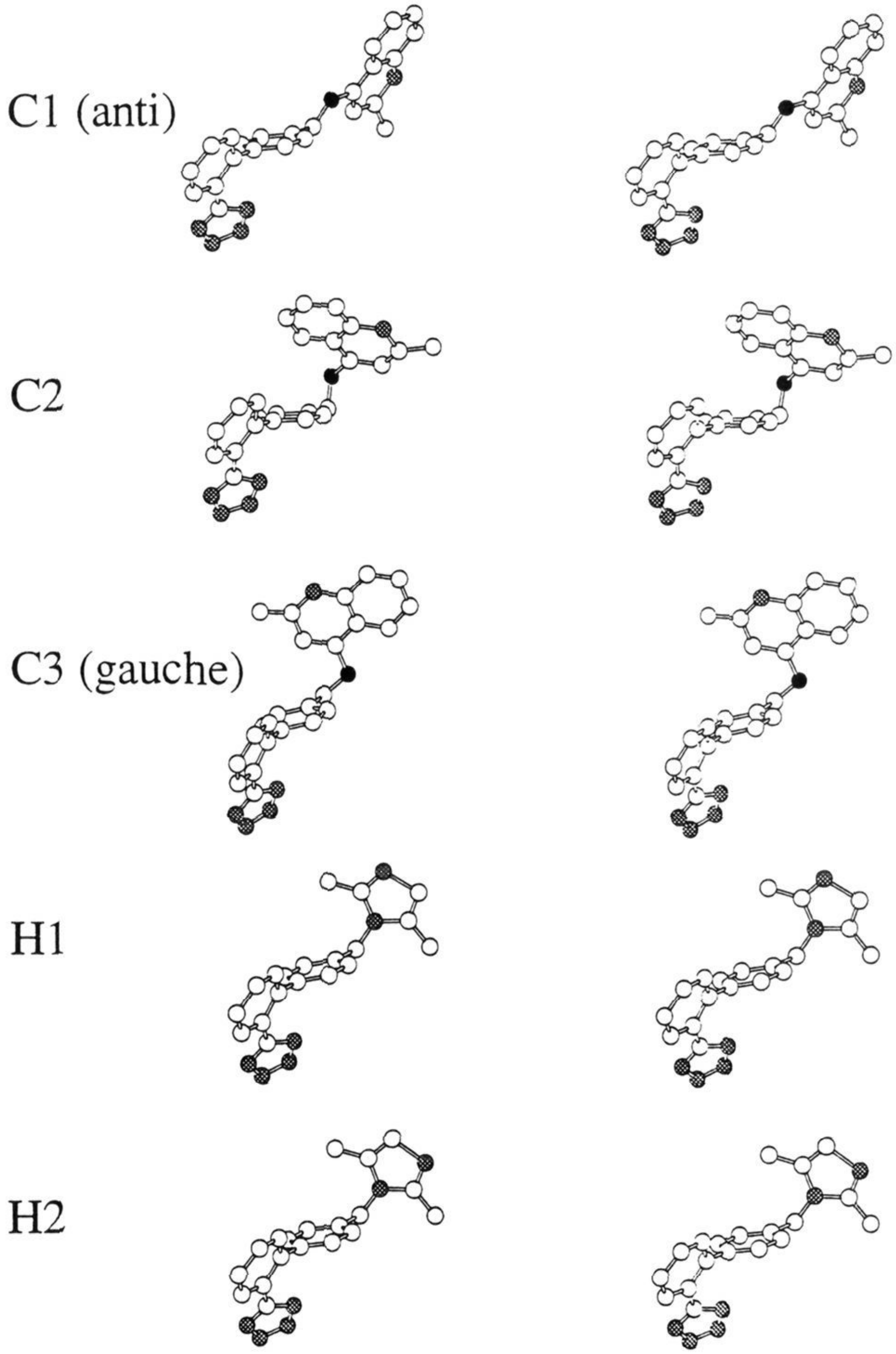

Figure 2. Conformations C1 (anti), C2, and C3 (gauche) of **1** and conformations H1 and H2 of **2**. Hydrogens are omitted for clarity: ○ = carbon, ▨ = nitrogen, ● = oxygen.

the two molecules. These conformations represent the candidates for the biologically active conformations of the two molecules. A model for the bioactive conformation of these AII antagonists has previously been proposed on the basis of chemical insight and extensive molecular graphics comparisons.[2] Our MSC, based on an unbiased comparison of total molecular volume, generates and prioritizes several alternative models for the AII antagonist pharmacophore. Among the best matches were structures closely related to the previously proposed model. In addition, our MSC has been enhanced with an option to allow matching of atoms or functional groups of similar chemical type. When MSC is done by chemical type, several interesting alternate models are again generated. Among the best models were structures nearly identical to the previously deduced pharmacophore model. The resulting pharmacophore model can be used as a frame of reference for the design of novel compounds. This allows the design of new ligands in the absence of structural information on the biological target.

II. Theory

A molecule is represented by a set of overlapping spherical atoms. The exposed surface of these spheres represents a molecular surface. Depending on the radii used, this could be, for example, a van der Waals (vdW) surface or an accessible surface as defined by Lee and Richards.[4] Here, we have chosen to use the all atom vdW surface. The vdW surface defines the boundary of a single molecules' volume.

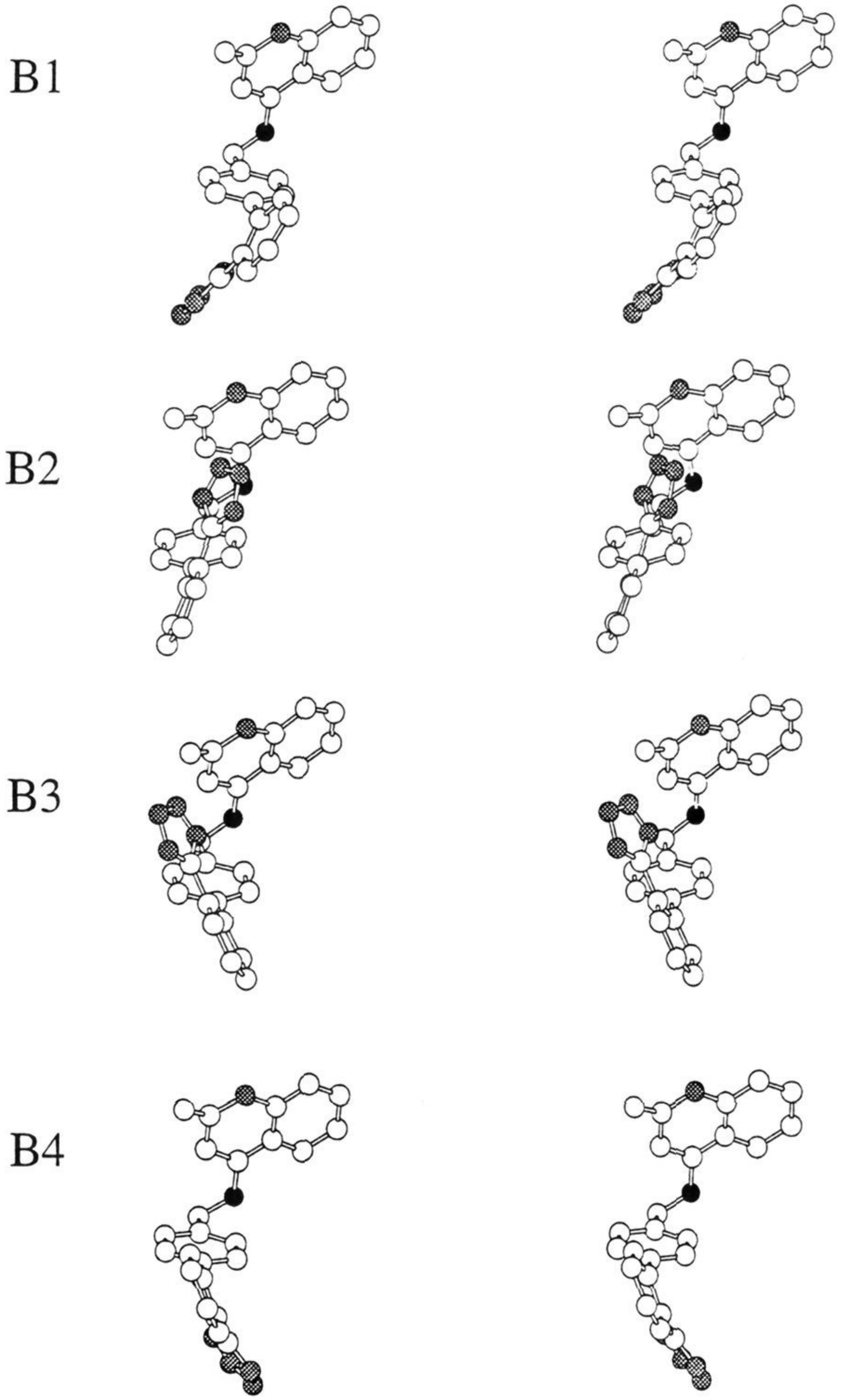

Figure 3. Four possible biphenyl orientations, B1–B4 shown for conformation C3 of 1. Hydrogens are omitted for clarity.

For two molecules, arranged such that their molecular surfaces intersect, the overlapping molecular volume can be calculated and used as a quantitative measure for shape comparison. Numerical approaches to this problem have been presented by a number of workers.[5] As pointed out by Connolly, the intersection of molecular volumes may be calculated using the inclusion–exclusion principle.[6] For two molecules (A and B), the intersection volume, $V(A \cap B)$, is given by

$$V(A \cap B) = V(A) + V(B) - V(A \cup B) \qquad (1)$$

where $V(A \cup B)$ is the volume of A union B, $V(A)$ is the volume of molecule A, and $V(B)$ is the volume of molecule B. The volumes required by eq 1 may be calculated analytically.[6,7] We have used the formulae of Connolly to calculate the terms $V(A)$, $V(B)$, and $V(A \cup B)$.[6] $V(A \cup B)$ can be calculated directly using Connolly's formulae; the formulae are applied to the A–B aggregate. Note, $V(A \cup B)$ can be calculated directly only when the surfaces are defined by a union of spheres, as in this work, but not with reentrant type surface representations.[8]

The magnitude of the intersection volume will depend on the relative position and orientation of the molecules being compared. In recent approaches to numerical volume matching, the relative position of the two molecules to be compared is fixed. In contrast, we maximize the volume overlap (the shape comparison measure) with respect to the relative positions of the two molecules. Mathematically, this is accomplished by inverting the sign of the function and then minimizing with respect to rigid

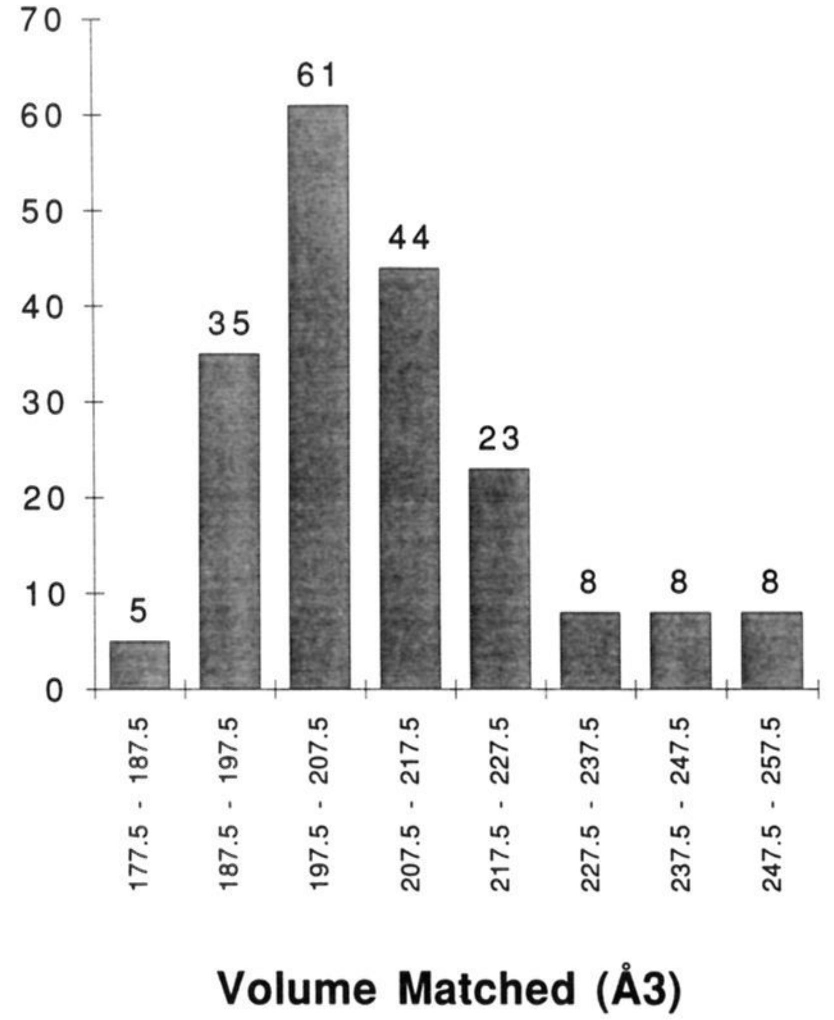

Figure 4. The highest intersection volume for each conformational pair plotted as a distribution for the 192 conformational pairs.

rotation and translation. For efficiency, analytical first and second derivatives of the volumes employed in eq 1 have been derived from Connolly's analytical equations for the volume.[6] These Cartesian derivatives are then transformed onto the variables for rigid rotation and translation. Expressions for the analytical derivatives (Cartesian and rotation/translation) were checked by comparison of the analytical and numerical first and second derivatives. Quaternions, a rotational coordinate system, have been used as the rotational variables in order to avoid discontinuities in the gradients which can occur with Euler angles.[9]

One concern in optimizing the alignment of two molecules based on molecular volume is the potential existence of many uninteresting local minima. Crevices and or atomic periodicity in the molecular surfaces might give rise to many minima which would trap the optimization. For example, consider optimizing the volume intersection of two diatomics, shown in Figure 1. It might be expected that a false minima, with only a single atom overlapped, would trap the optimization, but this problem is not observed.[10] Optimization of the diatomic system proceeds smoothly to a structure with both atoms of the diatomics overlapped. As a slightly more complicated example, overlap of one toluene molecule onto a second toluene molecule, starting with the two molecules in the same plane with edges touching slightly, proceeds smoothly to exactly superimpose the two aromatic rings. However, matching the methyl groups, for a complete shape match, is dependent on the starting orientation.

With energy minimization methods, optimization from a single starting geometry does not necessarily lead to the global energy minimum.[11] Similarly, with shape comparison, an optimization from a single starting alignment will not necessarily lead to the global maximum. A search procedure has been developed to identify the various local maxima. Alignments are selected randomly from a grid in rotation space to initiate multiple shape matching optimizations. A list of the unique maxima is kept. For any pair of molecules, the magnitude of the intersection volume is ultimately limited by the volume of the smallest molecule in the comparison.

We considered methods, such as simulated annealing, dynamics, and the genetic algorithm, as alternative approaches to explore if our volume optimizations were not well behaved due to discontinuities in the gradients or local trapping became an issue. We found the current gradient based optimization method to be well-behaved, fast, and useful.

Our MSC method has been enhanced with an option to allow discrimination between groups with different chemical properties. Atoms or groups of atoms may be assigned to different classes based on specific properties such as electrostatic potential, hydrogen-bonding ability, or hydrophobicity. This enables matches based on criteria such as alignment of hydrophobic groups or hydrogen-bond-acceptor groups. Considering each class separately, eq 1 is applied to the atoms in each class. The total intersection volume is then simply the sum of the intersection volumes for each class. By assigning each class a weighting, emphasis can be placed on different classes. Note, because atoms of one class are no longer "blocked" by atoms of other classes, overcounting of the overlap regions between classes does occur with this approach. In our experience, however, this has not affected the usefulness of the comparisons derived.

III. Computational Details

A program, Skinny, has been written to implement volume- and skin-based[12] molecular shape comparisons. Currently, only pairwise shape comparisons are implemented. One set of atoms (which may be one or more molecules) is rigidly rotated and translated to maximize the shape comparison to a second stationary set. Skinny is written in FORTRAN and runs on both Convex and Silicon Graphics platforms under UNIX. Computationally intensive subroutines have been coded with vector performance in mind. Scalar/vector CPU time ratios on the Convex C220 using a single processor are ca. 3. Skinny is controlled with keywords; atomic coordinates may be input via the Brookhaven Protein Databank file format.

A number of optimization algorithms are available within Skinny, including steepest descents, conjugate gradients, variable metric, and Newton–Rhapson methods.[13] Current experience indicates the best performance is obtained using the BFGS variable metric algorithm. At the start of each optimization, analytical second derivatives are obtained. The Hessian is then inverted and the BFGS numerical updates are done on the inverted Hessian. A line search with cubic interpolation is performed at each step of the optimization. Optimizations are terminated when the root mean square of the gradient vector falls below 0.2 $Å^3/Å$.

To identify multiple shape comparison maxima, a search procedure generates multiple starting geometries for optimization. We have found that the best shape match is often located by starting optimization from a "standard orientation". In the standard orientation, the principle axes of rotation for the moving molecule are aligned with the principle axes of rotation for the stationary molecule such that the axes of largest moments, the axes of

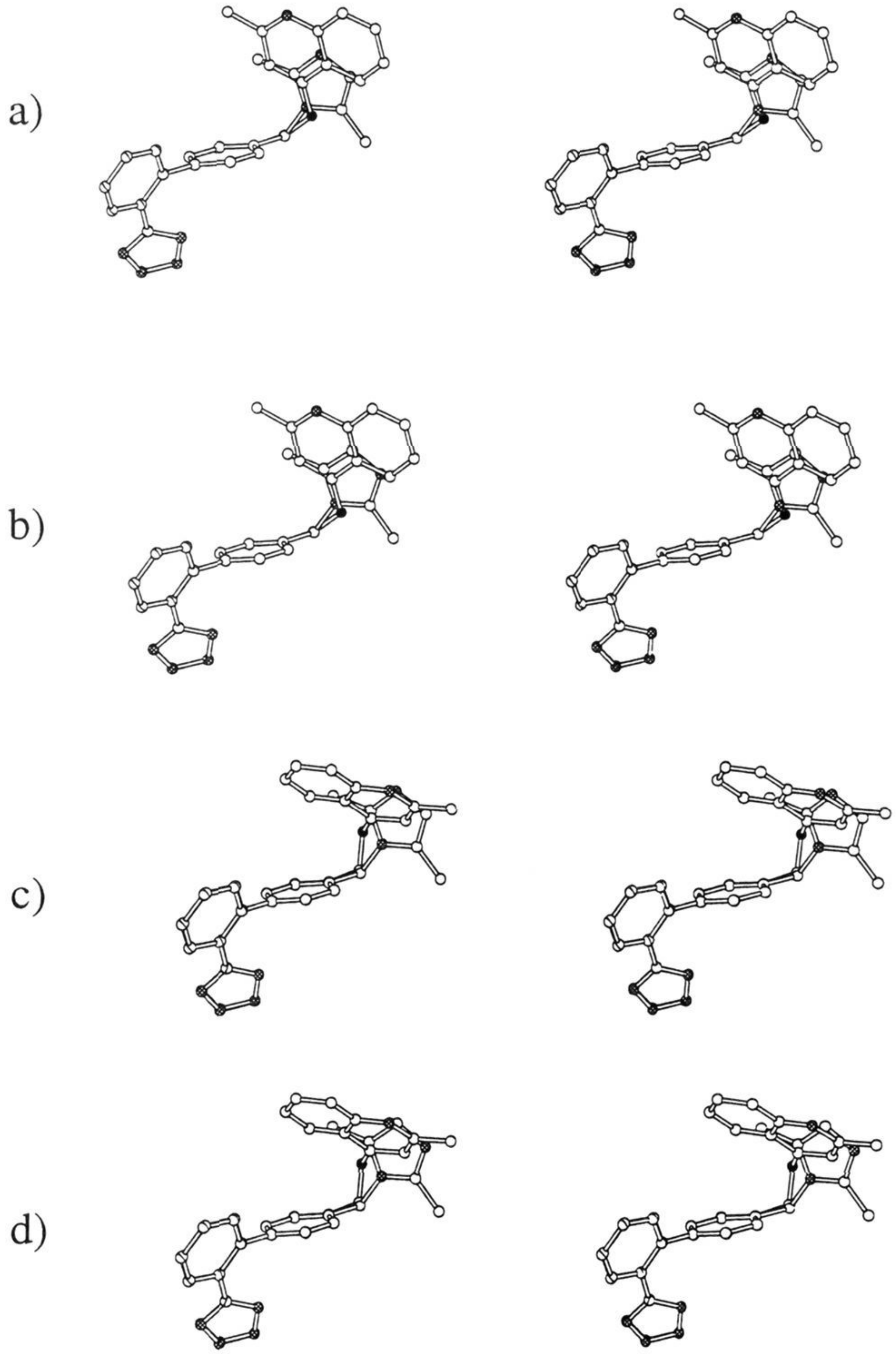

Figure 5. Four examples of shape match alignments for high scoring conformational pairs: (a) conformations C3B1 and H1B1 (252.9 Å^3), (b) conformations C3B1 and H2B1 (252.8 Å^3), (c) conformations C2B1 and H1B1 (240.1 Å^3), (d) conformations C2B1 and H2B1 (239.5 Å^3).

intermediate moments, and the axes of smallest moment are coincident and the centers of geometry are superimposed. From a standard orientation, rotation of 180° about any of the three principle axes generates another standard orientation. For efficiency, these four standard orientations are used as starting points for the first four shape comparison optimizations. Next, random searching is initiated. For subsequent optimizations, the moving molecule is first rotated randomly about its axis of largest moment in steps of 30° and then it is rotated about its center of geometry to align the axis of largest moment with a randomly selected direction. The vertices of a tesselated regular icosahedron[14] were used to define 42 evenly spaced directions. From each new starting alignment, a shape comparison optimization is initiated. The optimized intersection volume and the rotation and translation transformation are compared with those previously found and a list of unique matches is kept. A maxima is judged to be unique if its intersection volume differs from all previous maxima by more than 0.1 Å^3, if any element of the rotation transformations differs by more than 0.05, or if any element of the translation transformation differs by more than 0.05 Å. Searching can be halted when each unique maxima has been revisited a predetermined number of times or when a specified CPU

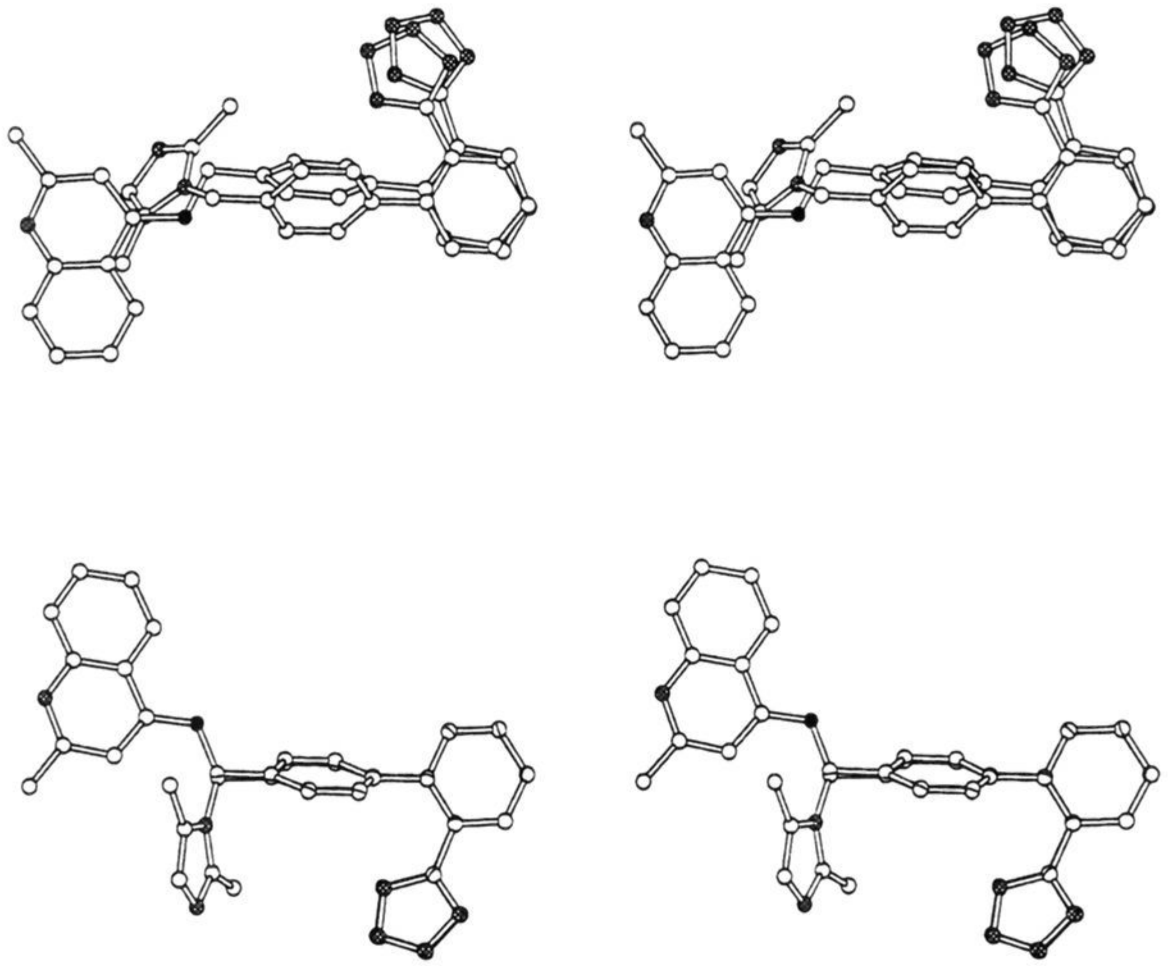

Figure 6. Two examples of less satisfying shape matched conformational pairs: upper, conformations C1B1 and H1B2 (199.6 Å^3); lower, conformations C1B1 and H1B3 (189.4 Å^3).

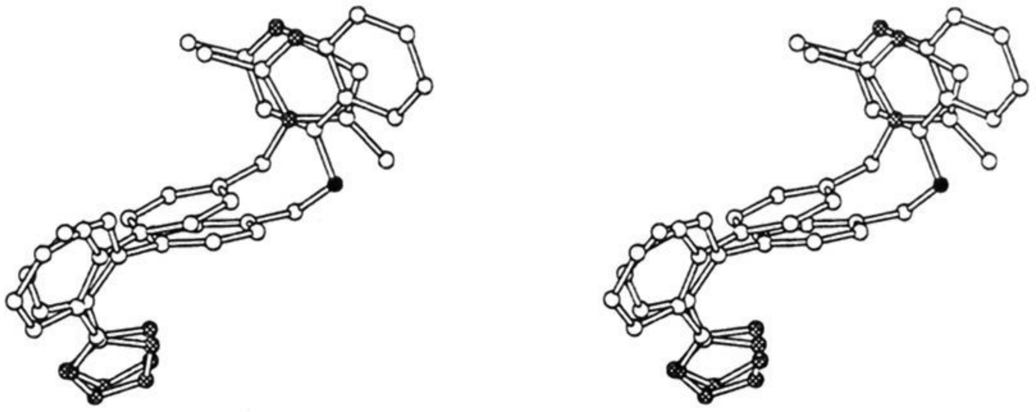

Figure 7. Stereoview of a previously proposed overlap of **1** and **2**.[2] Conformations C3B1 and H1B1 are shown. Similar alignments were found for the other three biphenyl orientations.

time limit has been exceeded. The atomic radii used in this work are provided as supplementary material.

IV. Results and Discussion

Both series of angiotensin II antagonists contain a number of single bonds which allow conformational flexibility. MSC, as discussed above, involves the rigid body overlap of a single molecular geometry with another. Our strategy for the inclusion of conformational flexibility in the MSC analysis starts with a list of candidate, low-energy conformations for the two molecules and assumes that the bioactive conformations of the two molecules are represented somewhere in the list. These conformations were chosen to follow previous work; however, alternate conformations could have been generated by systematic or random Monte Carlo methods. Each conformation of the first molecule is paired with each of the conformations of the second molecule. For each of these conformational pairs, an MSC comparison, which may generate multiple MSC maxima, is initiated. This process identifies the conformational pairs with high shape similarity, as measured by the intersection volume.

In the previous study, model systems **1** and **2** were used for the conformational analysis and comparison of the two series exemplified by ICI D8731 and DuP 753.[2] The conformations for **1** and **2**, previously derived by conformational analysis and molecular mechanics calculations, were also used for this work. For reference, these conformations are shown in Figures 2 and 3. For **1**, three conformations of the biphenylyl quinoline linkage (designated as C1–C3) were considered. For **2**, two conformations of the biphenylyl imidazole linkage (designated as H1 and H2) were considered. These are shown in Figure 2. Each of these may be combined with one of four possible orientations of the biphenylyl tetrazole moiety, designated as B1–B4. The four biphenylyl tetrazole orientations are shown in Figure 3 for the C3 conformation of **1**. The combination of the C3 conformation of the biphenylyl quinoline with the B1 conformation of the biphenylyl tetrazole moiety is then denoted as the C3B1 conformation of **1**. This leads to 12 conformations for **1** and eight conformations for **2**. Since each of these conformations is chiral, each conformation has a mirror image of equal energy on which it is not superposable. A complete MSC requires the comparison of each of the eight conformations of **2** with the 12 conformations of **1** and their 12 enantiomeric conformations. This yields a total of 192 unique pairings.

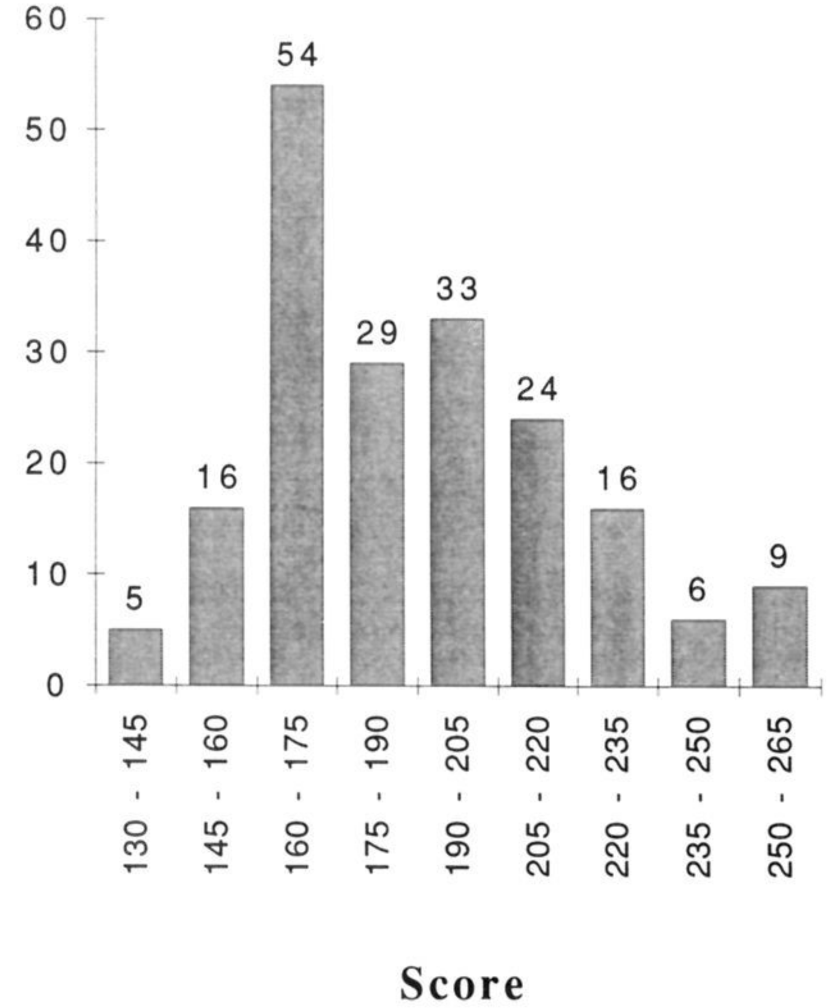

Figure 8. The highest score (volumes weighted by class) for each conformational pair plotted as a distribution for the 192 conformational pairs.

An MSC was carried out for each of the 192 unique pairings, with each search limited to 30 CPU min on a Silicon Graphics R4000-50 workstation: a total of 96 CPU h. For each conformational pair, 24–30 optimizations of the shape comparison were executed, starting from the four standard orientations + 20–26 random orientations. Typically, only four to eight unique shape match maxima were found. Not all optimizations lead to unique MSC maxima; duplicate maxima were often found. The theoretical maximum for the intersection volume was 287.1 Å^3, corresponding to the volume of the smallest molecule, **2**.

The highest intersection volumes found for each conformational pair are presented as a histogram in Figure 4 and also listed in Table II of the supplemental material. Eight conformational pairings are found with intersection volumes between 251.0 and 253.8 Å^3, the highest values obtained. For these eight pairs, the biphenylyl tetrazole moieties are nearly perfectly aligned. In addition, the quinoline and imidazole rings are also brought into coincidence (Figure 5a,b). All eight pairs involve conformation C3 of **1**. Similar alignments are obtained for the four biphenyl orientations (B1–B4). Thus, conformation C3B1 matches both H1B1 and H2B1, C3B3 matches H1B3 and H2B3, etc. As can be seen from Figure 2, conformation H1 and H2 of **2** are very similar in overall shape, differing only in the transposition of the NH and CH groups. The overall shape comparison is not sensitive to this interchange of NH for CH, and thus, the shape comparison of H1 or H2 with C3 yields similar results.

A second group of eight pairings with intersection volumes between 239.5 and 245.7 Å^3 is also obtained. These involve conformation C2 of **1**. A pattern similar to that observed for the eight best matches is repeated. The alignments generated by the shape match of C2B1 with H1B1 and H2B1 are shown in Figure 5c,d.

The remaining conformational pairings typically gave intersection volumes of ca. 190–215 Å^3. Several examples of these less satisfying shape matches are shown in Figure 6, for comparison.

It is interesting to compare these results with the previously derived model[2] (Figure 7). A good correspondence of the tetrazole groups and alignment of the quinoline and imidazole nitrogen atoms and of the 2-alkyl groups of the two molecules was sought since the structure–activity relationships of both series indicate that these groups are important to activity.[15] In the molecular graphics studies, four conformational pairings (C3B1–H1B1, C3B2–H1B2, C3B3–H1B3, and C3B4–H1B4) were identified as providing this correspondence. These same four conformational pairs are among the eight best matches found in the MSC of **1** and **2**. While the alignments from MSC are reasonably close to those in Figure 7, the matching of the bulky biphenylyl tetrazole moieties did not allow alignment of the quinoline and imidazole nitrogen atoms and of the 2-methyl groups. In addition, MSC identifies high scoring conformational pairings between C3 and H2. An alternative pairing using conformation C2 of **1** is also suggested by MSC (Figure 5c,d). The unbiased MSC overlays emphasize the alignment of the biphenylyl tetrazole moiety at the expense of overlaying the NH groups. Therefore, MSC does not discriminate between the match with H1 and the match with H2.

Our MSC has been adapted to allow for discrimination between various functional groups or atom classes, as described above, to facilitate matching based on known structure–activity relationships or chemical intuition. The atoms in **1** and **2** were divided into four classes; the four nitrogens of the tetrazole rings were assigned to the "red" class, the NH of the imidazole or quinoline were assigned to the "blue" class, the tetrazole rings were taken as unprotonated, and the remaining atoms were placed in the "yellow" class. MSC of the 192 possible conformational pairs using these classes gave results very similar to those outlined above. The MSC was again dominated by the biphenylyl tetrazole moiety. The red and blue classes in both **1** and **2** have volumes of 39.8 and 16.8 Å^3, respectively, and this does not vary significantly among conformations. The yellow class of **2** is smaller than that of **1** and has a theoretical maximum of 248.7 Å^3, which also does not vary with conformation. To emphasize the importance of matching the red and blue classes, a weighting scheme was adopted to place equal emphasis on each class. The weightings were chosen (red = 2.554, blue = 6.059, yellow = 0.409) to give similar theoretical "volume" maxima for each class (red = 2.554 × 39.8 = 101.6, blue = 6.059 × 16.8 = 101.8, and yellow = 0.409 × 248.7 = 101.7) and therefore a theoretical maximum for the "class weighted" MSC of 305.2 Å^3.

An MSC was carried out using the classes and weightings described above, for each of the 192 unique conformational pairings. Each pair was allowed to run for 30 CPU min, requiring a total of 96 CPU h for the entire set. For each pair, 9–18 shape comparison optimizations were completed. Again many duplicate maxima were found.

The highest MSC scores for each conformational pair are presented as a histogram in Figure 8 and are listed in Table III of the supplementary material. Nine high-scoring pairs (252.0–263.1 Å^3) were identified, with a gap to the next (10th) pair (240.2 Å^3). The tetrazole, NH, and biphenyl are all well-aligned for these pairs. These nine

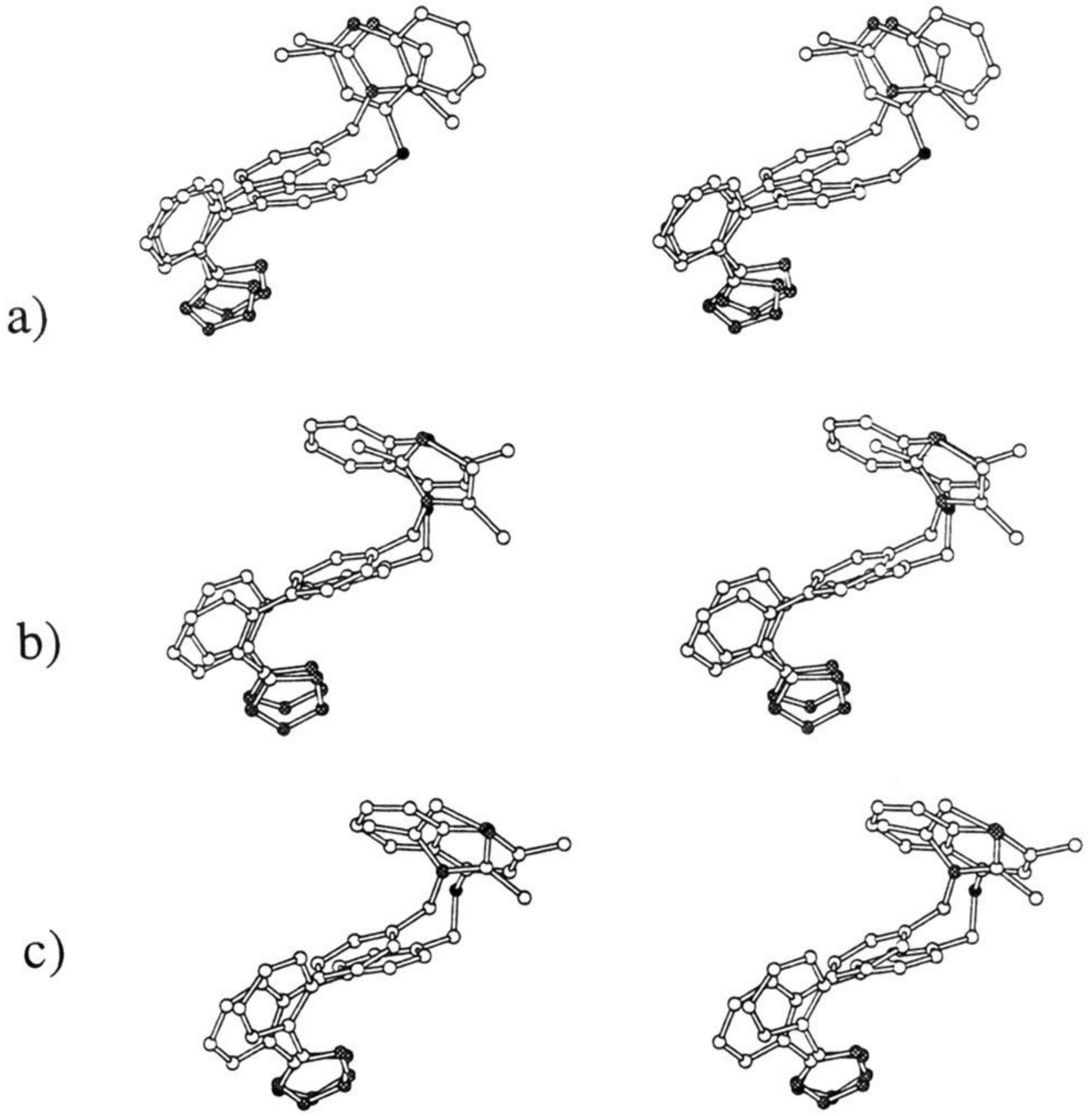

Figure 9. Three examplex of shape match alignments for high scoring conformational pairs using the classes and weighting defined in the text: (a) conformations C3B1 and H1B1 (262.8 Å^3), (b) conformations C2B1 and H1B1 (255.6 Å^3), (c) conformations C2B1 and H2B1 (257.9 Å^3).

pairs could be divided into three groups. Figure 9a shows the alignment of conformations C3B1 and H1B1 from the first group of four conformational pairs. Similar alignments are seen for the other biphenyl orientations (B2–B4). Thus, the first group paired C3B1 with H1B1, C3B2 with H1B2, etc. As can be seen, this alignment is nearly identical to the previously proposed model (Figure 7). The matches between C3 and H2 did not score as well in this case, presumably because it is not possible to overlay all three classes, NH, tetrazole, and biphenyl simultaneously with these conformational pairs. The second group pairs conformations C2B1 with H1B1 (Figure 9b) and C2B4 with H1B4. Interestingly, biphenyl orientations B2 and B3 did not score as well in this group. Graphical overlay confirmed that these two pairs do not match as well. Finally, the third group pairs conformations C2B1 with H2B1 (Figure 9c), C2B3 with H2B3, and C2B4 with H2B4. Thus, MSC has identified several alternative pharmacophore models.

V. Conclusion

We have developed a powerful new method for the analytical comparison of molecular shapes and the alignment of molecules based on this shape similarity. Our method allows discrimination between groups with different chemical properties and allows different emphasis of these groups in the shape similarity. Application to conformationally flexible angiotensin receptor antagonists identifies both a previously deduced pharmacophore model and also several attractive alternate models. Our MSC method provides a powerful tool for the systematic comparison of conformationally flexible molecules.

Acknowledgment. We thank Dr. Robert Hout for useful discussions on conformation searching and for pointing out the utility of quaternions. We also thank Dr. Robert Bradbury for useful comments on the manuscript.

Supplementary Material Available: A two-page elaboration of analytical derivative calculations for Connolly's molecular volume functions. Table I gives the atomic radii used and Tables II and III give the shape scores for the 192 conformational pairs which detail the data given in Figures 4 and 8 (5 pages). Ordering information is given on any current masthead page.

References

(1) See, for examples: (a) *Concepts and Applications of Molecular Similarity*; Johnson, M. A., Maggiora, G. M., Eds.; Wiley: New York, 1990. (b) Cardozo, M. G.; Kawai, T.; Iimura, Y.; Sugimoto, H.; Yamanishi, Y.; Hopfinger, A. J. Conformational Analyses and Molecular-Shape Comparison of a Series of Indanone-Benzylpiperidine Inhibitors of Acetylcholinesterase. *J. Med. Chem.* **1992**, *35*, 590–601. (c) Hermann, R. B.; Herron, D. K. OVID and SUPER: Two overlap programs for drug design. *J. Comput.-Aided Mol. Des.* **1991**, *5*, 511–524. (d) Meyer, A. Y.; Richards, W. G. Similarity of molecular shape. *J. Comput.-Aided Mol. Des.* **1991**, *5*, 427–439. (e) Papadopoulos, M. C.; Dean, P. M. Molecular structure matching by simulated annealing. *J. Comput.-Aided Mol. Des.* **1991**, *5*, 119–133. (f) Coloc'h, N.; Mornon, J. A new tool for the qualitative and quantitative analysis of protein surfaces using B-spline and density of surface neighborhood. *J. Mol. Graphics* **1990**, *8*, 133–140. (g) Burt, C.; Richards, W. G.; Huxley, P. The Application of Molecular Similarity Calculations. *J. Comput. Chem.* **1990**, *11*, 1139–1146. (h) Blandon, P. A rapid method for comparing and matching the spherical parameter