CHAPTER 10

Molecular Docking with *Candida antarctica* lipase B[1]

10.1 Introduction

In the previous chapters we have shown that enantiomerically pure building blocks like pyrrolinones and pyranones can be obtained by means of lipase catalyzed reactions or palladium catalyzed allylic substitutions. In addition of finding ways to obtain a chosen class of compounds enantiomerically pure, another challenge is to search for new compounds that can be obtained enantiomerically pure using a similar procedure.

In order to find new substrates for a particular enzyme one might perform a large number of enzymatic reactions screening different substrates. Such a process can be rather time consuming but nowadays this drawback might be partially overcome by using new techniques like combinatorial chemistry.[2] Another possibility for finding new substrates is to use computer calculations to screen large numbers of substrates.

The computer program DOCK[3] can be used for random screening of very large numbers of possible substrates (or inhibitors)[4] for a certain enzyme. This program is most often used in order to propose novel enzyme inhibitors and other therapeutic agents. With this program substrates are docked in a simplified version of the active site of the enzyme.[5] Compared to molecular modeling (see Chapter 9) this is a much faster but also a far less accurate technique to find out which substrates fit in the active site of the enzyme.

With the program DOCK points within the active site are identified where ligand[4] atoms may be located. From these points, called sphere centers, a set of overlapping spheres is generated which fill the active site. This is schematically represented in Figure 10.1.[5]

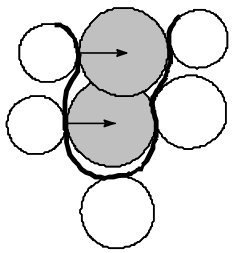

Figure 10.1 *Schematic representation of a small binding site of two spheres formed from five active site atoms. The overlapping spheres are constructed with their centers along the surface normals (the molecular surface is shown as a thick line).*

To orient a substrate within the active site, some of the sphere centers are matched (or paired) with substrate atoms. The orientation of the substrate can be evaluated with a function approximating the substrate-enzyme binding energy. The binding energy is derived from the sum of the van der Waals attractive, van der Waals dispersive, and Coulomb electrostatic energies. The energy score is the sum over all substrate atoms for these combined terms. Chemical scoring can also be applied. Therefore sphere chemical matching or coloring associates a chemical property to a sphere and a sphere of one color can only be matched with a substrate atom of a complementary color. The chemical properties include e.g. hydrogen bond donor, hydrogen bond acceptor, hydrophobic, electropositive, electronegative, and neutral.

Molecular docking has proven to be a useful computational technique for identifying small molecules from large databases, which are complementary to the active sites of enzymes[6]. Despite current limitations, such as the use of a single conformation for each molecule and the complexity of accurately predicting protein-ligand affinities, several successful applications have been reported over the last decade (see for instance Somoza et al.[7] for a recent example).

Because the crystal structure of lipase B from the yeast *Candida antarctica* is available,[8] we chose this lipase for our molecular docking experiment. *Candida antartica* lipase B, however, presents a difficult problem for molecular docking as its active site is large and mainly hydrophobic. Large active sites require extensive spatial sampling in order to identify the most favorable orientation of the ligand with respect to the protein. The hydrophobic nature of the active site yields few markers for specific interactions (hydrogen bonding) between ligands and protein. Although these limitations are likely to cause an unknown number of false negatives, molecular docking still offers the possibility to computationally screen large numbers of commercially available chemicals at limited costs. Therefore, the database docking program DOCK version 4.0[9] was used to propose novel substrates for *Candida antartica* lipase B from a collection of commercially available esters and anhydrides. The preliminary results towards the search for new CALB substrates are presented here.

10.2 Molecular Docking

All compounds containing an ester fragment R-CO_2-C-R' were taken from the Available Chemicals Directory (ACD), yielding a total of 40304 compounds (esters and anhydrides), which were transformed to mol2 format with Gasteiger charges. An X-ray structure of *Candida antarctica* B, cocrystallized with the detergent Tween 80, obtained from the Brookhaven Protein Data Bank, was used.[10,11] A sphere file containing 14 spheres was generated from 7 SPHGEN spheres and 7 atom positions of the DOCK 4 rigid-body minimized Tween 80 substrate located close to the catalytic triad. These were docked against the lipase crystal structure as rigid bodies using a united atom model and uniform sampling with maximally 200 conformations was scored. This is roughly double the amount of

sampling normally used in database searches since there is a fairly large active site to be searched. Both chemical and energy scoring were applied. All molecules with an sp^2 oxygen within 4 Å of the lipase Ser105 hydroxyl oxygen were selected from the highest ranking 2000 molecules obtained with both scoring functions. These lists were further reduced by visual inspection which removed compounds with sp^2 oxygens from non-ester groups close to the catalytic triad. Finally, 260 compounds were left from the chemical scoring list and 212 from the energy scoring list. These lists still contain some double entries and anhydrides. Note that since the active site is large and some substrates small, possible substrates may been missed because DOCK did not place the ester bond within 4 Å of the catalytic serine.

When the set of 260 possible substrates, which where selected on chemical basis, and the 212 compounds from the energy scoring list are combined and the double entries are removed, we obtain a list of 458 different compounds that can in principle be tested as a substrate for *Candida antartica* lipase B.

10.3 Possible New Substrates for CALB

Because an amount of 458 different substrates is still rather large to test and it was observed that there were many similarities between compounds, we have first ordered the substrates on basis of the number of stereogenic centers being present in each molecule. Secondly we have combined the compounds that differ only in one or more side chains (R) in one figure. From the 458 different compounds, the most prominent structures are summarized in Figures 10.2, 10.3, and 10.4.

In Figure 10.2, a number of achiral potential substrates are given. The diversity of structures on the basis of this rough screening and size of some of these substrates is remarkable. It is, however, known that *Candida antartica* lipase B is able to convert substrates as large as steroids.[12]

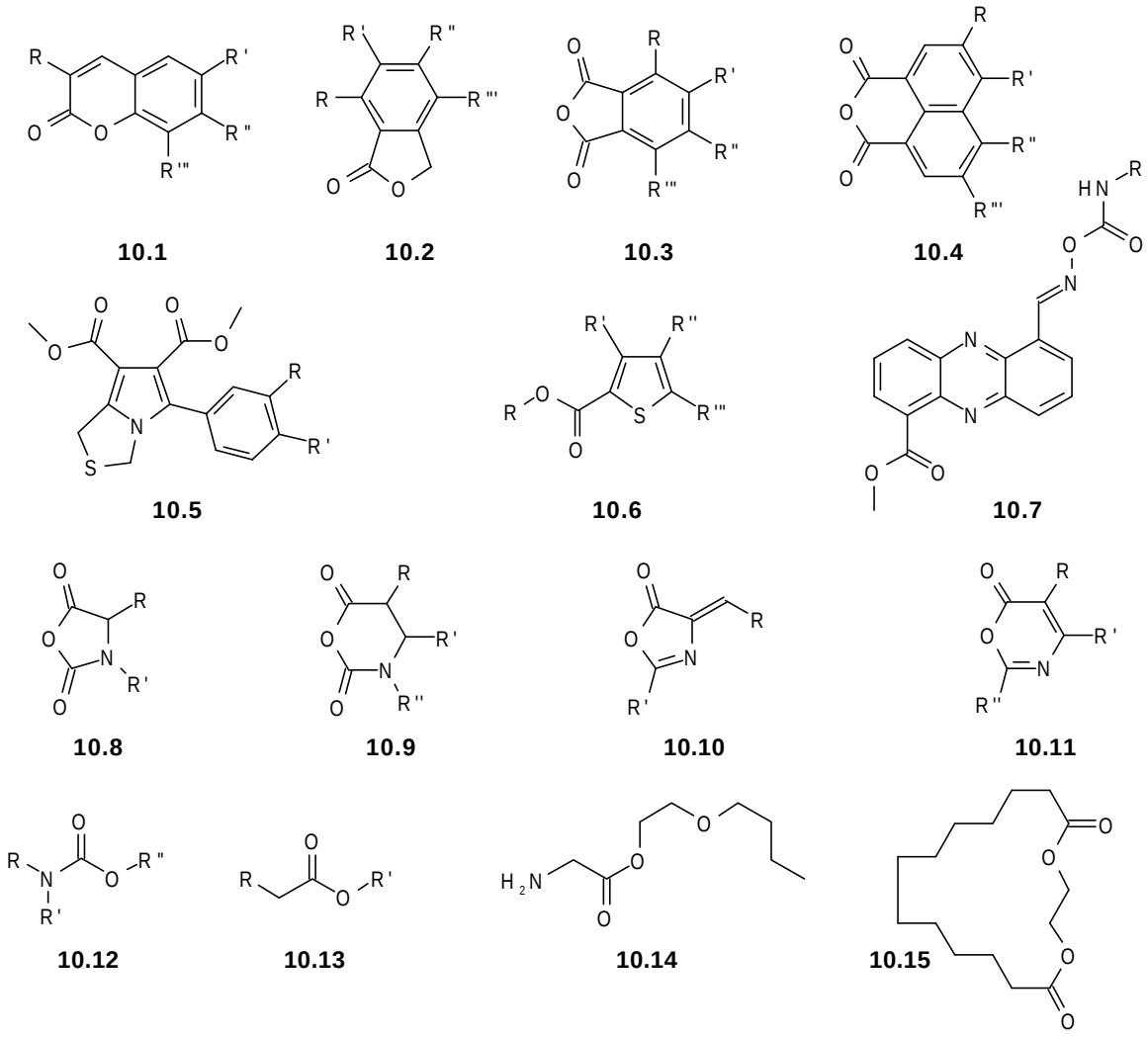

Figure 10.2 *Possible substrates for Candida antartica lipase B that contain no stereogenic center*

In Figure 10.3 substrates with one chiral center are presented. This group of compounds is very interesting for us because it contains potential simple chiral building blocks, that might be obtained enantiomerically pure with the use of CALB. It can also be observed from Figure 10.3 that among these compounds there are some very standard chiral synthons like several amino acid esters **10.16** (R = alkyl, aryl, CH_2CONH_2, CH_2CO_2H, CH_2OH, CH_2SH, R'= alkyl, Bn) and **10.17** – **10.20** and several lactones **10.21** – **10.23**.

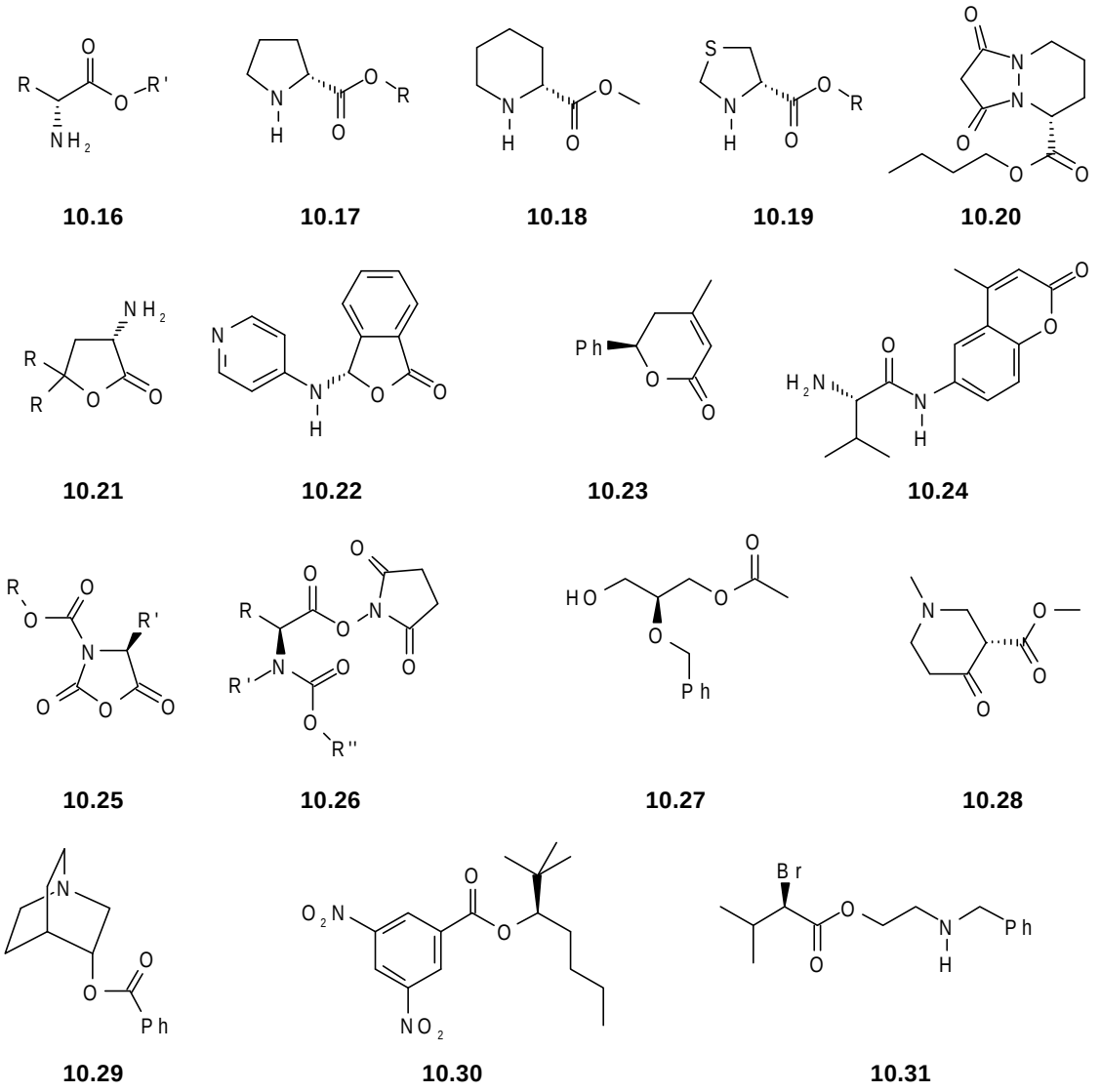

Figure 10.3 *Possible substrates for Candida antartica lipase B with one stereogenic center*

Most of the compounds with more than one chiral centers (Figure 10.4) are more difficult to use for testing. Among these possible substrates are, however, interesting compounds like an aziridine **10.33**, a lactone **10.41**, a sugar **10.44**, and several steroids **10.48**.

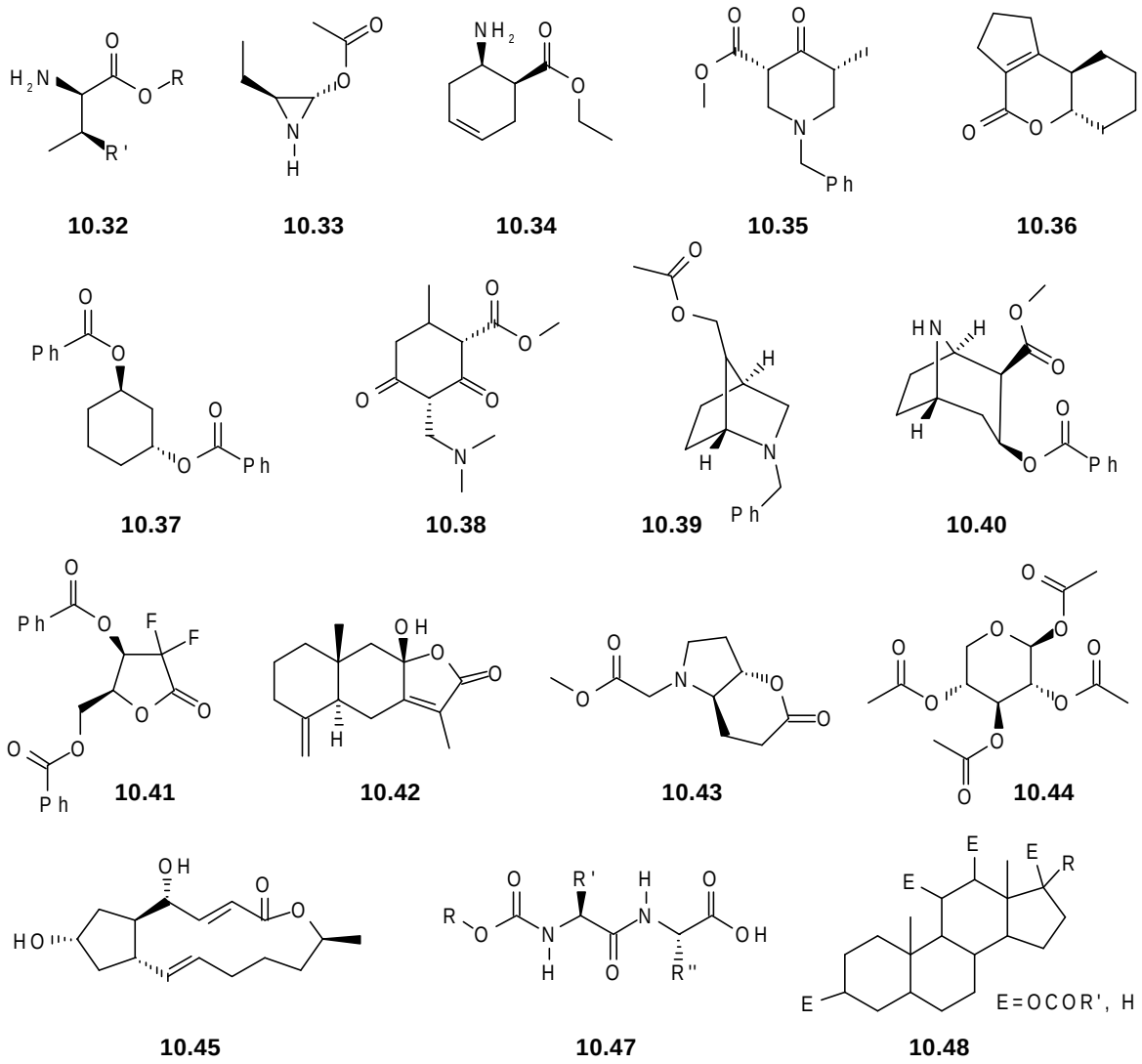

Figure 10.4 *Possible substrates for Candida antartica lipase B with two or more stereogenic centers.*

10.4 Enzymatic Reactions

From the list of compounds described above we made a selection of substrates that we planned to buy. However, we faced two major problems. First many of the compounds from the Available Chemicals Database were no longer available. Second many compounds were only available in enantiopure form.

From the set of compounds we were able to order a limited number were briefly tested in an enzymatic transesterification with CALB in *n*-hexane/ ethanol according to the procedure described for pyrrolinones in Chapter 3 (ethanol was used because it is easier to remove than *n*-butanol).

The first compound that we used was (+/-)-3-quinuclidinol benzyl ester **10.29** (Figure 10.3). After 72 h shaking at room temperature or 20 h refluxing in the presence of CALB there was still only starting material present according to NMR.

We also tested some amino acid esters. DL-alanine methylester **10.16** (R = R' = Me), DL-proline methyl ester **10.17** (R = Me), and DL-methylpipecolinate **10.17** were all converted to the ethyl esters according to NMR after 42 h shaking at room temperature. Although in these reactions both enantiomers were converted, this is a probably a much too long reaction time and in order to determine whether there is any kinetic resolution the progress of the reaction must be followed in time. So far a way to determine the e.e. has not been found. The reactions could not be followed with GC and these compounds are difficult to detect with UV making normal HPLC not a suitable method. There are, however, other methods to determine the e.e. of amino acid esters such as HPLC with fluorescence detection or NMR with chiral shift reagents.

10.5 Conclusions

From a total of 40304 molecules that have been used in a molecular docking experiment, 452 possible substrates for Candida *antartica* lipase B have been suggested. Because we only have performed a few initial experiments with a limited number of substrates it is not possible to conclude anything about the correctness of our predictions. However we recommend to screen several of the chiral heterocyclic compounds shown in Figures 10.3 and 10.4 for kinetic resolution with CALB.

10.6 Experimental Section

A total of 40304 molecules containing ester bond fragments were selected from the Available Chemicals Directory (ACD, MDL Inc., San Leandro, USA) on the basis of 2D similarity using Chem-X (Molecular Design Ltd., Oxon, UK). These were subsequently transformed to Sybyl mol2 format using the DOCK database processing utilities and Sybyl 6.3 (Tripos Inc., St. Louis, MO, USA). For docking the crystal structure of lipase from Candida *antartica* B cocrystallized with the detergent Tween 80 (obtained from the Brookhaven Protein Data Bank,[10] entry 1LBT[11]) determined at 2.5 Å resolution was used. A sphere set describing the active site for use with DOCK 4 was created by merging 7 spheres generated by SPHGEN[5] with 7 atom positions along the Tween-80 ester bond, obtained after rigid body minimization of the substrate with DOCK 4 in the lipase active site. These spheres were chosen to limit spatial sampling to the vicinity of the catalytic triad and to find favorable positions for the placement of ester bonds. Scoring grids were generated with CHEMGRID[13] for Amber[14] force field and chemical scoring, using a 1/4r distance dependent dielectric constant, a 10 Å distance cut-off for non-bonded interactions, a unified atom model and a grid spacing of 0.3 Å. Chemical scoring, as implemented in DOCK 4, is similar to force field

scoring but uses favorable weighting for interactions between chemically related atom types (e.g. hydrogen bond donors and acceptors, hydrophobic groups) and penalizes interactions between unrelated functionalities (such as hydrophobic atoms with polar groups)[15].
All molecules were docked as rigid bodies against the active site of lipase using uniform sampling with 200 orientations to be scored and minimized for 100 steps per molecule. Molecules with a carbonyl functionality within 4 Å of the Ser105 hydroxyl in the lipase active site were selected from the 2000 highest ranking docking solutions obtained with force field and chemical scoring. The resulting lists were further reduced by visual inspection with Sybyl 6.3 and the most promising compounds were saved to file. All calculations were performed on an eight R10000 processor SGI Power Challenge operating at 200 MHz.

10.7 References

1 This study was performed in cooperation with dr. R. Knegtel (Organon).

2 Combinatorial chemistry in combination with in-vitro evolution: (a) Reetz, M.T.; Zonta, A.; Schimossek, K.; Liebeton, K.; Jaeger, K-E *Angew. Chem. Int. Ed.* **1997**, *36*, 2830. (b) Reetz, M.T.; Becker, M.H.; Kühling, K.M.; Holzwarth, A. *Angew. Chem. Int. Ed.* **1998**, *37*, 2647.

3 (a) Shoichet, B.K.; Bodian, D.L.; Kuntz, I.D. *J. Comp. Chem.* **1992**, *13*, 380. (b) Kuntz, I.D.; Meng, E.C.; Shoichet, B.K. *Acc. Chem. Res.* **1994**, *27*, 117.

4 In the literature substrates or inhibitors of enzymes are often called ligands.

5 Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R. and Ferrin, T.E. *J. Mol. Biol.* **1982**, *161*, 269.

6 Kuntz, I.D., *Science* **1992**, *257*, 1078.

7 Somoza, J.R., Skillman, A.G., Munagala, N.R., Oshiro, C.M., Knegtel, R.M.A, Mpoke, S., Katakura, S., Fletterick, R.J., Kuntz, I.D.and Wang, C.C. *Biochemistry* **1998**, *37*, 5344.

8 (a) Uppenberg, J.; Hansen, M.T.; Patkar, S.; Jones, T.A. *Structure* **1994**, *2*, 293. (b) Uppenberg, J.; Patkar, S.; Bergfors, T.; Jones, T.A. *J. Mol. Biol.* **1994**, *235*, 790. (c) Uppenberg, J.; Öhrner, N.; Norin, M.; Hult, K.; Kleywegt, G.J.; Patkar, S.; Waagen, V.; Anthonsen, T.; Jones, T.A. *Biochemistry* **1995**, *33*, 16838.

9 Ewing, T.J.A. and Kuntz, I.D. *J. Comp. Chem.* **1997**, *18*, 1175.

10 (a)Abola, E.E.; Bernstein, F.C.; Bryant, S.H.; Koezle, T.F.; Weng, J. Protein Data Bank. In Crystallographic Databases Content, Software Systems, Scientific Applications. Allen, F.H.; Bergerhoff, G.; Sievers, R. (eds.). Data Commission of the International Union of Crystallography, Bonn/Cambridge/Chester. 1987: 107-132. (b) Bernstein, F.C.; Koezle, T.F.; Williams, G.J.B.; Meyer, Jr., E.F.; Brice, M.D.; Rodgers, J.R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The Protein Data Bank: a Computer-based Archival file for Macromolecular Structures. *J. Mol. Biol.* **1977**, *112*, 535. Entry 1lbt.

11 Uppenberg, J.; Öhrner, N.; Norin, M.; Hult, K.; Kleywegt, G.J.; Patkar, S.; Waagen, V.; Anthonsen, T.; Jones, T.A. *Biochemistry* **1995**, *33*, 16838

12 Bertinotti, A.; Carrea, G.; Ottolina, G.; Riva, S. *Tetrahedron* **1994**, *50*, 13165.

13 Meng, E.C., Shoichet, B.K. and Kuntz, I.D. *J. Comp. Chem.* **1992**, *13*,505.

14 Weiner, S.J., Kollman, P.A., Hguyen, D.T. and Case, D.A. *J. Comp. Chem.* **1986**, *7*, 230.

15 Ewing, T. (Ed.) DOCK Version 4.0 Manual, Regents of the University of California, San Francisco, **1997**.