure 1B, the behavior of the TS index (TSi; the % of protein atoms inside the TS) is monitored as the LBS goes from very shallow to very deep (green areas); the TSi decreases from 100% to 0%.

2.4 The CM-CM Method. This method is just an auxiliary method we tried to implement as an independent way to cross-check the TS and CP methods. This method is simply the determination of the distance between the GC and the LC, in Angstroms, and is a value that is expected to be inversely proportional to the LBS depth, just like the TSI. However, unlike CPI and TSI, this method does not take into consideration the percentage of protein atoms relative to the entire protein; hence this method is dependent upon the size of the protein, and is not reliable to use when the proteins being compared vary dramatically in sizes. On the contrary, CPI and TSI...
are directly comparable across proteins of widely varying sizes, since they are %’s of protein atoms relative to
the whole.

For thoroughness, we performed CPM and TSM in three slightly different ways depending on how the LC
coordinates are calculated, namely: (1.) based on the coordinates of the atoms of the amino acid residues that
contact the ligand; we designate this the ‘sidechain submethod’; (2.) based on the coordinates of all the atoms of
the amino acid residues that contact the ligand; we designate this method as the residue submethod’; and (3.)
based on the coordinates of the ligand itself, which we designate as the ‘ligand submethod’. In all cases, results
from these three submethods were very similar to each other, so essentially it may be concluded that the
submethod used does not affect the results (data not shown).

3 Results and Discussion

3.1 Theoretical Case: The Artificial Protein. The behavior of the artificial protein under the CP and TS
methods is illustrated in Figures 2, panels A, B and C. Binding sites lie along the z-axis in 1 unit increments,
and the opening of the LBS is at the top, at point (0,0,50) (Figure 2A). Starting from the upper left-hand corner
and following the small red arrows, we go from the artificial protein with the shallowest LBS at (0,0,50), to the
one with the deepest LBS at (0,0,-50). Midway between these two extremes is the case where the LBS
coincides with the protein molecular centroid at (0,0,0). In the first half of the graph, from the first to the 50th
LBS, we note that TSi starts from 100% and decreases quickly then more slowly until it reaches a minimum at
0% (Figure 2B, blue data points); meanwhile, CPi starts from 0% and increases to a maximum at 50% (Figure
2B, pink data points). The other half of the graph, that from 51st to 101th LBS, is a mirror image of the first
half, from x=1 to x=50. In this half, TSi increases from 0% to 100%, first slowly and then quickly (Figure 2B,
blue data points), while CPi decreases from 50% to 0% (Figure 2B, pink data points). Note that in this case, the
location of the LBS opening is known to be at (0,0,50). In this case, we can define the “anterior half” and the
“posterior half” of the protein. Imagine the CP passing through the molecular centroid and cutting the protein in
half; the half containing the LBS opening is the “anterior half” of the protein, whole the other half (not
containing the LBS opening) is its “posterior half.”

Had the location of the LBS opening not been known, or known but not taken into consideration, there would
have been an ambiguity in the locations of the individual LBS’s as to whether they are in the anterior or
posterior half of the protein. In that case, the proteins with the shallowest to deepest LBS’s would have been
determined based solely on their TS indices (largest to smallest) and CP indices (smallest to largest). Thus the
ambiguity between the anterior and posterior halves would have caused the right half of the plot in Figure 2B
corresponding to the posterior have) to merge with the left half (corresponding to the anterior half) and the plot
would have been that on Figure 2C. The TSM and CPM plots for the 67 test proteins (to be discussed shortly)
in Figures 4A and 4B, respectively, correspond to this situation.

3.2 Application to the Laskowski Protein Dataset. In applying the CP, TS and CM-to-CM methods we set
arbitrary limits for “shallow”, “medium” (intermediate depth) and “deep” LBS predictions. For the CP method,
they are: (%’s refer to % of protein atoms on the external side of CP): 0.0%-10.0% for shallow; 10.1%-20.0%
for medium; and 20.1%-42.0% for deep LBS’s (42% is a close upper bound for the results). For the TS method,
they are ((%’s refer to % of protein atoms inside the TS): 100.0%-40.1% for shallow; 40.0%-10.1% for
medium; and 10.0%-0.0% for deep LBS’s. For the CM-to-CM method, they are (Å’s refer to the distance
between the global centroid, GC, and the local centroid, LC): 38.0Å - 15.1Å for shallow; 15.0 Å-10.1Å for
medium; and 10.0Å -0.0Å for deep LBS’s. These were the best limit values as judged from the frequency
distributions obtained, as discussed below.

3.2.1 The Tangent Sphere Method as Applied to the Laskowski Dataset. We first determined the
frequency distribution of the prediction results (i.e., “shallow”, “intermediate” and “deep”) from the TS method,
using the ligand, residue and sidechain submethods described earlier. From the distributions (Figure 3A.1, 3A.2
and 3A.3, respectively), it may be concluded that the TS method has more resolving power at the lower bins
(shallow LBS’s) than at the higher bins (deep LBS’s), indicating that it has higher specificity for shallow LBS’s
than to deep LBS’s.
Figure 4A shows the results from the TS method from the 3 submethods: ligand (blue data points), residue (pink data points) and sidechain (yellow data points) submethods. All show the same trend in going from shallow to deep LBS’s: decreasing rather quickly in the first half, then more slowly on the second half, almost plateauing near 0% towards the deepest LBS’s. We note that this is very similar to the behavior of the artificial protein in Figure 2C (blue data points) and the first half of the plot in Figure 2B (blue data points).

Let us go back to the concept of “anterior” and “posterior” halves of a protein. These are both illustrated in the central object in Figure 1B. In the case of the artificial protein, the researcher knows the exact location of the LBS in relative to the LBS opening, hence it is possible to accurately arrange the proteins from those that have shallowest LBS’s (in the anterior half) to those that have deepest LBS’s (in the posterior half). In actual work as well in our application to the Laskowski dataset, however, the LBS opening is either unknown or known but not taken in consideration due to automation of the process. Thus in these cases, it is not possible to differentiate the LBS’s lying in the anterior half from those that lie in the posterior half. Thus there is such ambiguity in the case of the Laskowski test proteins (where the LBS opening may have been known, but not taken it into account because of automation).

3.2.2 The Cutting Plane Method as Applied to the Laskowski Dataset. As in the TS method, we determined the frequency distribution of the prediction results (i.e., “shallow”, “intermediate” and “deep”) using the ligand, residue and sidechain submethods described earlier. From the distributions (Figure 3B.1, 3B.2 and 3B.3, respectively), it may concluded that the CP method has more resolving power at the higher bins (deep LBS’s) than at the lower bins (shallow LBS’s), indicating that it has higher specificity for deep LBS’s than to shallow LBS’s. This is opposite behavior to that of the TS method.

Figure 4B shows the results from the CP method from the 3 submethods: ligand (blue data points), residue (pink data points) and sidechain (yellow data points) submethods. All show the same trend in going from shallow to deep LBS’s: increasing rather slowly at first, then faster and almost linearly thereafter, towards the deeper LBS’s, and then stops at between 39% and 44% at the deepest LBS’s. We note that this is very similar to the behavior of the artificial protein in Figure 2C (pink data points) and the first half of the plot in Figure 2B (pink data points).

As was explained in the previous section, there is an ambiguity in the latter case. However, as can be seen from Figure 4A, the CP results do not reach 50 but stops between 39%, 40% and 44% for the sidechain, ligand and residue submethods, respectively. This means that the LBS’s do not reach the depth of the protein molecular centroid. Therefore it is reasonable to conclude that there are no LBS’s on the posterior half of the protein, i.e., all LBS’s are in the anterior half. This must be so since it is quite improbable that no LBS’s are found on and around the global centroid while LBS’s on the posterior half exist. As for the case of the TS method in Figure 4B, we see that it does reach 0% near the deepest LBS’s (0.0%, 0.15% and 0.25% for the ligand, residue, and sidechain methods, respectively). However, as was discussed earlier, the TS method has low resolving power for the deep LBS’s, and in fact the curve plateaus near at 0% near the deepest LBS’s. The TS method seems unable to differentiate between LBS depths when the TS reaches a radius of within 5 Å from the global centroid. This appears to validate our conclusion that there are probably no LBS’s on the posterior halves of the 67 test proteins.

3.2.3 The CM-to-CM Method. Figure 3C.1, 3C.2 and 3C.3 show the frequency distribution of the results from the CM-to-CM method, using the ligand submethod, residue submethod and sidechain submethod, respectively. Here, the horizontal axis is the distance in Angstroms between the local CM and the global CM (“CM-to-CM value”), going from shallowest LBS at 38 Å on the extreme left, to deepest LBS at 0 Å on the extreme right, partitioned into intervals of width 1.9 Å, for a total number of intervals (“bins”) of 20. The vertical axis is the number of LBS’s (out of the 184 total) with CM-to-CM values corresponding to each interval or bin along the x-axis. For example, in the CM-to-CM method, ligand submethod, there is 1 LBS with a CM-to-CM value between 38 Å and 36.1 Å (first bin), while there are no LBS’s with a CM-to-CM value between 1.9 Å and 0% (last bin). Note that the resolving power of the CM-to-CM method is highest on the first half of the plot (shallow LBS’s), dipping to a minimum about halfway through the second half, and then increases again
towards the end of the second half (very deep LBS’s), indicating that it has higher specificity for shallow to almost-shallow LBS’s and for very deep LBS’s.

Figure 4C shows the results from the CM-to-CM method from the 3 submethods: ligand (blue data points), residue (pink data points) and sidechain (yellow data points) methods. All show the same trend in going from shallow to deep LBS’s: decreasing rather quickly in the first quarter, then more slowly on the second and third quarters, then more slowly again and almost plateauing near 0% towards the deepest LBS’s. That there are most probably no LBS’s on the posterior halves of the 67 test proteins is also supported by these CM-to-CM data since the LC to GC distance does not reach 0 Å (3.0 Å, 1.9 Å and 2.4 Å for the ligand, residue and sidechain submethods, respectively). This implies, as in the CP method, that no LBS is close to the GC, hence there are probably no LBS’s in the posteriori half of the proteins.

3.3 Summary of Final Results. From the 67 Laskowski dataset of protein structures, there are a total of 184 LBS’s since some of the proteins in the dataset have more than one ligand (from as low as none to as many as 11; see Table 1 and legend), as mentioned earlier. The determination of the mutual intersections of the shallow, medium (intermediate depth) and deep predictions by the TSM, CPM and CM-CM method, using the 3 submethods (ligand, residue and sidechain submethods), respectively, have been determined, and the results are shown and summarized in the Venn Diagrams in Figure 5, panels A, B and C for the TSM, CPM and CM-CM methods, respectively, and similarly in Figure 6, panels A, B and C. The overall results from the TSM, CPM and CM-CM are presented in combined form in Figure 7, Panel A while in Panel B, only the combined results from the TSM and CPM are presented (with the CM-CM results excluded).

Specifically, Figure 5A, B and C show the first step in the analysis process, which is the pairwise determination of the intersections among the three submethods (ligand, residue and sidechain) in the TSM, CPM and CM-CM methods, respectively. Meanwhile, Figure 6A, B and C show the second step in the process, which is the determination of the mutual 3-way intersections among the three submethods (ligand, residue and sidechain) in the TSM, CPM and CM-CM methods, respectively. Figure 7 shows the final step in the analysis process. In Panel A, the determination of the mutual three-way intersections among the TSM, CPM and CM-CM methods are shown. In this case, 37 structures are common to the TS, CP and CM-to-CM shallow LBS predictions, 8 are common to the medium LBS predictions, and finally 47 are common to the deep LBS predictions. The identities of these 37 shallow, 8 medium and 47 deep LBS’s are shown in Table 2. Thus a total of 37+8+47 = 92 LBS’s out of the 184 LBS’s, or 50%, were predicted identically by our three methods, TS, CP and CM-CM. Panel B is the same as Panel A except that the CM-CM results are excluded, as this method is disfavored relative to the TSM and CPM methods because it (the CM-CM method) is dependent on protein size, while the TSM and CPM are not. In this case, 38 structures are common to the TSM and CPM shallow LBS predictions, 11 are common to the medium LBS predictions, and finally 51 are common to the deep LBS predictions. The identities of these 38 shallow, 11 medium and 51 deep LBS’s may be gleaned from Table 2. Thus a total of 38+11+51 = 100 LBS’s out of the 184 LBS’s, or 54%, were predicted identically by the two methods about which this work is about, the TS and CP methods. Note that an additional 1, 3 and 4 LBSs are included in the shallow, medium and deep predictions, respectively compared to the previous tabulation. These additional LBSs represent the ones which were supposed to have been incorrectly predicted by the CM-CM method due to its dependence on protein size.

Finally in Table 3, panels A, B and C, we show the identities and the corresponding CPi and TSi values of the final predictions in Figure 7A and B. Panel A show the deep predictions, panel B the medium predictions, and panel C the shallow predictions. Note that the CPi’s of the LBSs in panels A, B and C decrease in that order, while the TSi’s increase in the same order. These results demonstrate the complementarity between these two LBS burial depth metrics. To the best of our knowledge, prior to this study, no such (numerical) metrics existed for LBS burial depth. With these two new metrics (TSi and CPi), LBS burial depth in different proteins, regardless of size variations among them, may be compared.

The results we obtained above are very similar to the results described by Laskowski et al.. (1996), wherein the ligand depths were determined by manual (visual) inspection. Hence we conclude that the Cutting Plane (CP) and Tangent Sphere (TS) methods are quite effective and reliable methods of determining and quantifying
ligand or LBS burial in proteins, provided that the coordinates of the LBS (local site) and that of the entire protein molecule (global structure) are available.

3.4 Advantages and Limitations of the Methods. Prior to this study, there were no numerical measures of LBS burial available that do not depend in protein size, and hence no way to directly and objectively compare LBS depths in different proteins. The CP and TS methods are two complementary methods that provide such a way. The fact that the TS and CP indices are complementary does not mean that they are redundant in the sense that if one is known, the other may be precisely calculated; that is only true for a perfectly spherical protein, but real proteins are never perfectly spherical. That being said, the CP and TS methods would work best if the test protein is roughly spherical, or globular, in shape. If the protein has an overall concavity, or is otherwise irregular in shape, the method becomes less accurate. Thus, we can expect the CP and TS methods to be less accurate for rod-shaped proteins, etc. But since the majority of structurally solved proteins (about 80% of them) are globular (“spheroproteins”), it may be concluded that the CP and TS methods are generally applicable. The complementarity of the two methods (i.e., the CP index is directly proportional to LBS burial depth, while the TS index is inversely proportional to it) is an advantage of employing both methods in parallel since they cross-validate each other. Although TSM and CPM were created out of a need for validating our own results from our prediction studies (Reyes, V.M., unpublished [a.], [b.] and [c.]), it is generally applicable to any protein whose atomic coordinates, as well as those of the relevant local sites in it (i.e., LBSs), are known.

4 Conclusions.

We have developed a couple of complementary methods, as well as an auxiliary method, to quantify the depth of burial of a ligand or its binding site in a protein structure. These methods, called the “Cutting Plane,” the “Tangent Sphere” and the “Centroid-to-Centroid” methods, respectively, are all implemented as Fortran 77/90 programs and are fast, efficient and amenable to high-throughput or big-batch applications. We tested the methods on a set of 67 previously characterized structures containing a combined total of 184 LBS’s, as well as on an artificial protein consisting of a grid of points in the shape of a sphere, where LBS’s can be arbitrarily assigned anywhere on it. Predictions on the LBSs by the three methods in the form of “shallow”, “medium” (intermediate depth) and “deep” are in line with previous characterizations done by human visual inspection. The predictions are also independent of protein size, as they are relative to the size of the entire protein (based on percentages of portions of protein relative to the whole). We believe that these methods may find useful applications in structural proteomics, especially protein dynamics, as ligand burial depth is indicative of the amount of conformational changes the protein undergoes in binding the ligand. To the best of our knowledge, this work is the first attempt at the relative quantification of the depth of burial of ligand or ligand binding sites in proteins using cutting planes and tangent spheres as described in this work.

Acknowledgments. This work was supported by an Institutional Research and Academic Career Development Award to the author, NIGMS/NIH grant number GM 68524. The author thanks Ms. Srujana Cheguri, one of his M.S. Bioinformatics graduate students at Rochester Institute of Technology (RIT), for verifying results. The author wishes to acknowledge the San Diego Supercomputer Center, the UCSD Academic Computing Services, and the UCSD Biomedical Library, for the help and support of their staff and personnel. He also acknowledges the Division of Research Computing at RIT, and computing resources from the Dept. of Biological Sciences, College of Science, at RIT.

References:
Andersen JF, Ding XD, Balfour C, Shokhireva TK, Champagne DE, Walker FA, Montfort WR. “Kinetics and equilibria in ligand binding by nitrophorins 1-4: evidence for stabilization of a nitric oxide-ferriheme complex through a ligand-induced conformational trap.” Biochemistry. 2000 Aug 22;39(33):10118-31.
Ben-Shimon, A., and Eisenstein, M. 2005. “Looking at enzymes from the inside out: the proximity of catalytic residues to the molecular centroid can be used for detection of active sites and enzyme-ligand interfaces.” J Mol Biol. 351:309-26.

Bui JM, McCammon JA. “Protein complex formation by acetylcholinesterase and the neurotoxin fasciculin-2 appears to involve an induced-fit mechanism.” Proc Natl Acad Sci U. S. A. 2006 Oct 17;103(42):15451-6. Epub 2006 Oct 4.

Coleman RG, Sharp KA. (2010) Protein pockets: inventory, shape, and comparison. J Chem Inf Model. 2010 Apr 26;50(4):589-603. doi: 10.1021/ci900397t.

Coleman RG, Sharp KA. (2006) Travel depth, a new shape descriptor for macromolecules: application to ligand binding. J Mol Biol. 2006 Sep 22;362(3):441-58. Epub 2006 Jul 15.

Goh CS, Milburn D, Gerstein M. “Conformational changes associated with protein-protein interactions.” Curr Opin Struct Biol. 2004 Feb;14(1):104-9.

Goodman, JE & O’Rourke, J. (eds.) In: “Handbook of Computational Geometry”, 2nd ed., © 2004, Chapman & Hall Pub. Co., Boca Raton, FL

Kalidas Y, Chandra N. (2008) PocketDepth: a new depth based algorithm for identification of ligand binding sites in proteins. J Struct Biol. 2008 Jan;161(1):31-42. Epub 2007 Sep 15.

Kawabata T, Go N. “Detection of pockets on protein surfaces using small and large probe spheres to find putative ligand binding sites.” Proteins. 2007 Apr 19; [Epub ahead of print]

Kimura SR, Brower RC, Vajda S, Camacho CJ. “Dynamical view of the positions of key side chains in protein-protein recognition.” Biophys J. 2001 Feb;80(2):635-42.

Laskowski, RA, Luscombe, NM, Swindells MB and Thornton JM (1996) “Protein clefts in molecular recognition and function”, Protein Sci., 5: 2438-2452

Mitchell, J.C., Kerr, R. and Ten Eyck, L.F. (2001) “Rapid Atomic Density Methods for Molecular Shape Characterization,” J. Mol. Graph. Model, 19(3): 324-329.

Rauh D, Klebe G, Stubbs MT. “Understanding protein-ligand interactions: the price of protein flexibility.” J Mol Biol. 2004 Jan 30;335(5):1325-41.

Reyes, V.M. & Sheth, V.N., "Visualization of Protein 3D Structures in 'Double-Centroid' Reduced Representation: Application to Ligand Binding Site Modeling and Screening", 2011, Handbook of Research in Computational and Systems Biology: Interdisciplinary Approaches, IGI-Global/Springer, pp. 583-598.

Reyes, V.M. (unpublished [a.]) "An Automatable Analytical Algorithm for Structure-Based Protein Functional Annotation via Detection of Specific Ligand 3D Binding Sites: Application to ATP Binding Sites in ser/thr Protein Kinases and GTP Binding Sites in Small Ras-type G-Proteins" (forthcoming); Abstract: J. Biomol. Struct. & Dyn., Book of Abstracts, Albany 2009: The 16th Conversation, June 16-20, 2009, Vol. 26 (6) June 2009, p. 874

Reyes, V.M. (unpublished [b.]) “Structure-Based Prediction of Sialic Acid, Retinoic Acid and Heme-Bound and Unbound Nitric Oxide Protein Binding Sites” (forthcoming); Abstract: J. Biomol. Struct. & Dyn., Book of Abstracts, Albany 2009: The 16th Conversation, June 16-20, 2009, Vol. 26 (6) June 2009, p. 874

Reyes, V.M. (unpublished [c.]) “A Novel, Automatable and Analytical (Exact) Method for Structure-Based Prediction of Protein-Protein Interaction Partners” (forthcoming); Abstract: J. Biomol. Struct. & Dyn., Book of Abstracts, Albany 2009: The 16th Conversation, June 16-20, 2009, Vol. 26 (6) June 2009, p. 873
Tan KP, Nguyen TB, Patel S, Varadarajan R, Madhusudhan MS. (2013) Depth: a web server to compute depth, cavity sizes, detect potential small-molecule ligand-binding cavities and predict the pKa of ionizable residues in proteins. Nucleic Acids Res. 2013 Jul;41(Web Server issue):W314-21. doi: 10.1093/nar/gkt503. Epub 2013 Jun 12.

Tan KP, Varadarajan R, Madhusudhan MS. (2011) DEPTH: a web server to compute depth and predict small-molecule binding cavities in proteins. Nucleic Acids Res. 2011 Jul;39 (Web Server issue):W242-8. doi: 10.1093/nar/gkr356. Epub 2011 May 16.

Tronrud DE. (1997) “TNT refinement package.” Methods Enzymol. 277:306-19.

Varrazzo D, Bernini A, Spiga O, Ciutti A, Chiellini S, Venditti V, Bracci L, Niccolai N. (2005) Three-dimensional computation of atom depth in complex molecular structures. Bioinformatics. 2005 Jun 15;21(12):2856-60. Epub 2005 Apr 12.

**FIGURE LEGENDS:**

**FIGURE 1, Panel A:** Definition and Illustration of the Cutting Plane and Tangent Sphere Methods
Illustration of the Cutting Plane and Tangent Sphere Methods and how they complement each other. The brown circle represents a spherical protein, with a ligand binding site whose opening is at the top; the ligand is shown as a small red circle. The blue horizontal line is the cutting plane (CP), cutting the protein through its ligand binding site (LBS) centroid. This centroid is different from the protein molecular centroid, shown at the center of the protein as a ©: the former is a local centroid, while the latter is a global centroid. The tangent sphere (TS) is shown as a green circle tangent to the CP at the LBS centroid. The exterior side of the CP is defined to be the side opposite the protein molecular centroid, while its interior side is that containing the protein molecular centroid. The volume of the protein lying on the CP exterior side is shown in purple, while the volume of the protein lying inside the TS, on the interior side of the CP, is shown in green.

**Panel B:** Illustration of the CP and TS Methods on an Artificial Protein in the form of a Sphere. To start with, the protein is shown with a shallow ligand binding site. The CP (blue line) and TS (green circle) are shown as well, and the volume of the protein lying on the exterior of the CP and in the interior of the TS are shown in purple and green, respectively. These volumes, expressed as percentages of the total protein volume, are also shown above (%CP) and below (%TS) the cutting plane, respectively; we designate them as the CP and TS indices, respectively. In the succeeding panels, the LBS is made deeper and deeper into the protein; the behavior of %CP and %TS follow two different trends and are discussed in the text and plotted in Figure 3. In both panels, the V-shaped concavity in the circle represents the LBS, but shown only for purposes of illustration. In all our calculations, the receptor protein is conceptualized as a perfect sphere without any such concavity embodying the LBS.

**FIGURE 2:** Plots of CPM and TSM Results for the Theoretical Case ("Artificial Protein")
**Panel A:** The artificial protein is shown as a spherical grid of points with center at the origin and the LBS assumed to be at the “top” i.e., at its maximum point along the positive z-axis. Plots of the CP and TS indices for this case where the LBS opening is precisely known is shown in Panel B. Plots of the CP and TS indices for the case where the LBS opening is not known is shown in Panel C.

**Panel B:** %CP and %TS data from the artificial protein as LBS is made to vary from very shallow (near the surface), to shallow, to very deep (near the protein centroid), to deepest (near the other end of the protein opposite the LBS opening). The %CP (pink data points) and the %TS (dark blue data points) were plotted against the LBS depth, with the shallowest on the far left and the deepest on the far right. LBS depth exactly midway between the shallowest and deepest depths corresponds to the situation where the LBS is at or very near the protein molecular centroid. See text for a discussion of the %CP and %TS trends as LBS depth varies.

**Panel C:** This is what the plot in Panel A would look like if the LBS opening is unknown in the case of the artificial proteins. The right half of the plot will fold into the left half due to the resulting ambiguity between the
anterior and posterior halves of the proteins (see text). This figure shows a basic difference between the artificial proteins and the 67 test proteins. In the artificial proteins the location of the LBS opening is known, and there are LBS’s on both the anterior and posterior half of the protein. In the test (real) proteins, the location of the LBS opening is not known, hence there is ambiguity between the anterior half and the posterior half of the proteins in terms of the location of the LBS’s. However, data from the CP and CM-to-CM methods indicate that there are no LBS’s on the posterior half of the proteins; data from the TS method does not disagree with such conclusion.

**Figure 3:** Frequency Distribution Plots for CPM, TSM and CM-CM Methods for the 67 Test Structures
Panels A1, A2 and A3: Frequency distribution of the classification results obtained from the Tangent Sphere method using the three submethods: ligand CM (Panel A1), residue CM (Panel A2), and side chain CM (Panel A3). In the TS method, the LBS depth is measured in terms of the fraction of protein volume lying inside the tangent sphere, and the range of this parameter is (theoretically and experimentally) from 1% to 100% (see text). Each bin or interval on the horizontal axis is 5% in all 3 panels, giving rise to 20 bins. The vertical axes measure the number of structures in each particular bin. Note that TSM has higher resolving power for shallower LBSs than for deeper LBSs.

Panels B1, B2 and B3: Frequency distribution of the classification results obtained from the Cutting Plane method using the three submethods: ligand CM (Panel B1), residue CM (Panel B2), and side chain CM (Panel B3). In the CP method, the LBS depth is measured in terms of the fraction of protein volume on the external side of the cutting plane, and the range of this parameter is theoretically from 0% to 50%, but experimentally the upper bound only goes to 42% (see text). Each bin or interval on the horizontal axis is 2.1% in all 3 panels, giving rise to 20 bins. The vertical axes measure the number of structures in each particular bin. Note that CPM has higher resolving power for deeper LBSs than for shallower LBSs.

Panels C1, C2 and C3: Frequency distribution of the classification results obtained from the CM-CM method using the three submethods: ligand CM (Panel C1), residue CM (Panel C2), and side chain CM (Panel C3). In the CM-to-CM method, the LBS depth is measured in terms of the distance between the local and the global centroids (LBS and protein CM’s, respectively), and the range of this parameter is from 0 Å (local and global CM’s coincide), and there is no theoretical upper bound, but in this data set, the upper bound is 38 Å (see text). Each bin or interval on the horizontal axis is 1.9 Å in all 3 panels, giving rise to 20 bins. The vertical axes measure the number of structures in each particular bin. Note that the maximal resolving power of the CM-CM method is intermediate to those of the TS and CP methods.

**Figure 4:** Plots of CPM and TSM Results for the 67 Test Structures
Panel A: Results of LBS depth quantification using the CP method done using the 3 submethods, ligand CM (black), residue CM (pink) and side chain CM (yellow). LBS depth is directly proportional to % protein atoms on the external side of the CP, with a lower bound of 0%, and an upper bound of about 42%. Along the horizontal axis are the 184 protein-ligand pairs from the Laskowski dataset arranged from those with lowest CP index (leftmost) to those with highest CP index (rightmost).

Panel B: Results of LBS depth quantification using the TS method done using the 3 submethods, ligand CM (black), residue CM (pink) and side chain CM (yellow). LBS depth is inversely proportional to % protein atoms lying inside the TS, with a lower bound of about 3%, and an upper bound of about 42%. Along the horizontal axis are the 184 protein-ligand pairs from the Laskowski dataset arranged from those with highest TS index (leftmost) to those with lowest TS index (rightmost).

Panel C: Results of LBS depth quantification using the CM-CM method done using the 3 submethods, ligand CM (black), residue CM (pink) and side chain CM (yellow). LBS depth is inversely proportional to the distance between the global and the local centroids, and in this data set, this parameter has a lower bound of about 2 Å and an upper bound of about 38 Å. Along the horizontal axis are the 184 protein-ligand pairs from the Laskowski dataset arranged from those with largest CM-CM distance (leftmost) to those with smallest CM-CM distance (rightmost).

**Figure 5A,B,C: Summary Two-Way Venn Diagrams.** Nine Venn Diagrams showing pairwise intersections of the results from the ligand, residue and side chain submethods for those structures classified as “deep”, “medium (intermediate depth), and “shallow.” Data in panel A are from the TS method, those in panel B from the CP method, and those in panel C from the CM-CM method.
Figure 6A,B,C: Summary Three-Way Venn Diagrams. Three Venn Diagrams showing the mutual (three-way) intersections of the results from the ligand, residue and side chain submethods for those structures classified as “deep”, “medium”, and “shallow”. Data were derived from those in Figure 5, panels A, B and C. The entries in the center of each Venn Diagram are the 3-way intersections among the three submethods (ligand, residue and side chain). Data in panel A are from the TS method, those in panel B from the CP method, and those in panel C from the CM-CM method.

Figure 7, Panel A: Venn Diagrams Illustrating Commonalities of Prediction by the CP, TS and CM-CM Methods. The final results from the TS, CP and CM-CM methods of determining LBS burial depth is shown. The top Venn Diagram shows the intersections of the “shallow” predictions by these three methods, and we see that 37 are common among them. The lower left Venn Diagram shows the intersections of the “medium” (i.e., intermediate LBS depth) predictions of these three methods, and we see that 8 are common among them. The lower right Venn Diagram shows the intersections of the “deep” predictions of these three methods, and we see that 47 are common among them. In all 92 cases (=37+8+47), the predictions were identical with those of Laskowski et al. (1996), whose assessments were derived by inspection (visual).

Panel B: If the CM-CM results are not considered, we arrive at the Venn Diagram shown here. The common intersection will then be 100 (=38+11+51) instead of 92. The CM-CM method is oftentimes best left out because it is dependent on protein size, since it measures absolute distances in Angstroms instead of percentages of atoms in the protein (relative to that in the entire molecule) as the TS and CP indices do.

TABLE LEGENDS:

Table 1: Tabulation of the 67 Test Structures from the Laskowski Dataset. Many of the 67 structures (column 1) have more than one bound ligand; each was broken down into 1:1 protein-ligand pairs. For example, consider a structure that has a single protein chain P, with 3 bound ligand molecules, A, B and C. This structure is broken down into P-A, P-B and P-C, and the 1:1 protein-ligand pairs are designated 1, 2 and 3. If a structure contains a single ligand, the protein-ligand pair is designated “0”. These are the entries in the third column of the table. Broken down in this way, the 67 structures in the Laskowski data set gave rise to 184 protein-ligand pairs. The second column of the table are the short-hand designations of the 67 structures. For example, “Bl.lig4” would refer to structure with PDB ID “1ELA” and its ligand “ACY” in chain B, residue number 300; “Bllbs4” would mean the same, except that the ligand binding site, not the ligand, is the entity on question.

Table 2: LBS’s Judged “Shallow”, Medium depth”, and “Deep” by the Three Methods. These are the identities of the 1:1 protein-ligand pairs classified as “shallow”, “medium”, and “deep” by the 3 methods as shown in the center of each Venn Diagrams in Figure 6, panels A B and C, respectively. The first 3 columns show the results from the TS method, the next three from the CP method, and the last three from the CM-CM method. Each column is the mutual (three-way) intersection of the ligand, residue and side chain submethods for each pertinent classification (deep, medium and shallow) and method (TS, CP and CM-CM).

Table 3A,B,C: TS and CP Indices of Common Predictions. The TS indices (first 4 columns), CP indices (next 4 columns) and CM-to-CM distances (last 4 columns) of the Laskowski structures classified as “deep” (panel A), “medium” (panel B), and “shallow” (panel C). In each panel, the three indices from the three submethods (ligand, “lig”; residue, “res”; side chain, “sdc”; average, “ave”) are shown in black, and their averages are shown in red. The numbers in blue (8 in all) in the average column of the CM-CM section are the ones that become excluded if the CM-CM method is not considered (see Figure 7A,B), for reasons discussed previously. These indices (the TS and CP indices, in particular.) are precisely the quantitative measures of LBS burial that are referred to in the text. Prior to this study, there were no such numerical measures of LBS burial available, and hence no way to directly and objectively compare LBS depths in different proteins.
FIGURES:

Figure 1A
Figure 2A

Artificial Protein:
cubic 3D lattice of points w/ overall spherical shape subjected to TSM and CPM

Figure 2B

Theoretical Case (Artificial Protein)

LBS at protein centroid

TSM results
CPM results
Figure 2C

Figure 3A.1
Figure 3A.2

TSM data (residue CM submethod)

Figure 3A.3

TSM data (sidechain CM submethod)
Figure 3B.1

Figure 3B.2
Figure 3B.3

Figure 3C.1
Figure 3C.2

Figure 3C.3
Figure 4A

Figure 4B
Figure 4C

![Local CM to Global CM Lengths](image)

Figure 5A

![Venn Diagrams](image)
Figure 5B

Figure 5C
Figure 6A
CP method:

Figure 6B
Figure 6C
Figure 7A  (CM-to-CM method included)
Figure 7B  (CM-to-CM method excluded)

TABLES:

Table 1 (part 1 of 4)
| Structure PDB ID | Shorthand designation | no. of domains | class | LIG # or LBS # | Ligand Name |
|-----------------|----------------------|---------------|-------|---------------|-------------|
| 1ADS            | Aa                   | 1             | 1     | 0             | NAP-350     |
| 9ICD            | Ab                   | 1             | 1     | 0             | NAP-1       |
| 1IPD            | Ac                   | 1             | 1     | 1             | SO4-346     |
| 1GOX            | Ad                   | 1             | 1     | 2             | SO4-347     |
| 1TDE            | Ae                   | 2             | 1     | 0             | FMN-1       |
| 1OYB            | Af                   | 1             | 1     | 1             | FAD-500     |
| 2NPX            | Ag                   | 3             | 1     | 1             | FMN-401     |
| 1CCA            | Ah                   | 2             | 1     | 2             | HBA-402     |
| 1ARP            | Ai                   | 2             | 1     | 3             | NAD-818     |
| 1PBE            | Aj                   | 2             | 1     | 3             | CYC-42      |
| 1HMY            | Ak                   | 2             | 1     | 1             | HEM-1       |
| 3CLA            | Al                   | 1             | 1     | 1             | HEM-345     |
| 1GPB            | Am                   | 1             | 1     | 2             | NAG-348     |
| 1ULA            | An                   | 1             | 1     | 3             | NAG-349     |
| 1STO            | Ao                   | 1             | 1     | 1             | FAD-396     |
| 2YHK            | Ap                   | 3             | 1     | 2             | PHB-396     |
| 1PHP            | Aq                   | 2             | 2     | 3             | SAM-328     |
| 1GKY            | Ar                   | 2             | 1     | 1             | CLM-221     |
| 2CUT            | As                   | 1             | 1     | 2             | CO-222      |
| 1THG            | At                   | 1             | 1     | 1             | PLP-999     |
| 1RPA            | Au                   | 1             | 1     | 1             | SO4-290     |
| 1RNH            | Av                   | 1             | 3     | 2             | SO4-291     |
| 1ONC            | Aw                   | 1             | 2     | 1             | OMP-216     |
| 1FUT            | Ax                   | 1             | 2     | 0             | OH-214      |
| 1RGP            | Ay                   | 1             | 1     | 0             | ADP-365     |
| 1SNC            | Az                   | 1             | 1     | 2             | MG-1        |
|                 |                      |               |       | 1             | 5GP-187     |
|                 |                      |               |       | 2             | SO4-188     |
|                 |                      |               |       | 3             | DEP-401     |
|                 |                      |               |       | 4             | NAG-990     |
|                 |                      |               |       | 5             | NAG-991     |
|                 |                      |               |       | 6             | NAG-994     |
|                 |                      |               |       | 7             | NAG-996     |
|                 |                      |               |       | 8             | TAR-343     |
|                 |                      |               |       | 9             | NAG-344     |
|                 |                      |               |       | 10            | NAG-347     |
|                 |                      |               |       | 11            | MSE-47      |
|                 |                      |               |       | 12            | MSE-142     |
|                 |                      |               |       | 13            | SO4-156     |
|                 |                      |               |       | 14            | SO4-105     |
|                 |                      |               |       | 15            | 2GP-108     |
|                 |                      |               |       | 16            | C2P-126     |
|                 |                      |               |       | 17            | PTP-143     |
|                 |                      |               |       | 18            | CA-142      |

Table 1 (part 2 of 4)
| 1CDG  | Ba  | 4   | 1   | 1  |
|-------|-----|-----|-----|----|
| 2     | MAL-888 |
| 3     | MAL-889 |
| 4     | MAL-660 |
| 5     | GLC-496 |
| 6     | GLC-497 |
| 7     | GLC-499 |
| 8     | SO4-860 |
| 9     | SO4-101 |
| 10    | DAN-800 |
| 11    | BUT-216 |
| 12    | GLC-216 |
| 13    | GLC-217 |
| 14    | FMP-301 |
| 15    | ZN-488 |
| 16    | ZN-486 |
| 17    | LEU(I)-1 |
| 18    | FOR(I)-2 |
| 19    | VAL(I)-3 |
| 20    | VAL(I)-4 |
| 21    | ASP(I)-5 |
| 22    | LOF-309 |
| 23    | ZN-308 |
| 24    | HIN-247 |
| 25    | NAS(I)-1 |
| 26    | APH(I)-3 |
| 27    | PIP(I)-4 |
| 28    | GLY(I)-2 |
| 29    | SO4-1 |
| 30    | SO4-2 |
| 31    | TFA(B)-258 |
| 32    | ISO(B)-259 |
| 33    | SO4(B)-260 |
| 34    | ACY(B)-300 |
| 35    | LYS(B)-257 |
| 36    | PRO(B)-258 |
| 37    | ALM(I)-5 |
| 38    | MSU(I)-1 |
| 39    | ALA(I)-2 |
| 40    | ALA(I)-3 |
| 41    | PRO(I)-4 |
| 42    | PRO(C)-1 |
| 43    | ALA(C)-2 |
| 44    | PRO(C)-3 |
| 45    | PHE(C)-4 |
| 46    | ALA(D)-5 |
| 47    | ALA(D)-6 |
| 48    | PHA(P)-1 |
| 49    | ACE(P)-5 |
| 50    | PRO(P)-4 |
| 51    | ALA(P)-3 |
| 52    | PRO(P)-2 |

Table 1 (part 3 of 4)
|   |   |   |   |
|---|---|---|---|
| 1PI| Bp | 2 | 1 |
|   |   |   |   |
| 1SMR | Bq | 2 | 1 |
|   |   |   |   |
| 1PPI | Br | 2 | 1 |
|   |   |   |   |
| 3APR | Bs | 2 | 1 |
|   |   |   |   |
| 1EPM | Bt | 2 | 1 |
|   |   |   |   |
| 1MPP | Bu | 2 | 1 |
|   |   |   |   |
| 1HYT | Bv | 2 | 1 |
|   |   |   |   |
| 1IAG | Bw | 1 | 1 |
|   |   |   |   |
| 1ADD | Bx | 1 | 3 |
|   |   |   |   |
| 2DHC | By | 1 | 3 |
|   |   |   |   |
| 1PII | Bz | 2 | 1 |
|   |   |   |   |
| 2DKB | Ca | 2 | 1 |

Table 1 (part 4 of 4)
| 1C7H  | Cb  | 1   | 1   | 1   | AMX-700 |
|-------|-----|-----|-----|-----|---------|
| POR   | Cc  | 1   | 1   | 2   | OAA-702 |
| 1CIL  |Cd   | 3   | 3   | 1   | OTE-546 |
| 8ACN  |Ce   | 3   | 3   | 2   | OTE-547 |
| 5ENL  |Cf   | 2   | 1   | 1   | OTE-548 |
| 1PDA  |Cg   | 3   | 1   | 0   | ETS-262 |
| 1MNS  |Ch   | 2   | 1   | 2   | ZN-261  |
| 3PGM  |Ci   | 1   | 1   | 1   | NIC-755 |
| 1BIB  |Cj   | 3   | 1   | 0   | FS4-999 |
| 2ABK  |Cl   | 2   | 1   | 0   | FS4-999 |
| 2ACK  |Cl   | 1   | 1   | 0   | EDR-999 |
| 2CND  |Cm   | 2   | 1   | 0   | FAD-271 |
| 2PGD  |Cn   | 2   | 2   | 1   | BTN-500 |
| 4BCL  |Co   | 1   | 1   | 0   | BTN-500 |


### Table 3A: Deep LBSs

| Tangent Sphere Method | Cutting Plane Method | CM-CM Method |
|-----------------------|----------------------|--------------|
| shallow               | medium   | deep     | shallow | medium | deep     | shallow | medium | deep     |
| Ac.ibs1               | Ab.ibs0          | Aa.ibs0   | Bt.ibs1 | Ab.ibs0 | Bm.ibs2  | An.ibs2 | Ad.ibs0 | Br.ibs2  |
| Ac.ibs2               | Ab.ibs0          | Aa.ibs0   | Bt.ibs1 | Ab.ibs0 | Bm.ibs2  | An.ibs2 | Ad.ibs0 | Br.ibs3  |
| Al.ibs2               | Ae.ibs0          | Aa.ibs0   | Bt.ibs1 | Ab.ibs0 | Bm.ibs2  | An.ibs2 | Ad.ibs0 | Br.ibs7  |
| Al.ibs3               | Ae.ibs0          | Aa.ibs0   | Bt.ibs1 | Ab.ibs0 | Bm.ibs2  | An.ibs2 | Ad.ibs0 | Br.ibs7  |
| Al.ibs2               | Ae.ibs0          | Aa.ibs0   | Bt.ibs1 | Ab.ibs0 | Bm.ibs2  | An.ibs2 | Ad.ibs0 | Br.ibs3  |
| Al.ibs1               | Ae.ibs0          | Aa.ibs0   | Bt.ibs1 | Ab.ibs0 | Bm.ibs2  | An.ibs2 | Ad.ibs0 | Br.ibs7  |
| Ae.ibs0               | Ae.ibs0          | Aa.ibs0   | Bt.ibs1 | Ab.ibs0 | Bm.ibs2  | An.ibs2 | Ad.ibs0 | Br.ibs7  |
| Al.ibs1               | Ae.ibs0          | Aa.ibs0   | Bt.ibs1 | Ab.ibs0 | Bm.ibs2  | An.ibs2 | Ad.ibs0 | Br.ibs7  |
| Ae.ibs1               | Ae.ibs0          | Aa.ibs0   | Bt.ibs1 | Ab.ibs0 | Bm.ibs2  | An.ibs2 | Ad.ibs0 | Br.ibs7  |
| Ae.ibs1               | Ae.ibs0          | Aa.ibs0   | Bt.ibs1 | Ab.ibs0 | Bm.ibs2  | An.ibs2 | Ad.ibs0 | Br.ibs7  |
| Ae.ibs1               | Ae.ibs0          | Aa.ibs0   | Bt.ibs1 | Ab.ibs0 | Bm.ibs2  | An.ibs2 | Ad.ibs0 | Br.ibs7  |
| Ae.ibs1               | Ae.ibs0          | Aa.ibs0   | Bt.ibs1 | Ab.ibs0 | Bm.ibs2  | An.ibs2 | Ad.ibs0 | Br.ibs7  |
| Ae.ibs1               | Ae.ibs0          | Aa.ibs0   | Bt.ibs1 | Ab.ibs0 | Bm.ibs2  | An.ibs2 | Ad.ibs0 | Br.ibs7  |
| Ae.ibs1               | Ae.ibs0          | Aa.ibs0   | Bt.ibs1 | Ab.ibs0 | Bm.ibs2  | An.ibs2 | Ad.ibs0 | Br.ibs7  |
| Ae.ibs1               | Ae.ibs0          | Aa.ibs0   | Bt.ibs1 | Ab.ibs0 | Bm.ibs2  | An.ibs2 | Ad.ibs0 | Br.ibs7  |
| Ae.ibs1               | Ae.ibs0          | Aa.ibs0   | Bt.ibs1 | Ab.ibs0 | Bm.ibs2  | An.ibs2 | Ad.ibs0 | Br.ibs7  |
| Ae.ibs1               | Ae.ibs0          | Aa.ibs0   | Bt.ibs1 | Ab.ibs0 | Bm.ibs2  | An.ibs2 | Ad.ibs0 | Br.ibs7  |
| Ae.ibs1               | Ae.ibs0          | Aa.ibs0   | Bt.ibs1 | Ab.ibs0 | Bm.ibs2  | An.ibs2 | Ad.ibs0 | Br.ibs7  |
| Ae.ibs1               | Ae.ibs0          | Aa.ibs0   | Bt.ibs1 | Ab.ibs0 | Bm.ibs2  | An.ibs2 | Ad.ibs0 | Br.ibs7  |
| Ae.ibs1               | Ae.ibs0          | Aa.ibs0   | Bt.ibs1 | Ab.ibs0 | Bm.ibs2  | An.ibs2 | Ad.ibs0 | Br.ibs7  |
| Ae.ibs1               | Ae.ibs0          | Aa.ibs0   | Bt.ibs1 | Ab.ibs0 | Bm.ibs2  | An.ibs2 | Ad.ibs0 | Br.ibs7  |
| Ae.ibs1               | Ae.ibs0          | Aa.ibs0   | Bt.ibs1 | Ab.ibs0 | Bm.ibs2  | An.ibs2 | Ad.ibs0 | Br.ibs7  |
| Ae.ibs1               | Ae.ibs0          | Aa.ibs0   | Bt.ibs1 | Ab.ibs0 | Bm.ibs2  | An.ibs2 | Ad.ibs0 | Br.ibs7  |
| Ae.ibs1               | Ae.ibs0          | Aa.ibs0   | Bt.ibs1 | Ab.ibs0 | Bm.ibs2  | An.ibs2 | Ad.ibs0 | Br.ibs7  |
| Ae.ibs1               | Ae.ibs0          | Aa.ibs0   | Bt.ibs1 | Ab.ibs0 | Bm.ibs2  | An.ibs2 | Ad.ibs0 | Br.ibs7  |
| Ae.ibs1               | Ae.ibs0          | Aa.ibs0   | Bt.ibs1 | Ab.ibs0 | Bm.ibs2  | An.ibs2 | Ad.ibs0 | Br.ibs7  |
| Ae.ibs1               | Ae.ibs0          | Aa.ibs0   | Bt.ibs1 | Ab.ibs0 | Bm.ibs2  | An.ibs2 | Ad.ibs0 | Br.ibs7  |
| Ae.ibs1               | Ae.ibs0          | Aa.ibs0   | Bt.ibs1 | Ab.ibs0 | Bm.ibs2  | An.ibs2 | Ad.ibs0 | Br.ibs7  |
| Ae.ibs1               | Ae.ibs0          | Aa.ibs0   | Bt.ibs1 | Ab.ibs0 | Bm.ibs2  | An.ibs2 | Ad.ibs0 | Br.ibs7  |
| Ae.ibs1               | Ae.ibs0          | Aa.ibs0   | Bt.ibs1 | Ab.ibs0 | Bm.ibs2  | An.ibs2 | Ad.ibs0 | Br.ibs7  |
| Ae.ibs1               | Ae.ibs0          | Aa.ibs0   | Bt.ibs1 | Ab.ibs0 | Bm.ibs2  | An.ibs2 | Ad.ibs0 | Br.ibs7  |
| Ae.ibs1               | Ae.ibs0          | Aa.ibs0   | Bt.ibs1 | Ab.ibs0 | Bm.ibs2  | An.ibs2 | Ad.ibs0 | Br.ibs7  |
| Ae.ibs1               | Ae.ibs0          | Aa.ibs0   | Bt.ibs1 | Ab.ibs0 | Bm.ibs2  | An.ibs2 | Ad.ibs0 | Br.ibs7  |
| Ae.ibs1               | Ae.ibs0          | Aa.ibs0   | Bt.ibs1 | Ab.ibs0 | Bm.ibs2  | An.ibs2 | Ad.ibs0 | Br.ibs7  |
| Ae.ibs1               | Ae.ibs0          | Aa.ibs0   | Bt.ibs1 | Ab.ibs0 | Bm.ibs2  | An.ibs2 | Ad.ibs0 | Br.ibs7  |
| Ae.ibs1               | Ae.ibs0          | Aa.ibs0   | Bt.ibs1 | Ab.ibs0 | Bm.ibs2  | An.ibs2 | Ad.ibs0 | Br.ibs7  |
| Ae.ibs1               | Ae.ibs0          | Aa.ibs0   | Bt.ibs1 | Ab.ibs0 | Bm.ibs2  | An.ibs2 | Ad.ibs0 | Br
| STRUCTURE | T.S. Method | C.F. Method | C.h-to-C.m Method |
|-----------|-------------|-------------|-------------------|
| Ms | lig | res | sdc | ave | lig | res | sdc | ave | lig | res | sdc | ave | lig | res | sdc | ave |
| As. hm 50 | 0.5075 | 2.6671 | 0.6026 | 0.6585 | 0.2017 | 1.0205 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 |
| As. Lhm 47.5 | 0.5075 | 2.6671 | 0.6026 | 0.6585 | 0.2017 | 1.0205 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 |
| As. Lhm 55.0 | 0.5075 | 2.6671 | 0.6026 | 0.6585 | 0.2017 | 1.0205 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 |
| As. Lhm 62.5 | 0.5075 | 2.6671 | 0.6026 | 0.6585 | 0.2017 | 1.0205 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 |
| As. Lhm 70.0 | 0.5075 | 2.6671 | 0.6026 | 0.6585 | 0.2017 | 1.0205 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 |
| As. Lhm 77.5 | 0.5075 | 2.6671 | 0.6026 | 0.6585 | 0.2017 | 1.0205 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 |
| As. Lhm 85.0 | 0.5075 | 2.6671 | 0.6026 | 0.6585 | 0.2017 | 1.0205 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 |
| As. Lhm 92.5 | 0.5075 | 2.6671 | 0.6026 | 0.6585 | 0.2017 | 1.0205 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 |
| As. Lhm 100.0 | 0.5075 | 2.6671 | 0.6026 | 0.6585 | 0.2017 | 1.0205 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 | 0.2017 |
| Table 3B: Medium LBSs

**Table 3B:** Medium LBSs
| STRUCTURE | T.S. Method | C.I. Method | CN-to-CN Method |
|-----------|-------------|-------------|-----------------|
|           | lig res sdc ave | lig res sdc ave | lig res sdc ave |
| Ac.LSB1  | 0.6554 8.9412 8.0295 3.1997 | 0.6554 8.9412 8.0295 3.1997 | 0.6554 8.9412 8.0295 3.1997 |
| Ac.LSB2  | 0.5930 9.3951 8.0231 3.1986 | 0.5930 9.3951 8.0231 3.1986 | 0.5930 9.3951 8.0231 3.1986 |
| Ac.LSB3  | 0.5930 9.3951 8.0231 3.1986 | 0.5930 9.3951 8.0231 3.1986 | 0.5930 9.3951 8.0231 3.1986 |
| Ac.LSB4  | 0.5930 9.3951 8.0231 3.1986 | 0.5930 9.3951 8.0231 3.1986 | 0.5930 9.3951 8.0231 3.1986 |
| Ac.LSB5  | 0.5930 9.3951 8.0231 3.1986 | 0.5930 9.3951 8.0231 3.1986 | 0.5930 9.3951 8.0231 3.1986 |
| Ac.LSB6  | 0.5930 9.3951 8.0231 3.1986 | 0.5930 9.3951 8.0231 3.1986 | 0.5930 9.3951 8.0231 3.1986 |
| Ac.LSB7  | 0.5930 9.3951 8.0231 3.1986 | 0.5930 9.3951 8.0231 3.1986 | 0.5930 9.3951 8.0231 3.1986 |
| Ac.LSB8  | 0.5930 9.3951 8.0231 3.1986 | 0.5930 9.3951 8.0231 3.1986 | 0.5930 9.3951 8.0231 3.1986 |
| Ac.LSB9  | 0.5930 9.3951 8.0231 3.1986 | 0.5930 9.3951 8.0231 3.1986 | 0.5930 9.3951 8.0231 3.1986 |
| Ac.LSB10 | 0.5930 9.3951 8.0231 3.1986 | 0.5930 9.3951 8.0231 3.1986 | 0.5930 9.3951 8.0231 3.1986 |
| Ac.LSB11 | 0.5930 9.3951 8.0231 3.1986 | 0.5930 9.3951 8.0231 3.1986 | 0.5930 9.3951 8.0231 3.1986 |
| Ac.LSB12 | 0.5930 9.3951 8.0231 3.1986 | 0.5930 9.3951 8.0231 3.1986 | 0.5930 9.3951 8.0231 3.1986 |
| Ac.LSB13 | 0.5930 9.3951 8.0231 3.1986 | 0.5930 9.3951 8.0231 3.1986 | 0.5930 9.3951 8.0231 3.1986 |
| Ac.LSB14 | 0.5930 9.3951 8.0231 3.1986 | 0.5930 9.3951 8.0231 3.1986 | 0.5930 9.3951 8.0231 3.1986 |
| Ac.LSB15 | 0.5930 9.3951 8.0231 3.1986 | 0.5930 9.3951 8.0231 3.1986 | 0.5930 9.3951 8.0231 3.1986 |
| Ac.LSB16 | 0.5930 9.3951 8.0231 3.1986 | 0.5930 9.3951 8.0231 3.1986 | 0.5930 9.3951 8.0231 3.1986 |
| Ac.LSB17 | 0.5930 9.3951 8.0231 3.1986 | 0.5930 9.3951 8.0231 3.1986 | 0.5930 9.3951 8.0231 3.1986 |
| Ac.LSB18 | 0.5930 9.3951 8.0231 3.1986 | 0.5930 9.3951 8.0231 3.1986 | 0.5930 9.3951 8.0231 3.1986 |
| Ac.LSB19 | 0.5930 9.3951 8.0231 3.1986 | 0.5930 9.3951 8.0231 3.1986 | 0.5930 9.3951 8.0231 3.1986 |
| Ac.LSB20 | 0.5930 9.3951 8.0231 3.1986 | 0.5930 9.3951 8.0231 3.1986 | 0.5930 9.3951 8.0231 3.1986 |

Table 3C: Shallow LBSs