ELSEVIER

Journal of Molecular Graphics and Modelling 20 (2002) 399–409

Journal of Molecular Graphics and Modelling

www.elsevier.com/locate/JMGM

Computer-aided design of chiral ligands Part I. Database search methods to identify chiral ligand types for asymmetric reactions

Marisa C. Kozlowski*, Manoranjan Panda

Department of Chemistry, Roy and Diana Vagelos Laboratories, University of Pennsylvania, Philadelphia, PA 19104, USA

Received 14 September 2001; accepted 19 November 2001

Abstract

The utility of database searching to identify chiral ligand motifs is outlined. The key elements of three known chiral ligands have been described as bond vectors. The CAVEAT program was then used to screen the Cambridge Structural Database (CSD), portions of the Chemical Abstracts Services three-dimensional database (CAS-3D), and the TRIAD tricyclic structure database for scaffolds containing these elements. Scaffolds corresponding to the known starting points were identified indicating that this method can be used to identify chiral ligand structural motifs. In addition, alternate structural motifs were found that suggested alternative possible ligands.

Keywords: Computer-aided design; Chiral ligands; Database screening; Transition state; Molecular mechanics; Asymmetric reactions

1. Introduction

Diastereo- and enantioselective chemical reactions have become essential for the efficient synthesis of complex chiral targets, including natural products, novel materials, biological probes, and pharmaceuticals. While much progress has been made in developing nonenzymatic asymmetric methodology, good stereoselective versions of many reactions remain unavailable. The development of effective chiral auxiliaries (stoichiometric) and catalysts (substoichiometric) for asymmetric reactions has largely been driven by the identification of the appropriate chiral ligands. The conventional approach towards ligand identification relies upon intuitive or random screening approaches. These empirical approaches bias which ligands are considered. Usually researchers consider motifs with which they are familiar or which can be made from commercially available chiral materials. When an entirely new ligand type is identified, the impact can be enormous. An example is the axially chiral ligand class typified by BINOL and BINAP [1]. An accelerated approach toward identification of such ligands for selective reactions would be extremely valuable. To this end, we are developing a computational method to generate new chiral ligand classes for asymmetric reactions by regularly identifying motifs that conventional ligand designs usually do not consider.

Computational methods can aid in the generation of chiral auxiliaries and catalysts by providing increased efficiency, an improved rational basis, and leads usually available only by intuitive leaps. With this motivation, we have developed a systematic methodology that could lead to novel ligands for a number of asymmetric reactions. A schematic diagram of the steps is presented in Fig. 1. In the multiple step procedure (1) the transition state for the particular reaction is modeled, (2) stereodiscriminating groups (SGs) are selected from models of previously successful chiral ligands or are generated around the desired transition state by functionality mapping, (3) vectors to define relative placement and orientation are chosen to the metal and stereodiscriminating groups, (4) the CAVEAT program uses the vectors to screen large databases for molecules that contain bonds with matching vector relationships, and (5) the hit compounds are classified into structural subgroups and the important motifs are identified. The actual ligand is constructed by incorporating the identified motif into a synthetically reasonable structure and then appending the appropriate substituents to produce the chiral ligand. In the last part (Step 6), the ligands are evaluated theoretically and/or experimentally for potential selectivity and utility.

In this paper, the concept in Steps 3 and 4 (Fig. 1) has been benchmarked using a series of known chiral ligands. In particular, the question of whether it is possible to identify lead structures containing motifs corresponding to known chiral ligands is addressed. In addition, a reasonable method for

* Corresponding author. Tel.: +1-215-898-3048; fax: +1-215-573-7165.
E-mail address: marisa@sas.upenn.edu (M.C. Kozlowski).

1093-3263/02/$ – see front matter
PII: S1093-3263(01)00138-3

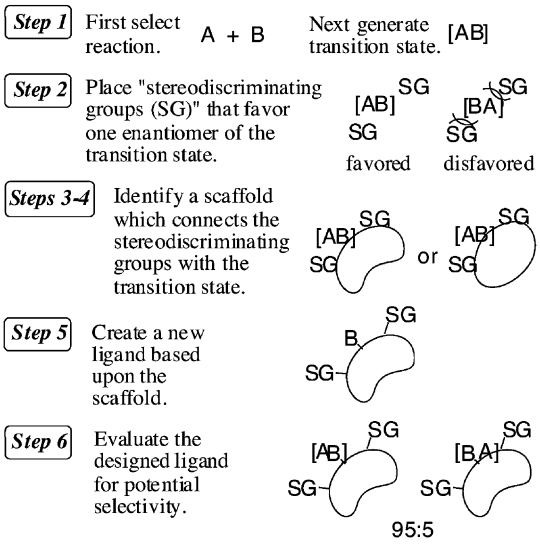
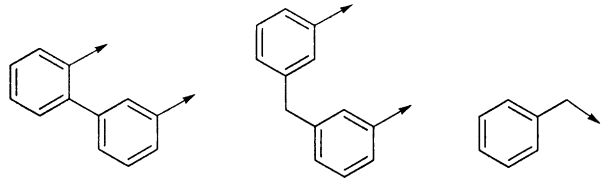


Fig. 1. Generalized schematic of the ligand design protocol.

classifying the identified scaffolds is presented. The number and types of compounds in a given class provide useful insight into whether a scaffold is general enough to allow modification and still retain the desired geometric parameters.

2. Computational methods

All calculations were performed using a SGI OCTANE workstation with an R10,000 processor. Molecular mechanics calculations of the transition states were performed using MacroModel [2] with a modified MM2* force field [3,4]. The CAVEAT [5] program was used, which includes TRIAD (a database of tricyclic ring structures) and ILIAD (a database of acyclic structures). The Cambridge Structural Database (CSD) [6] and the Chemical Abstracts Services three-dimensional database (CAS-3D) [1] were obtained separately and were also used as source databases for CAVEAT. CAVEAT vector databases were constructed with the cavinitvector script supplied with CAVEAT using metals and lone pairs as searchable elements in addition to nonmetallic atoms as previously described [5]. The *linked* database (vprep option: -e) was prepared from nonmetallic cyclic structures in the CSD using "one-bond linked-ring" vector pairs; all the bases of the vectors were part of a ring system. The *met* database (vprep options:-z, -o, -h) was constructed from all structures in the CSD containing metals. All bonds were chosen as vectors provided that the base is not in the same cyclic system as the tip (with the exception of metallocycles). The net result was that vector pairs which possessed tips that shared a metal atom were allowed, if there was a molecular fragment connecting their bases by a path that did not involve the tip atoms. The *cast* database (vprep options: -g, -f) was generated from a sample set of the CAS-3D database processed for "two-bond linked-ring" vector pairs and "one-out" vectors. The *triad* database (vprep option: -f) was made from the TRIAD database processed for "one-out" vectors. Leads from the database searches are reported with the CSD code for structures from *met*, with the CAS registry number for structures from *cast*, and with a unique identifying code denoting the ring type for structures from *triad*.



The CLASS subroutine of CAVEAT implements two kinds of clustering: partial comparison in which not all pairs of structures are compared and the full comparison in which every structure is compared to every other structure. While slightly slower, the latter approach was chosen since it is more accurate and since the classification protocols were typically rapid overall (<10 min). Of the three available algorithms to differentiate the structures into groups (graph, threshold and Jarvis–Patric), graph clustering was employed [5]. In this method, each group contains all the structures sharing one core geometry. A new structure is added to the group if its core matches the core of any of the existing structures in the group within a user specified RMS deviation. For our systems, we found use of full comparison and graph clustering gave a good overview of the structural diversity of the hits.

The features described are available in the CAVEAT 2.2b release. Visualization of CAVEAT input and output was accomplished using a MacroModel interface.

3. Results and discussion

3.1. Boron aldol and allylation reactions

The boron aldol and allylation reactions were chosen as initial cases for study because the stereochemical factors important to these reactions are relatively well understood. These reactions, which find a wide variety of applications in synthetic organic chemistry, have been examined extensively experimentally and theoretically [7,8]. In addition, the absence of turnover and any elements past the second row of the periodic table greatly simplifies the computational analysis.

Stereochemical control of the aldol and allylation reactions with unsubstituted alkene derivatives involves the selective formation of one enantiomer, whereas substituted

[1] We thank J.W. Macko and Chemical Abstract Services for making CAS-3D available to us.

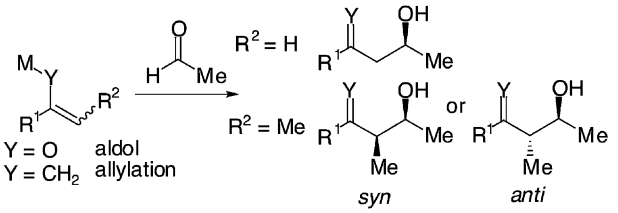


Fig. 2. Stereochemical outcomes of the aldol and allylation reactions.

alkene derivatives also require control of diastereoselectivity to achieve a selective reaction (Fig. 2). Numerous auxiliaries and catalysts have been developed to facilitate many aldol and allylation reactions with a remarkable degree of selectivity [1,7]; however, some variants, particularly *anti* additions, remain problematic.

A reasonable mechanistic understanding of the aldol and allylation reactions has been achieved [9]. Some of the most studied variants are the reactions involving enol- and allylboron compounds. Different kinds of transition states, both cyclic and acyclic have been reported to explain the observed selectivity. The cyclic transition states TS-A1 and TS-A2 lead to different enantiomers of the product since the allyl or enolate is delivered to a different faces of the prochiral aldehyde. When the ligands of the boron (BL_2) are chiral, TS-A1 and TS-A2 are diastereomeric with different energies. The differences in energy between these chair transition states, illustrated schematically in Fig. 3, generally correlate well with the experimental enantioselectivities, although in certain instances boat forms of these

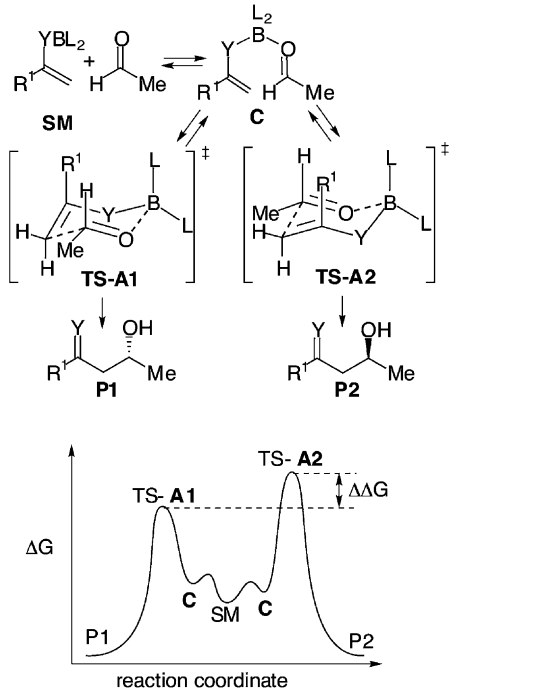


Fig. 3. Correlation between transition states for the boron aldol/allylation reaction and product ratios.

cyclic transition states may predominate [10]. Thus, boron ligands which can maximize the energy difference through selective stabilization or destabilization of one diastereomeric pathway will cause one enantiomeric product to predominate leading to an asymmetric transformation.

Progress has been made toward applying computational tools to the development of such asymmetric variants with the implementation of boron aldol [3] and allylation [4] force fields for molecular mechanics transition state calculations; however, prior computational efforts have been restricted to the *prediction* of stereoselectivity for given substrates and chiral ligands [11,12]. The use of computational methods for the *generation*, as well as for the evaluation, of chiral auxiliaries and catalysts has not been documented and could provide a useful tool for the design of new chiral ligands.

Analysis of the means by which a given ligand causes stereochemical discrimination in a boron aldol or allylation reaction indicates that there are two key features to any ligand. The first key feature is a set of "stereodiscriminating groups" which cause either stabilizing interactions in the favored enantiomeric transition state (TS-A1, Fig. 3) and/or destabilizing interactions in the disfavored enantiomeric transition state (TS-A2, Fig. 3). The second key feature is a ligand backbone or motif which confers enough structural rigidity (or fixes conformational mobility) such that the stereodiscriminating groups are presented uniformly towards the reactant partners. Examples of these two sets of structural features for three highly enantioselective boron chiral reagents (Fig. 4) are examined in the following paragraphs.

The dimethylborolane (**I**) developed by Masamune et al. has been used in the aldol addition of acetate and propionate enolates (E) [13a] as well as allyls [14]. Examination of the illustrated favorable and disfavorable transition structures (Fig. 4) proposed for this reaction [13] reveals that the stereodiscriminating groups are the two methyl substituents of the borolane while the five-membered borolane ring is the rigid scaffold element which presents these two stereodiscriminating groups.

In a second example, the Stein reagent **II** developed by Corey et al. for the aldol additions [15] of thioester enolates as well as for allylation reactions [16] contains a similar but slightly larger scaffold. Corey et al. have proposed that the *trans* phenyl substituents of the five-membered boron ring force the sulfonamide aryl groups to orient as shown [15]. The chair transition state where the SPh group on the enolate does not approach the sulfonamide aryl group then leads to the major product (Appendix A). In this case, the scaffolding element is the five-membered boranesulfonamide and the pendant sulfonyl groups. The stereodiscriminating groups are now the two aryl groups of the sulfonamide.

In a third example, the dimenthylborane reagent **III** which has been developed for use in aldol reactions [11], the scaffolding element is less obvious but arises from the positions of the two cyclohexane groups which are a result of the minimization of steric and torsional interactions between the two

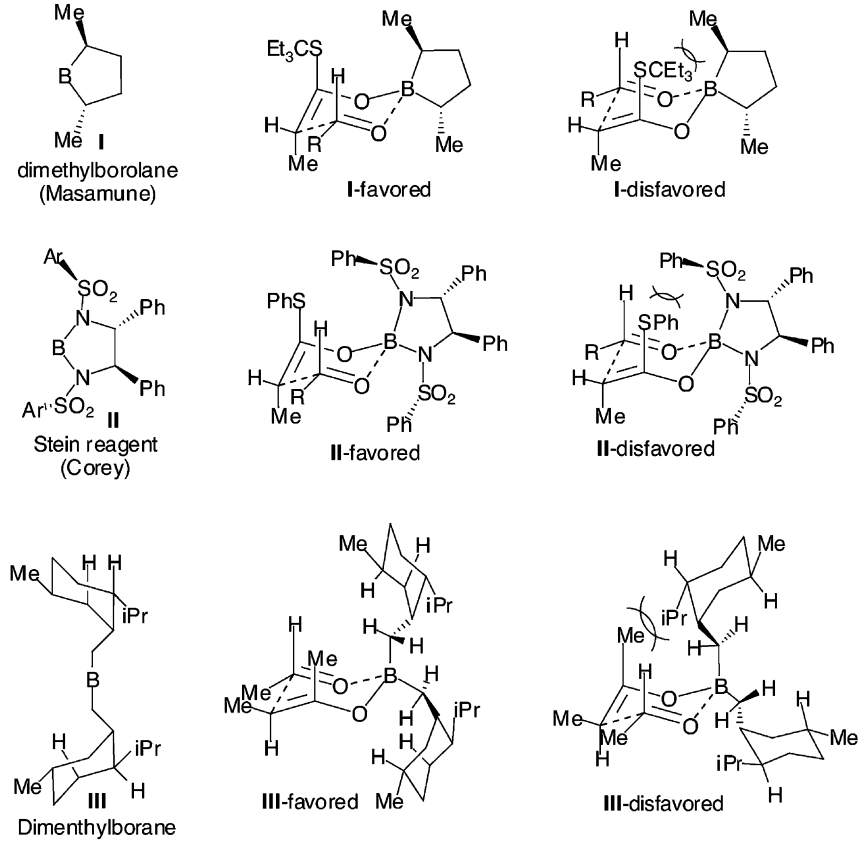


Fig. 4. Several known chiral auxiliaries along with transition structures showing favored and disfavored reaction pathways.

rings. In this case, the identification of a more rigid scaffold would be advantageous since a less conformational mobile auxiliary would allow fewer reactive species. Further examination of the transition structures identified by Gennari et al. indicate that the relevant stereodiscriminating groups are the isopropyl groups.

As can be seen from the aforementioned transition structures, the groups residing on the carbon atom adjacent to the boron atom (i.e. the Me groups for the Masamune dimethylborolane) incur disfavorable 1,3-diaxial interactions with select groups (i.e. $SCEt_3$ for the Masamune dimethylborolane) making one of the possible transition states less stable. A common feature of all the three examples is that substituents on the carbon atom adjacent to the boron play a vital role in discriminating one pathway over the other. The position and orientation of these substituents (stereodiscriminating groups) were assumed to be the predominant factors steering the stereochemical course of the reaction. The remaining part of the ligand can be viewed as a simple scaffold which fixes these stereodiscriminating groups in certain positions.

For any set of such key stereodiscriminating groups, it then seems reasonable that a number of different ligand scaffolds could be employed which present these groups in a similar manner. In an effort to determine if database searching could be used to identify such scaffolds, the three known chiral boron auxiliaries described have been examined in this context. As a first step, the question of whether it is possible to identify lead structures containing motifs corresponding to the known chiral ligands has been addressed. To this end, the program CAVEAT was used to extract structures from three-dimensional databases using search criteria which define the key elements of several boron aldol/allylation auxiliaries.

3.2. CAVEAT overview

The CAVEAT program is a search engine which retrieves molecules from a database of three-dimensional structures with bonds that match a vector relationship specified in the query. Bartlett and coworkers have used CAVEAT to identify lead structures in the design of a number of enzyme inhibitors [17]. In a typical CAVEAT search a bond is treated as a vector; one of the atoms in the bond, the base, defines the location of the vector and the tip atom determines the orientation. The relative orientation of several bonds is defined by the combination of such vectors taken in a pairwise fashion. For more vectors, more pairwise combinations are required to define the problem. CAVEAT can then be used to search structural databases for these vector pair relationships. A vector pair is defined uniquely by four parameters—

the distance between the two base atoms, the dihedral angle between the vectors and the two exterior angles for each vector. Queries that involve specific relationships among three or more vectors are defined as combination of vector pairs. The user needs to specify the tolerance for each pair of vectors in the base to base distance parameter (Å) and for the permissible angle of variation in each vector (rad). This last parameter defines a cone in which the vector must lie. The program searches and retrieves the structures that fall within the specified tolerances. Since CAVEAT does not rely on an atom-by-atom match as would be the case for a search engine that matches Cartesian coordinates,[2] the length of the bond is irrelevant and the search will identify relationships even between bonds of different lengths.

Several source databases have been used with CAVEAT to address various problems [5,17] including the CSD [6], the CAS-3D (see footnote 1) the RCSB Protein Databank (PDB) [18], as well as computer-generated databases of tricyclic hydrocarbons (TRIAD) [5] and acyclic fragments (ILIAD) [5]. In the searches for rigid scaffolding elements, we were primarily looking for cyclic structures where the vectors to bonds which contained one atom that was attached to a ring. As such, the CSD, CAS-3D, and TRIAD source databases were employed in this study. These source databases need to be processed into vector databases prior to commencing with CAVEAT searches. In doing so, vectors that are not of interest are excluded. For example, vectors that are part of long alkyl chains are not useful in identifying rigid scaffolds. The vector databases used in this work, *linked* (from the CSD), *met* (from the CSD), *cast* (from CAS-3D), and *triad* (from TRIAD) were comprised of vectors that contained one atom that was attached to or part of a ring.

CAVEAT initially performs a very rapid search of the large vector databases retrieving molecules with bonds that match the input vectors as the sole criteria for retrieval. Further, time-intensive classification and evaluation is then performed on the relatively small number of hit structures retrieved. CAVEAT facilitates examination of the total list of hits by classifying them on the basis of several factors that can be modulated by the user. The template core from which the vectors originate is identified, cores are analyzed for degree of similarity, and the leads ("hits") are clustered into several groups according to core type. The details of each step have been described by Lauri and Bartlett [5a]. The final results are a file containing representative structures ("top hits") for each group in which the parent framework is highlighted.

3.3. Screening for chiral ligand scaffolds

Using CAVEAT we desired to identify useful ligand scaffolding elements that can present the stereodiscriminating groups needed in stereoselective boron aldol and allylation reactions. The structures of the known chiral auxiliaries that exhibit high stereoselectivity were taken as the starting point. Among the many possible reagents that could be used as starting points, the Masamune dimethylborolane, the Corey Stein reagent, and dimenthylborane were employed here. As discussed (Fig. 4), these chiral ligands can be divided into stereodiscriminating elements and scaffolding elements. With CAVEAT, searches were initiated to identify structures containing vectors corresponding to the stereodiscriminating groups and the attachment point of the chiral ligand to the reactants (i.e. the B–C bonds) from known ligands. It was anticipated that the cores of the hit structures could be translated into scaffolds which contain the boron and present the necessary stereodiscriminating groups in a manner that would provide a stereoselective reaction.

Vectors for the database searches were taken from the geometries of the known ligands as found in the lowest energy minimized aldol and allylation transition states. These transition states were calculated for simplified structures using methyl substituted aldehydes and enolates/allyls (Fig. 5, $R^1 = R^2 = R^3 = Me$). The vectors for target **I** (the Masamune reagent) are the C–B bonds in the five-membered ring and the stereodiscriminating C–CH_3 bonds of the five-membered ring (Fig. 5). Similarly, the vectors employed for other ligands are used to define the connection of the scaffold in the transition structure and the stereodiscriminating groups (Fig. 5). For Stein reagent **II**, these are the C–B bonds for the five-membered ring for the former and the sulfur–aryl bond for the latter. For dimenthylborane **III**, the cylcohexane axial C–C bonds connecting toward the boron were chosen as the vectors to define the relation of the scaffolding cyclohexane rings to the transition structure. Use of these vectors rather than the C–B bonds would lead to a wider variety of structures since the tips of the

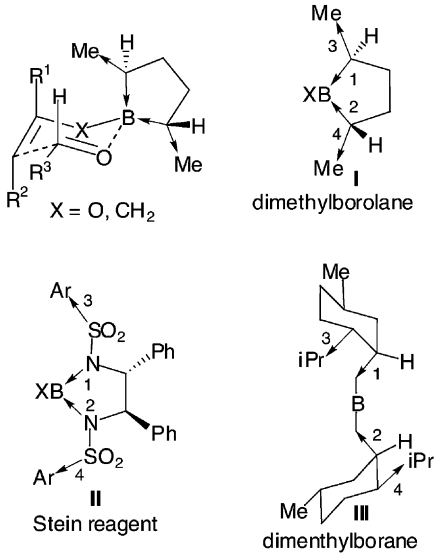


Fig. 5. Vectors used for the database searching (indicated by arrows) for **I–III**.

[2] We are currently examining whether Cartesian coordinates searching offers any advantage.

vectors do not share an atom. The C–C bonds leading from the cyclohexane scaffolds to the iPr groups were chosen as the second set of vectors since an analysis of the transition structures of Gennari et al. [11a] indicated that these groups were important to the stereochemical course of the reaction.

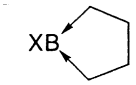
This analysis assumes that the most critical features of a stereodiscriminating group are its ultimate placement and orientation. An advantage of using a vector-based search to identify scaffold cores is that the position of the stereodiscriminating groups and the direction in which these groups are pointed are incorporated into the search. For example, if a database search for target **II** (the Stein reagent) were performed using the Cartesian coordinates corresponding to the boron and the carbons of the aryl attached to the sulfonyls, the resultant leads will not necessarily contain a bond that could orient the aryl groups as illustrated in Fig. 5.

CAVEAT finds molecules with matching vector relationships by taking combinations of vector pairs from the query and comparing with vector pairs in a database. The actual database that is searched is comprised of vector pairs pregenerated from a source database (i.e. the CSD) by the cavinitvector script supplied with the CAVEAT program. In defining a new search with a set of vectors, CAVEAT takes all possible combinations of the vectors as vector pairs by default. For example, the vector pairs of the target **II** (Corey Stein reagent) are (1, 2), (2, 3), (3, 4), (1, 4), (1, 3) and (2, 4). With the exception of the Masamune dimethylborolane, the default vector pairing was found to work well in that the searches were rapid (<10 min) and generated reasonable matches. For the Masamune dimethylborolane **I**, the pairwise combinations of four vectors led to an undefined core because vector pairs (2, 4) and (1, 3) originate from the same base. To correct this problem, the following vector pairs for the CAVEAT search, (1, 2), (1, 4), (3, 2) and (3, 4) were employed for this structural framework. The other target structures (Corey Stein reagent **II** and dimenthylborane **III**) do not have vectors that originate from the same base.

Classification was done by taking the smallest core geometry (eliminates any fused rings from the whole core which do not directly connect to the vectors) as the similarity measure to compare the molecules. To aid in the evaluation of the various structures, the hits were scored by RMSD relative to the input vectors.

3.4. Target *I*

Initial searches of the *linked* database with the vectors from the Masamune dimethylborolane **I** provided few lead structures. In addition, these leads did not correspond well with the framework in **I**. This result is a consequence of the vector pairs allowed during the construction of the linked database from the CSD; namely, convergent vector pairs within a ring were not allowed. As such, a new vector database, *met*, was constructed from the CSD in which such convergent vectors were allowed provided the vectors converged on a metal atom and the bases were linked by another molecular fragment.



convergent vectors from I

A search of the *met* vector database was carried out using the Masamune dimethylborolane **I** vectors with a 0.5 Å vector pair distance tolerance and 0.4 rad angle tolerance for each of the vectors in a vector pair, which is abbreviated as (0.5, 0.4, 0.4). With these tolerances 17,035 structures were retrieved. When classified, these hits divided into 444 clusters with 60 main groups and 384 subgroups. As expected, smaller vector tolerances (0.35, 0.25, 0.25) resulted in fewer hits: 4902 structures which classified into 92 clusters with 18 main groups and 74 subgroups. Since the template cores of the subgroups were subsumed by the template cores of the main groups, only the main groups needed to be examined to get a good overview of the scaffolding templates identified. The structures VAMNOM, ACACBE02, RBORMY, and CUNKEB with scores 0.132, 0.186, 0.188, and 0.191, respectively are the representatives for the first five groups from the search of the *met* database (Fig. 6). The template core characteristic of each group as identified by the classification protocol is highlighted in bold.

Group 1 with the five-membered ring "core" contained the largest number of hits (9404). A few illustrative structures from Group 1 are shown in Fig. 7. The large number of structurally diverse hits in this group indicates that the desired vector parameters are maintained in a five-membered ring scaffold even with different substitution patterns. In addition, the five-membered ring is also a key component of the template cores of Groups 3, 4, and 5. The six-membered ring core found in Group 2 suggests that boron reagents based on this ring system may also function as useful chiral auxiliaries since the same geometric elements needed to appropriately place the boron and stereodiscriminating groups are present.

It is gratifying to note that the best hit (VAMNOM) from the search that started with the Masamune dimethylborolane **I** is the crystalline complex of 2-(trimethylsilyl) borolane reported by Masamune (Fig. 6) [14]. The parent Masamune dimethylborolane, or derivatives thereof, are not present in the CSD. In the absence of experimental structural data, the vectors used in the database search were obtained from a theoretical structure (the favored transition state for the Masamune dimethylborolane in Fig. 4). The identification of this closely related compound validates (1) the use of vectors from calculated structures and (2) the ability of this database searching method to retrieve usable structures. The allylboration with the allyl reagent derived from 2-(trimethylsilyl) borolane has been shown to exhibit higher enantioselectivity than the dimethylborolane due to the larger steric effect induced by the bulkier trimethylsilyl group. Furthermore, the compounds of Group 60 such as COKRAV (Fig. 6) are

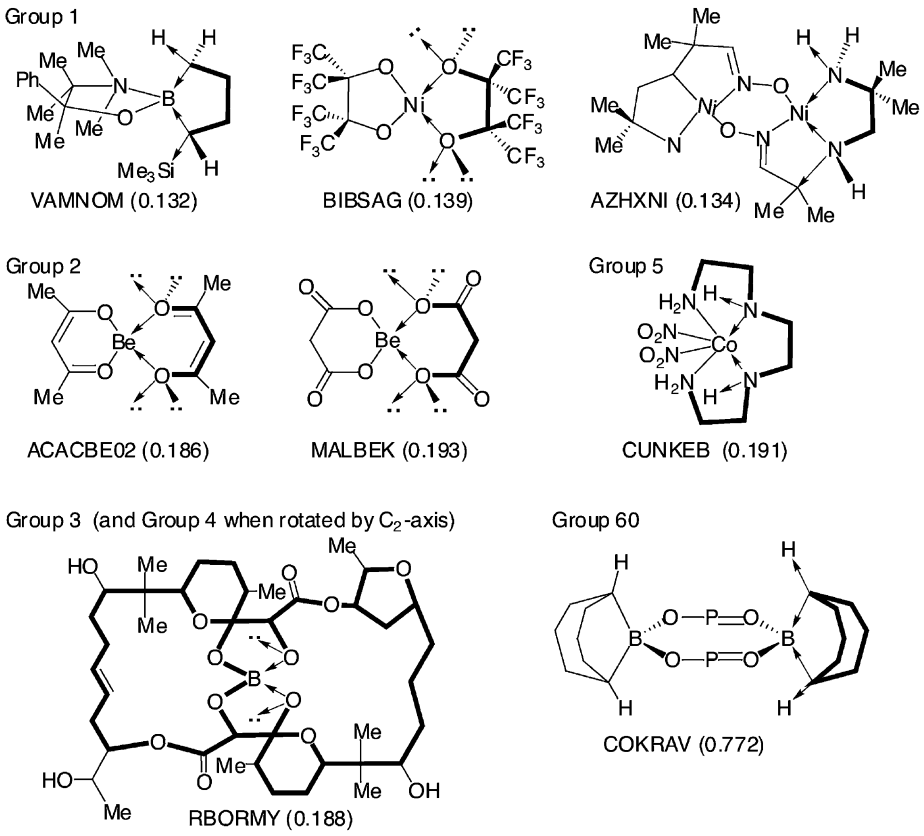


Fig. 6. CAVEAT hits from *met* for Masamune dimethylborolane **I**.

reminiscent of the 9-borabicyclodecane chiral auxiliaries developed by Soderquist et al. [19]. These results support the validity of this method for the identification of relevant ligand scaffolds.

The *triad* and *cast* vector databases do not contain vector pairs in which the tips can converge in a ring as is possible with the *met* database. Nevertheless, these databases were examined against the vectors from **I**. The results will be fundamentally different since vector pairs converging on the same atoms (i.e. vectors 1 and 2 in targets **I** and **II**) are not present. These searches were done to determine if useful new scaffold leads could be suggested from databases

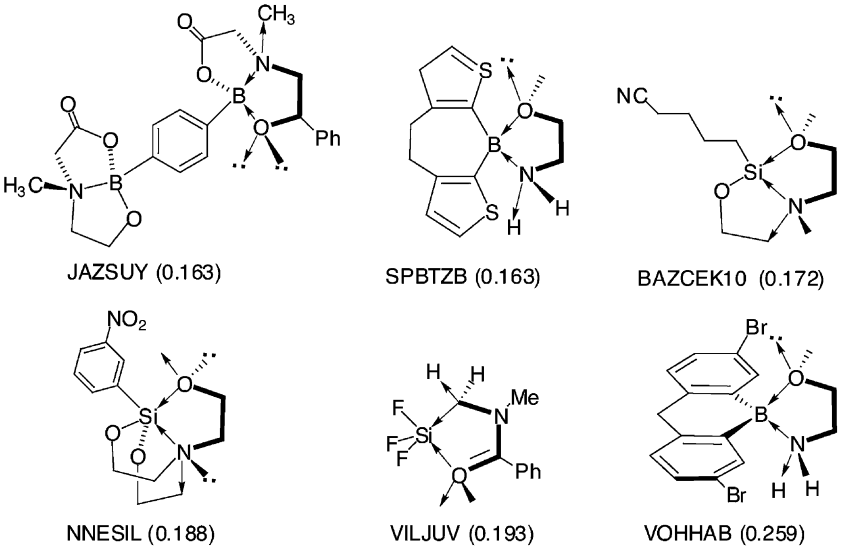


Fig. 7. Group 1 CAVEAT hits for the Masamune dimethylborolane **I** using the *met* database.

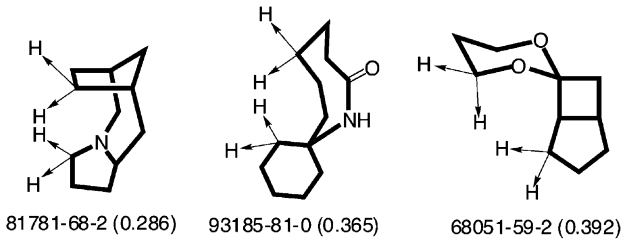


Fig. 8. CAVEAT hits for the Masamune dimethylborolane **I** using the *cast* database.

lacking the exact connectivity found in the target structure. Indeed, searches of these databases using the vectors from **I** generated quite a number of interesting molecular frameworks. A search of *cast* generated 5406 hits when a tolerance specification of (0.4, 0.3, 0.3) was employed. These hits divided into 15 main groups and 280 subgroups (Fig. 8). A search of *triad* with a tolerance specification of (0.5, 0.4, 0.4) provided 13,806 hits comprising 98 main groups and 139 subgroups (Fig. 9). Due to the nature of the vectors selected in the construction of the *triad* and *cast* databases, the matching vectors are exocyclic C–H bonds in contrast to the *met* database where many vectors correspond to endocyclic bonds. Although structurally rigid, the leads from the *triad* database (Fig. 9) are complex ring systems which are not reasonable starting points. The leads arising from the *cast* database (Fig. 8) are more interesting, especially the relatively simple spiro structure 93185-81-0 from which a potentially interesting boron auxiliary could be easily designed.

3.5. Target *II*

The basic structural frameworks of **I** and **II** are similar since both are based on five-membered rings containing a boron atom; however, the vector arrangements are different. In **II**, vectors 3 and 4 are one bond away from vectors 1 and 2, respectively. In **I**, vectors 3 and 4 share base atoms with vectors 1 and 2, respectively (see Fig. 5). Because of the additional distance between the vectors in **II**, a greater number of different structural scaffolds can span these vectors. Overall, the same tolerances for **II** gives a smaller number of hits than **I** and the vector errors are larger; however, the number of groups is larger for **II** compared to **I**. With tolerances of

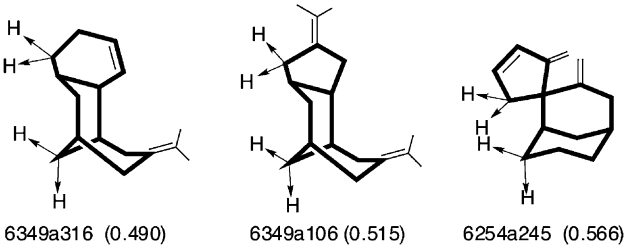


Fig. 9. Group 1 CAVEAT hits for the Masamune dimethylborolane **I** using the *triad* database.

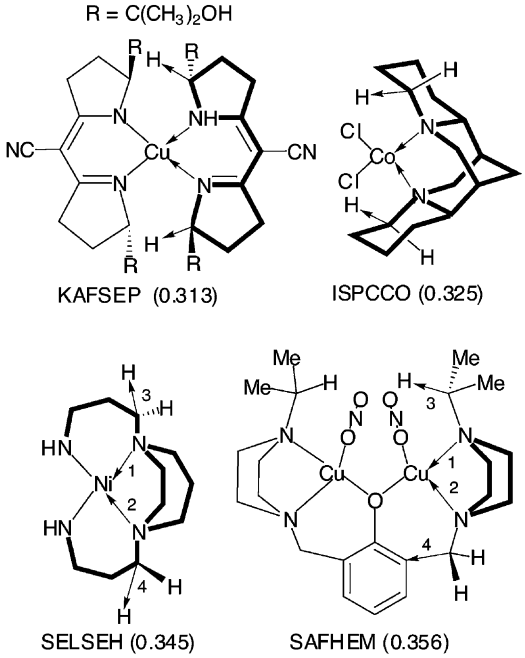


Fig. 10. Group 1 CAVEAT hits for the Stein reagent **II** using the *met* database.

(0.5, 0.4, 0.4) for the vectors in **II**, 2391 hits were retrieved from the *met* vector database that could be classified into 121 main groups and 54 subgroups. The best hits are a number of interesting organometallic complexes where the metal binds to ligands comprised of rigid frameworks (Fig. 10).

KAFSEP and ISPCCO are interesting leads which suggest creating a multicyclic boron auxiliary that would ensure limited conformational flexibility of the stereodiscriminating group substituents. Overall, KAFSEP and ISPCCO are better leads than SELSEH or SAFHEM because the vectors all originate from rigid rings that could ultimately be engineered into a chiral auxiliary. On the other hand, SELSEH is poor because the rings containing vectors 3 and 4 in this search would have to be removed due to overlap with the transition state; two open valences are necessary on a tetrahedral boron center which would replace the Ni in SELSEH. SAFHEM is a poor lead because vector 3 (corresponding to the vector defining one of the stereodiscriminating groups) is part of a flexible acyclic portion of the molecule. When incorporated into a chiral auxiliary, such a scaffold would not place the stereodiscriminating group in a single well-defined position. If the stereodiscriminating group is not fixed, it may flex to accommodate either enantiomer of the transition state equally well resulting in low product selectivity.

A search of *cast* with a tolerance specification of (0.5, 0.4, 0.4) for the vectors from **II** retrieved 3299 hits that were classified into 296 main groups and 112 subgroups (Fig. 11). The structures found are polycyclic hydrocarbons with complex fused and spiro ring systems which are not useful. Such leads satisfy the criteria of a rigid scaffold, but synthesis of boron auxiliaries based on these motifs will likely be difficult.

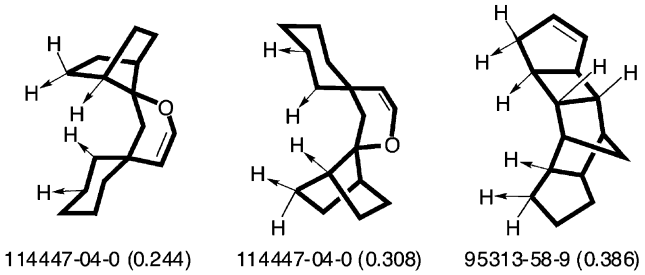


Fig. 11. CAVEAT hits for the Stein reagent **II** using the *cast* database.

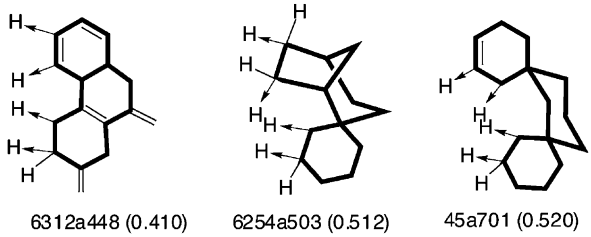


Fig. 12. CAVEAT hits for the Stein reagent **II** using the *triad* database.

A search of TRIAD with a (0.6, 0.5, 0.5) tolerance specification for the vectors in **II** yielded 3279 structures with 94 main groups and two subgroups (Fig. 12). Notably, when the same tolerance specification (0.5, 0.4, 0.4) used in other searches was employed, only eight hits were found. Examination of the resultant structures revealed that some hits were poor leads. For example, the sp^2 hybridized centers of 6312a3448 and 45a701 would be difficult to incorporate into a synthetically viable ligand and would mean that the stereodiscriminating group would not originate from an asymmetric center. While it may be possible to still create a usable chiral auxiliary by enforcing the position of the stereodiscriminating groups from distal stereogenic centers, such leads can be excluded during the CAVEAT search by specifying in the input parameters that the ends of the vectors be comprised of a certain atom types or hybridizations. Tricyclic compound 6254a503 is the best lead of the structures illustrated in Fig. 12. While synthesis of an auxiliary corresponding to 6254a503 would be difficult, the simpler spirocyclic substructure common to 6254a503 and 45a701 is eminently reasonable and is also similar to the motif found in 93185-81-0 (Fig. 8) when the vectors from **I** were screened against the *cast* database.

3.6. Target *III*

Target **III** is quite different from the first two targets since the vectors describing the attachment to the boron are not bonds to the boron atom. A search of the *met* database for the vectors in **III** yielded 2299 hits which divided into 257 clusters (179 main groups and 78 subgroups). The top hits were mostly complexes of porphyrins or complexes of substituted ethylene diamines that did not suggest useful scaffolds for the reactions under investigation.

A search of the *cast* database provided a number of leads suggesting potentially interesting scaffolds (Fig. 13). With tolerances of (0.6, 0.5, 0.5) for the vectors from **III**, 11,668 hits were retrieved which were classified into 928 clusters (704 main groups and 224 subgroups). When tolerances of (0.5, 0.4, 0.4) were employed, 542 hits were found which were classified into 18 main groups and 31 subgroups. The best hit from this search, 115187-31-0, as defined by the lowest RMSD value, reproduces the cyclohexane rings in the same orientation as in **III** (Fig. 13). Indeed, the correlation

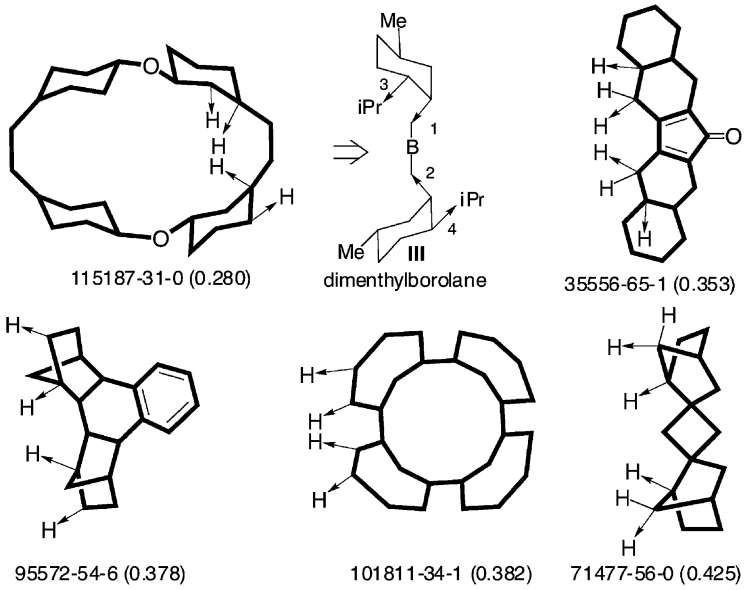


Fig. 13. Top CAVEAT hits for the dimenthylborane **III** using the *cast* database.

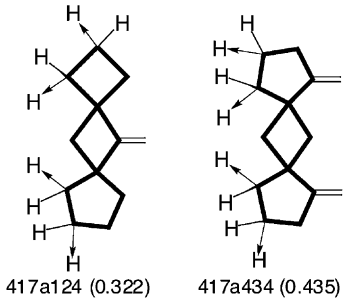


Fig. 14. Top CAVEAT hits for the dimenthylborane **III** using the *triad* database.

between the starting structure used to generate the vector inputs and the found structure is striking. Other top hits (35556-65-1, 95572-54-6, 101811-34-1 and 71477-56-0) are quite interesting in that the vectors originate from sp^3 centers that are part of a rigid molecular framework. Upon suitable modification with the synthetic viability in mind, these scaffolds may lead to a novel chiral ligands for selective aldol or allylation reactions.

A search of the *triad* database employing the tolerances (0.6, 0.5, 0.5) for the vectors in **III** provided 3279 hits which were classified into 94 main groups and two subgroups. A variety of fused as well as spiro tricyclic structures were observed, two examples of which are illustrated in Fig. 14. In contrast to the triad hits for **I** and **II**, these structures are more reasonable starting points for the design of a chiral auxiliary by virtue of their symmetry which simplifies possible syntheses.

4. Conclusion

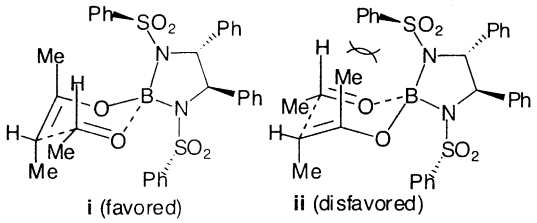
The utility of searching three-dimensional databases of small molecules to identify the motifs corresponding to a series of known chiral auxiliaries is outlined. As part of this effort, an analysis of the key elements of three known chiral auxiliaries is presented. From these connecting elements and stereodiscriminating elements a series of bond vectors could be selected which uniquely defined the chiral auxiliaries. The CAVEAT program was then used to screen the Cambridge Structural Database, portions of the CAS three-dimensional database, and the TRIAD tricyclic structure database for scaffolds containing these elements. A good number of structures containing the scaffolds characteristic to the known starting points were identified indicating that this method can be used to identify chiral ligand structural motifs. For one of the auxiliaries investigated, the best hit obtained from CSD-based database, *met* reproduced an auxiliary which has been reported to have a better enantioselectivity. In addition, alternate structural motifs were found that suggested alternative possible ligands. The use of this protocol to identify new ligand motifs with these and other databases of three-dimensional structures may prove fruitful. Further work in this direction is under way and will be reported separately.

Acknowledgements

Financial support was provided by the University of Pennsylvania, the National Institute of Health (GM-59945), Merck Research Laboratories, and DuPont. Acknowledgment is made to the donors of the Petroleum Research Fund, administered by the American Chemical Society, for partial support of this research. We thank Georges Lauri for helpful discussions and advice regarding the implementation of the CAVEAT program and Paul A. Bartlett for access to CAVEAT.

Appendix A

The two simplified diastereomeric chair transition states **i** and **ii** illustrated next were examined using molecular mechanics calculations as models of **II**-favored and **II**-disfavored (Fig. 4) for the aldol reactions of the Corey Stein reagent. The transition state **i** which leads to the major product was found to be 2.6 kJ/mol lower in energy than the transition state **ii** which leads to the minor product. This energy difference corresponds to an 83:17 product ratio as calculated using a Boltzmann distribution. These results compare well to the closely related aldol reaction with propionaldehyde (CH_3CH_2CHO) and diethyl ketone examined experimentally by Corey et al. where a 99:1 ratio was observed (see [15]).



The geometries of **i** and **ii** were optimized using an MM2* derived force field in MacroModelwith the following additional parameters:

1	B3-N3	1.4200	4.0000	0.0000	0000	0000	A 3		
1	B3-N2	1.6500	2.0000	0.0000	0000	0000	A 3		
2	00-B3-00	120.0000	0.4000	0000	0000	0000	A 3		
4	00-N3-B3-00	0.0000	0.0000	0.3500	0000	0000	0000	0000	A 3
4	00-N2-B3-00	0.0000	0.0000	0.4000	0000	0000	0000	0000	A 3

References

[1] E.N. Jacobsen, A. Pfaltz, H. Yamamoto (Eds.), Comprehensive Asymmetric Catalysis, Vol. I–III, Springer-Verlag, Berlin, 1999.

[2] (a) W.C. Still, MacroModel V6.5, Columbia University, Columbia, 1998;
(b) F. Mohamdi, N.G. Richards, W.C. Guida, R. Liskamp, M. Lipton, C. Caufield, G. Chang, T. Hendrickson, W.C. Still, J. Comput. Chem. 11 (1990) 440.

[3] (a) A. Bernardi, A.M. Capelli, C. Gennari, Aldol force field, J. Org. Chem. 55 (1990) 3576;
(b) A. Bernardi, A.M. Capelli, A. Comotti, C. Gennari, M. Gardner, J.M. Goodman, I. Paterson, Aldol force field, Tetrahedron 47 (1991) 3471;
(c) A. Bernardi, A. Cassinari, A. Comotti, M. Gardner, C. Gennari, J.M. Goodman, I. Paterson, Aldol force field, Tetrahedron 48 (1992) 4183;
(d) C. Gennari, S. Vieth, A. Comotti, A. Vulpetti, J.M. Goodman, I. Paterson, Aldol force field, Tetrahedron 48 (1992) 4439;
(e) A. Vulpetti, A. Bernardi, C. Gennari, J.M. Goodman, I. Paterson, Aldol force field, Tetrahedron 49 (1993) 685.

[4] A. Vulpetti, M. Gardner, C. Gennari, A. Bernardi, J.M. Goodman, I. Paterson, Allylation force field, J. Org. Chem. 58 (1993) 1711.

[5] (a) G. Lauri, P.A. Bartlett, J. Comput.-Aided Mol. Design 8 (1994) 51;
(b) P. A. Bartlett, CAVEAT V2.2, University of California at Berkeley, CA, 1995.

[6] F.H. Allen, O. Kennard, R. Taylor, Acc. Chem. Res. 26 (1983) 146.

[7] B.M. Trost, I. Fleming, C.H. Heathcock (Eds.), Comprehensive Organic Synthesis, Vol. 2, Pergamon Press, Oxford, 1991.

[8] (a) C. Gennari, R. Todeschini, M.G. Beretta, G. Favini, C. Scolastico, J. Org. Chem. 51 (1986) 612;
(b) Y. Li, M.N. Paddon-Row, K.N. Houk, J. Am. Chem. Soc. 110 (1988) 3684;
(c) Y. Li, M.N. Paddon-Row, K.N. Houk, J. Org. Chem. 55 (1990) 481, and references therein;
(d) F. Bernardi, M.A. Robb, G. Suzi-Valli, E. Tagliavini, C. Tromboni, A. Umani-Ronchi, J. Org. Chem. 56 (1991) 6472.

[9] (a) W.R. Roush, A.E. Walts, L.K. Hoong, J. Am. Chem. Soc. 107 (1985) 8186;
(b) W.R. Roush, A.M. Ratz, J.A. Jablonowski, J. Org. Chem. 57 (1992) 2047;
(c) R.E. Gawley, J. Aube, Principles of Asymmetric Synthesis, Pergamon Press, Oxford, 1996, p. 161 (Chapter 5).

[10] (a) D. Seebach, V. Prelog, Angew. Chem. Int. Ed. Engl. 21 (1982) 654;
(b) A. Bernardi, C. Gennari, L. Raimondi, M.B. Villa, Tetrahedron 53 (1997) 7705–7714.

[11] (a) C. Gennari, C.T. Hewkin, F. Molinari, A. Bernardi, A. Comotti, J.M. Goodman, I. Paterson, J. Org. Chem. 57 (1992) 5173;
(b) I. Paterson, Pure & Appl. Chem. 64 (1992) 1821;
(c) C. Gennari, D. Moresca, S. Vieth, A. Vulpetti, Angew. Chem. Int. Ed. 32 (1993) 1618;
(d) A. Bernardi, A. Comotti, C. Gennari, C.T. Hwekin, J.M. Goodman, A. Schlapbach, I. Paterson, Tetrahedron 50 (1994) 1227;
(e) A. Bernardi, C. Gennari, J.M. Goodman, I. Paterson, Tetrahedron: Asymmetry 6 (1995) 2613;
(f) C. Gennari, Pure & Appl. Chem. 69 (1997) 507;
(g) C. Gennari, S. Ceccarelli, U. Piarulli, K. Aboutayab, J. Braz. Chem. Soc. 9 (1998) 319.

[12] (a) D.G. Truhlar, K. Morokuma (Eds.), in: Proceedings of the ACS Symposium Series 721 on Transition State Modeling for Catalysis, American Chemical Society, Washington, DC, 1999, and references therein;
(b) E. Peña-Cabrera, P.-O. Norrby, M. Sjögren, A. Vitagliano, V. De Felice, J. Oslob, S. Ishii, D. O'Neill, B. Åkermark, P. Helquist, J. Am. Chem. Soc. 118 (1996) 4299;
(c) A. Vidal-Ferran, A. Moyano, M.A. Pericas, A. Riera, Tetrahedron Lett. 38 (1997) 8773.

[13] (a) S. Masamune, T. Sata, B.M. Kim, T.A. Wollmann, J. Am. Chem. Soc. 108 (1986) 8279;
(b) M.T. Reetz, P. Heitmann, Tetrahedron Lett. 27 (1986) 4721.

[14] R.P. Short, S. Masamune, J. Am. Chem. 111 (1989) 1892.

[15] E.J. Corey, R. Imwinkelried, S. Pikul, Y.B. Xiang, J. Am. Chem. Soc. 111 (1989) 5493.

[16] E.J. Corey, C.-M. Yu, S.S. Kim, J. Am. Chem. Soc. 111 (1989) 5495.

[17] (a) F.A. Etzkorn, T. Guo, M.A. Lipton, S.D. Goldberg, P.A. Bartlett, J. Am. Chem. Soc. 116 (1994) 10412;
(b) Z.Q. Tian, P.A. Bartlett, J. Am. Chem. Soc. 118 (1996) 943;
(c) J.H. Meyer, P.A. Bartlett, J. Am. Chem. Soc. 120 (1998) 4600;
(d) W.W. Smith, P.A. Bartlett, J. Am. Chem. Soc. 120 (1998) 4622.

[18] (a) H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, Nucl. Acids Res. 28 (2000) 235;
(b) Further information about the PDB is available at the website: http://pdb.ccdc.cam.ac.uk/pdb/.

[19] J.A. Soderquist, K. Matos, C.H. Burgos, C. Lai, J. Vaquer, J.R. Medina, S.D. Huang, Organoboranes for syntheses, ACS Symp. Ser. 783 (2001) 176–194.