Expert System Assisted Pharmacophore Identification

Attilla Ting, Ross McGuire,† A. Peter Johnson,* and Stuart Green

ICAMS, School of Chemistry, University of Leeds, Leeds, LS2 9JT, U.K.

Received August 15, 1999

An expert system for automatic perception of pharmacophoric groups is presented. Important features include consideration of the protonation state at physiological pH and detection of potential tautomerism. This perception information is used in the generation of pharmacophores using clique detection.

1. INTRODUCTION

For many interesting receptor families detailed three-dimensional (3-D) structural information is not currently available. In these cases, indirect approaches to drug design such as the pharmacophore method are particularly useful.[1] With some notable exceptions e.g., ALADDIN,[2] many of the existing methods for pharmacophore generation depend on the manual identification of the atoms to be superimposed and sometimes neglect important chemical issues such as tautomerism and protonation state. Moreover, as combinatorial techniques are now widely used in drug discovery,[3] automatic perception of pharmacophoric groups is absolutely essential because of the large numbers of compounds involved.

An expert system has been developed and used for automatic perception of pharmacophoric groups. This system includes the identification of potential tautomerism and consequent assignment of the protonation state and identification of pharmacophoric groups. The overall approach has three main stages: (i) the generation of tautomers, if present, (ii) automatic perception of protonation state, and (iii) identification of pharmacophoric groups. Conformational flexibility is handled by the generation of a small library of low energy conformers of active compounds using conformational search. Possible pharmacophores are then generated using a clique detection algorithm.

2. AN EXPERT SYSTEM FOR CHEMICAL PERCEPTION

To allow intelligent, accurate, and flexible identification of pharmacophoric groups, an expert system based approach was chosen both for perception of the pharmacophoric groups and for the generation of tautomers. An expert system generally consists of a reasoning engine and one or more knowledge bases. The reasoning engine which was developed is based on that used in the CAESA (computer-assisted estimation of synthetic accessibility) system, which estimates ease of synthesis using a systematic knowledge based approach.[4,5]

The knowledge bases are text files containing a number of rules, each consisting of one of more antecedents (substructures described in PATRAN[6] strings) followed by a number of consequences. PATRAN, which is used to specify required substructures, is similar to the SMILES language,[7] with additional properties that increase the flexibility when specifying atoms and bonds. [Details of the PATRAN used in the knowledge bases are available upon request.] Consequences instruct the reasoning engine which actions to perform if the specified antecedent(s) is (are) present. Although one reasoning engine is used throughout, several knowledge bases have been developed for different parts of the system. They are named tautomerism, protonation, and hydrogen bonding knowledge bases. [Please refer to Supporting Information for details of the knowledge bases.] An overview of the expert system presented is summarized in Figure 1.

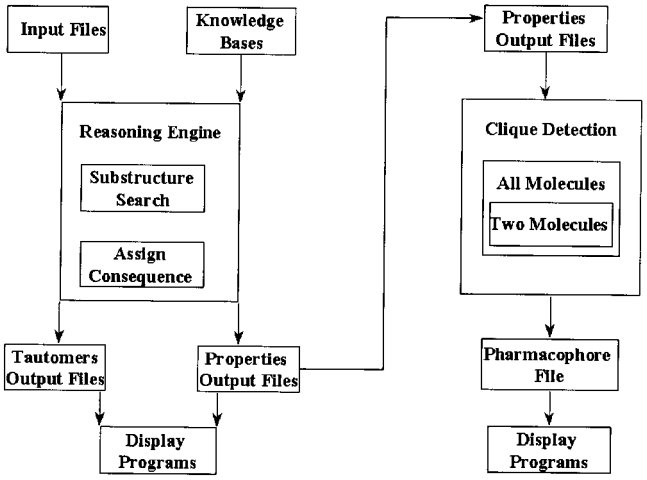

Figure 1. Overview of the programs presented.

3. CHEMICAL PERCEPTION

3.1. Tautomerism. Many molecules can potentially exist in several tautomeric forms such as keto/enol and enamino/imino. One result of such tautomerism is that hydrogen bonding properties of functional groups such as enamino, imino, enol, and ketone can change dramatically; for instance a given atom can be a hydrogen donor in one form and an acceptor in another.[8] Tautomerism is an important issue, but at present most pharmacophore identification programs do not take it into consideration, perhaps because of the complicated nature of the problem.

3.1.1. Overall Method. The overall approach used to tackle tautomerism is outlined below. First a filter is used

* To whom correspondence should be addressed.
† Organon Labs Ltd., Newhouse ML1 5SH, Scotland, U.K.

10.1021/ci990106h CCC: $19.00
Published on Web 02/05/2000

Table 1. Imidazole Rule in Tautomerism Knowledge Base[a]

```
PATRAN string :

CHEMICAL-LABEL <Imidazole ring>
...STARTP
...N[HS=0];[ARYL=YES]=C[ARYL=YES]
…-N[HS>0];[ARYL=YES]-C[ARYL=YES]
…=C[ARYL=YES]-@1
...ENDP

RULE: IF antecedent THEN consequences

RULE
EXPLANATION: prototropic tautomerism
with proton jumping from N3 to N1.
IF Imidazole ring
THEN
        substitute-bond 1 with -
        substitute-bond 2 with =
        add-taut-hydrogen 1
END-THEN
```

[a] Refer to Figure 2 for structures.

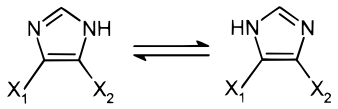

Figure 2. Prototropic tautomerism of imidazole. Tautomerism rule shown in Table 1.

to identify those atoms in a molecule which the knowledge base suggests may participate in tautomerism. This is followed by generation of possible tautomeric forms which are stored as new structures. Pharmacophoric perception is then carried out and the resulting information stored in a list of tautomer structures.

A knowledge base has been developed in order to identify molecules that may undergo tautomerism. Common tautomeric cases that are covered range from keto/enol tautomeric forms to tautomerism in five- and six-membered heteroaromatic rings. A typical rule in the tautomerism knowledge base is shown in Table 1, with an example of its operation shown in Figure 2. In general, the parent molecule is transformed into one or more tautomers using the rules in the knowledge base. For example, molecule 1 in Figure 3 can undergo tautomerism to give rise to molecule 2 by the imidazole ring tautomeric rule (Table 1). Molecule 1 can also transform into molecule 3 through the pyrimidone type 1 tautomeric rule and to molecule 4 by the amide to imino alcohol in aromatic system tautomeric rule on the pyrimidone ring. Each of these newly generated tautomers from the parent molecule is then checked for the possibility of further tautomerization, with the process applied iteratively until no more unique structures are generated. In Figure 3, molecule 2 can also be transformed into molecules 5 and 6 respectively by applying the same tautomeric rules used by molecule 1. Moreover, molecule 3 can form tautomers 4 and 5, while both molecules 4 and 5 can also give rise to molecule 6.

3.1.2. Tautomer Relationships. The relationship between tautomers and the parent molecule can be quite complex due to the recursive nature of the structure generation. Tautomers are classified into one of four main groups: forward

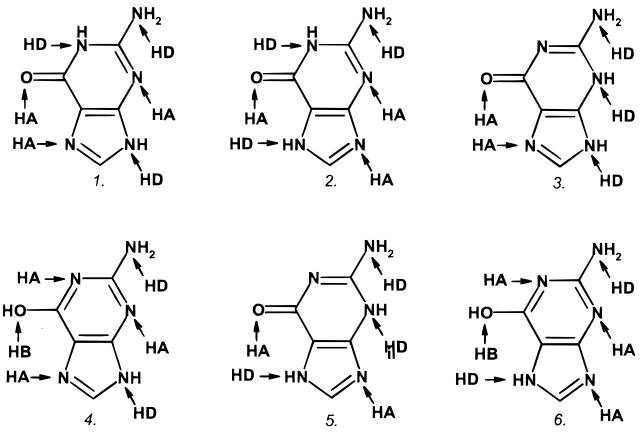

Figure 3. Perceived hydrogen bonding properties for guanine and its generated tautomers. HD, hydrogen donor *(N.B. the hydrogen bearing atom is labeled)*; HA, hydrogen acceptor; HB, both hydrogen donor and acceptor.

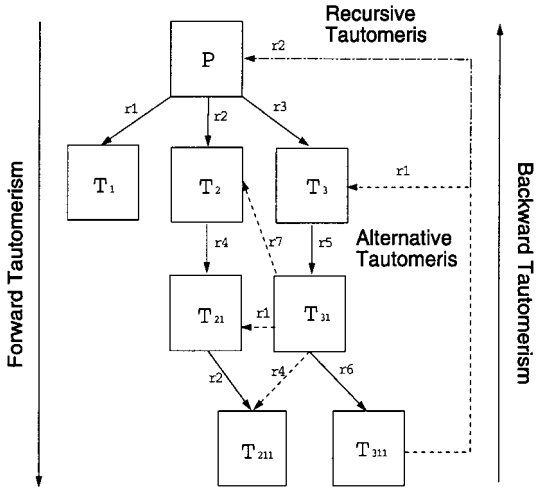

Figure 4. Tautomer relationships.

generated tautomers, backward generated tautomers, recursive tautomers, and alternative tautomers.

As shown in Figure 4, new structures generated from the parent molecule or an appropriate ancestor are called forward generated tautomers. For example, in Figure 4 parent molecule P generates T_1. The reverse of this process gives backward generated tautomers, e.g., regeneration of the parent from a child node. However, there is the case where a child may be able to regenerate a molecule in the previous generation that is not its direct parent or vice versa. This is called alternative tautomerism (T_{311} to T_3), and is the case for any molecule than has more than one incoming route (T_3). Therefore, alternative tautomerism can be forward, backward, or go from right to left in the same generation. As an example, T_{311} can regenerate molecule P. However, due to the fact that P only has one incoming edge, it is not classified as an alternative tautomerism but rather as a recursive tautomerism. Recursive tautomerism always goes back to the starting molecule, forming a complete loop. Therefore, a recursive tautomer can be an alternative tautomer, but the reverse is not necessarily the case.

3.1.3. Implementation and Testing. In order for the program to identify these relationships, trees are generated, and to avoid being trapped in an endless loop, the branches of the tree are pruned as soon as they generate repeated

Table 2. Some pK_a Values of Heterocyclic Compounds from the Literature, Which Were Used To Construct the Protonation Knowledge Base

	pK_a		reference	
name	nonsubstituted	substituted	nonsubstituted	substituted
imidazole	7.00	−0.81 to 8.65	p 21 ref 13	p 20 ref 11
pyrazole	2.52	−3.79 to 3.51	p 321 ref 10	p 20 ref 11
oxazole	0.8 ± 0.2[a]	−1.21 to 3.56	p 21 ref 13	p 24 ref 11
isoxazole	0.23	−0.76 to −3.22	p 321 ref 10	p 24 ref 11
thiazole	2.50	1.20 to 9.53	p 21 ref 13	p 24 ref 11
isothiazole	<0		p 29 ref 13	
1,2,3-triazole		−0.45 to 8.85		p 21 ref 11
1,2,4-triazole	2.27,[b] 10.26[a]	2.05 to 11.08		p 21 ref 11
thiadiazole	−0.90 to −4.90		p 456 ref 15	
tetrazole	4.89	1.70 to 1.73	p 19 ref 12	p 21 ref 11
pyridine	5.16	−4.00 to 9.57	p 9 ref 11	pp 9−10 ref 11
α-pyridone	0.8		p 62 ref 10	
γ-pyridone	3.3d		p 62 ref 10	
piperidine	11.29	4.5 to 11.22	p 17 ref 11	p 17 ref 11
2-pyrone	<−0.3		p 154 ref 10	
4-pyrone	−0.3		p 154 ref 10	
pyrimidine	1.31, −6.3[c]	−4.4 to 13.68	p 12 ref 11	pp 12−15 ref 11
pyridazine	2.44	1.61 to 3.70	p 73 ref 13	p 70 ref 12
pyrazine	0.51	0.55 to 3.55	p 73 ref 13	p 70 ref 12
piperazine	9.82		p 63 ref 12	
morpholine	8.70	7.70 to 8.70	p 63 ref 12	p 63 ref 12
1,4-dioxan	−2.92		p 63 ref 12	
1,3,5-triazine		3.91 to 8.15		p 17 ref 11
1,2,4-triazine		6 to 12.19		p 17 ref 11

[a]Basic ionization. [b]Acidic ionization. [c]The weaker of two basic groups.[10−13,15]

structures. Each tautomer generated has a record of its intermediate parent for fast comparison and to ensure that forward and backward generated relationships for the tautomer only match once. The reasoning engine generates a unique identifier (canonical number) for each tautomer. This number is compared with a list of previously calculated numbers to prevent generation of repeated structures. The canonicalization technique used was originally developed in CAESA and is based on an extended Morgan algorithm.[9]

A set of nucleic acid bases were chosen as test data due to their "druglike" structures, and these provide a demonstration of complex tautomerism. Figure 3 shows the consequences of tautomerism on the assignment of hydrogen bonding properties in this case.

3.2. Protonation State. It is known that electrostatics plays an important role in protein−ligand binding and so should be taken into consideration in order to provide good estimation of protein−ligand interactions. Therefore, this study included perception of the protonation behavior of ligands at physiological pH. The measured pK_a values of different functional groups in many different environments were extracted from the literature in order to construct the protonation knowledge base[8,10−17] (Table 2). Using this information, the knowledge base provides an estimate of the protonation states of various common acidic and basic functional groups and heteroaromatic rings at physiological pH (pH7). Functional groups contained in the knowledge base include aliphatic and aromatic amines, carboxylic acids, nitro groups, sulfoxides, guanidines, amidines, and heterocyclic compounds, mainly five- and six-membered rings. The expected protonation states of key functional groups are summarized in Table 3a and key heterocyclic systems in Table 3b. A typical rule in the protonation knowledge base is shown in Table 4, with corresponding structures shown in Figure 5. For example, pyridine has a pK_a of 5.16 and is not protonated to a significant extend at pH7 but 4-amino-pyridine has a pK_a value of 9.29[11] (Figure 5) and

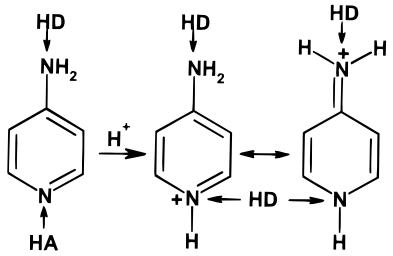

Figure 5. Protonation of 4-aminopyridine. Perceived hydrogen bonding properties are labeled as in Figure 3. Protonation rule see Table 4. "+" only represents the sign of the charge, not the magnitude.

is protonated at physiological pH. For these substituted heterocyclic systems the more specific rules override the more general rules, which are applied when the particular substituents are not identified. For example, most imidazole type molecules have pK_a above 7 and are protonated in physiological pH, except for chloro or nitro substituted imidazoles. Therefore, the knowledge base uses a general imidazole rule to set all imidazole molecules to their protonated forms and then more specific chloroimidazole and nitroimidazole rules are used to set chloroimidazoles and nitroimidazoles to deprotonated forms.

3.3. Hydrogen Bonding Properties. The hydrogen bonding properties knowledge base was constructed in order to assign hydrogen donor/acceptor properties to functional groups. The entries in the knowledge base contain rules that cover functional groups ranging from simple amides to sulfonamides and guanidines.[8] The knowledge base is constructed in such a way that the more generic rules match first followed by increasingly specific ones. A typical rule in the hydrogen bonding properties knowledge base is shown in Table 5. Single heteroatom functional groups covered include alcohols, ethers, aldehydes, ketones, amines, amides, nitriles, thiols, and sulfides. More complex functional groups with multiple heteroatoms, including carboxylic acid, ester,

Table 3. Protonation States of Functional Groups Used To Construct the Protonation Knowledge Base

(a) Simple Functional Groups

name	protonation	structures
aliphatic amine	√	RNH_3^+
aromatic amine	×	$ArNH_2$
amide	×	$RCONH_2$
carboxylic acid	×	CO_2^-
nitro	×	NO_2^-
sulfoxide	×	S^+O^-
guanidine	√	$CN^+H_2(NH_2)_2$
amidine	√	$RCN^+H_2NH_2$

(b) Heterocyclic Compounds[a]

size	hets	N	O	S	protonation	name
5	2	2			√	imidazole
5	2	2			×	pyrazole
5	2	1	1		×	oxazole
5	2	1	1		×	isoxazole
5	2	1		1	√ (s)	thiazole
5	2	1		1	×	isothiazole
5	3	3			√ (s)	triazole
5	3	2	1		?	oxadiazole
5	3	2		1	×	thiadiazole
5	4	4			×	tetrazole
5	4	3	1		?	oxatriazole
5	4	3		1	?	thiatriazole
6	1	1			√ (s)	pyridine
6	1	1			×	pyridone
6	1	1			√	piperidine
6	1		1		×	pyrylium
6	1		1		×	pyran
6	1		1		×	pyrone
6	1			1	×	thiopyrylium
6	2	2			√ (s)	pyrimidine
6	2	2			×	pyridazine
6	2	2			×	pyrazine
6	2	2			?	pyrimidone
6	2	2			√	piperazine
6	2	1	1		?	oxazine
6	2	1	1		√	morpholine
6	2	1		1	?	thiazine
6	2		2		×	1,4-dioxan
6	3	3			√ (s)	triazine

[a] Size, the size of the rings of the heterocyclic compounds; hets, the number of heteroatoms present in the ring, which separated into number of N, O, and S; protonation, compound protonation state at pH 7. √, protonated; √ (s), only substituted form protonated; ×, not protonated; ?, no information; name, chemical name.

nitro, guanidine, amidine, sulfoxide, sulfones, sulfonamide, and phosphoramide are also covered by the knowledge base. This rule based approach allows the extended molecular environment of the group to be taken into consideration, and therefore enables a more accurate modeling of the ligands than would be the case with an isolated atom based approach. For example, the sp^3 nitrogen in piperidine is not a hydrogen acceptor because it is protonated but the sp^2 nitrogen in pyridine is. Perception of hydrogen bonding properties follows perception of tautomerism and protonation. It is necessary to perform perception in this order since the hydrogen bonding properties of atoms are strongly affected by tautomeric and protonation states (Figures 3, 5 and 6).

4. CONFORMATIONAL SEARCH

All of the conformational searches in this study were carried out using MacroModel Version 5.5.[18] The molecules under consideration were drawn in manually in the input mode followed by minimization using Merck molecular force field (MMFF)[19] with the Polak−Ribiere conjugate gradient (PRCG) minimization algorithm. The molecules used were subjected to a short conformation search using default parameters. The lowest energy conformation found was subjected to the full Monte Carlo conformational search,[20] using the Monte Carlo multiple minimum (MCMM) search method with MMFF force field and PRCG minimization algorithm. Further minimization cycles were carried out for some flexible molecules which initially gave poorly converging results. For each active compound all of the conformations within a specified energy window were then considered in the pharmacophore search strategy.

Table 4. 4-Aminopyridine Rule in Protonation Knowledge Base[a]

```
PATRAN string :

CHEMICAL-LABEL <4-Amino-Pyridine>
...STARTP
...N-,%C=,%C-,%C(-N[EPS=1])=,%C-,%C
…=,%@1
...ENDP

RULE
EXPLANATION: 4-Amino-Pyridine
protonated at ring Nitrogen.
IF 4-Amino-Pyridine
THEN
        set-charge 1 to +1
        add-prot-hydrogen 1
END-THEN
```

[a] Refer to Figure 5 for structures.

Table 5. Thiol Rule in Hydrogen Bonding Properties Knowledge Base[a]

```
PATRAN string :

CHEMICAL-LABEL <Sulphur with H & EPS>
...STARTP
...C&S[HS>0];[EPS>0]
...ENDP

RULE
EXPLANATION Sulphur with hydrogen and
electron pairs (EPS) is both HAcceptor
and HDonor.
IF Sulphur with H & EPS
THEN
        assign-atom 2 as HBOTH
END-THEN
```

[a] For example, the ACE inhibitors (Figure 6) used in this study all have thiol groups.

5. PHARMACOPHORE MAPPING

Following chemical perception, pharmacophores are generated by the application of a clique detection pharmacophore mapping algorithm on the library of conformers in a manner similar to that used in the DISCO program. Several applications of DISCO in pharmacophore mapping have been published, including the dopaminergic, benzodiazepine agonists[21] and platelet-activating factor antagonists.[22] Willett et al. have also applied clique detection in several areas[23−25] including 3-D substructure search.[24]

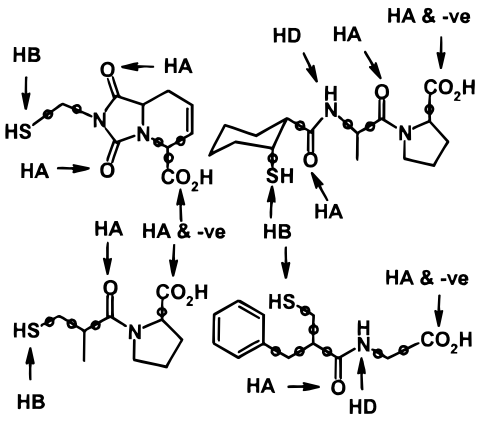

Figure 6. The four ACE inhibitors used in our test set labeled with rotatable bonds and pharmacophoric properties. The molecules are drawn as they are used in the conformational search. The perception results shown have taken into account the protonation state of the molecules. Rotatable bonds used in conformational search are indicated by a circle across the bond. Perceived hydrogen bonding properties perceived are labeled as in Figure 3. −ve indicates a negatively charged group.

The clique detection algorithm finds maximal common subgraphs (cliques) in an input graph that matches cliques in a template (reference) graph. In another words, it finds the common points shared by the two 3-D structures. These points are matched if they have the same pharmacophoric types and share the same distances within a user-defined tolerance.

The clique detection algorithm has the advantages of being fast and less computationally demanding than, for example, simulated annealing. It also has some very powerful features compared with other pharmacophore mapping methods. In most pharmacophore mapping algorithms, there is a requirement for both prior knowledge of the pharmacophoric pattern and the alignment of the molecules in some manner, but these are not necessary for the clique detection algorithm. Brint and Willett[26] compared several different clique detection algorithms and found that the Bron−Kerbosch algorithm[27] was the fastest for automatic pharmacophore detection.

The clique detection method used in this approach is capable of handling many molecules, each with multiple conformers. It is based on the Bron−Kerbosch algorithm,[27] with a branch and bound check used to improve performance by detecting and removing those branches on the search tree that do not lead to a successful solution. In contrast to many other pharmacophore mapping procedures which need predefined pharmacophoric points before matching, clique detection not only identifies those conformations that fit the pharmacophore but also generates the pharmacophore itself.

5.1. Tolerance Handling. A user-defined tolerance is used in two different aspects of the clique detection process: first, in identifying cliques between two molecules to decide whether the distances between pharmacophoric groups in molecules match, which is relatively straightforward; second, when classifying cliques, the tolerance influences whether two cliques are classified as identical, which is more complex.

At present, cliques are identified by comparing molecules with a template and classifying them using the pharmacophoric distances in the template. We currently use a tight checking method, meaning that two cliques are classified as

Table 6. Running Time for the Conformational Search Experiments[a]

template	speed (h)	R bond	no. atoms
ACE1	3.13	5	26
ACE2	17.77	5	46
ACE3	3.05	4	29
ACE4	15.17	7	32

[a] Experiments were conducted using MacroModel v5.5 on an SGI Indy R5000.

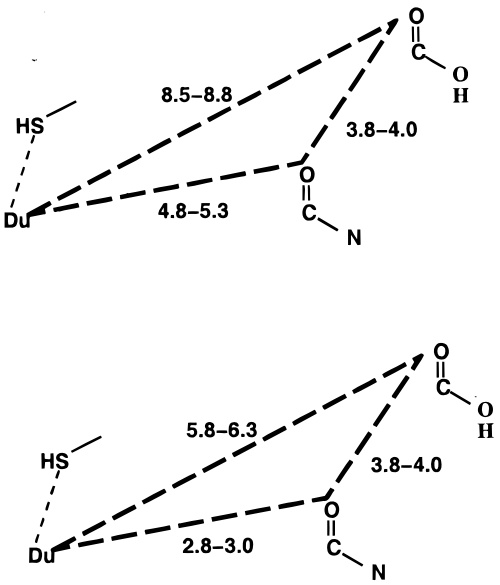

Figure 7. The two pharmacophores for the ACE inhibitors proposed by Dammkoehler et al. using constrained systematic search.[28,29] Du, the approximate location of the enzyme Zn atom. All distances are in Å.

Table 7. Running Time for the Pharmacophore Mapping Experiments[a]

template	speed (s)	no conf	phar pts
ACE1	18.3	6	5
ACE2	53.4	14	6
ACE3	1.6	2	4
ACE4	38.5	15	5

[a] Experiments were conducted using an SGI Indy R5000.

the same only if their template distances are identical. As a result, all members of a clique will be within the user-defined tolerance of the template. The only problem with this approach is that sometimes two cliques that are classified as different may have similar distances and actually belong to a single larger clique. Therefore, a clustering algorithm is used to reclassify all the cliques following the initial classification; thus, small cliques are grouped into larger cliques.

5.2. Identical and Chiral Clique Classification. As mentioned previously, when a new clique is generated it is necessary to determine whether it is identical to or a mirror image of any existing cliques. A hybrid distance and transformation method has been used to allow efficient and effective classification.

A distance based approach was initially used, but though fast, this was found to have limitations when dealing with cliques with identical distances, thus being unable to differentiate chiral cliques. The transformation approach has the

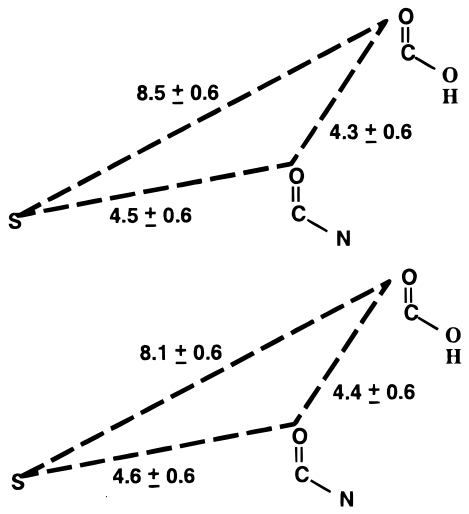

Figure 8. The two pharmacophores generated by our program which contain the pharmacophoric groups shown in Figure 7. ACE inhibitor 4 was used as the template molecule. User-defined tolerance was 0.6 Å. Distances are measured from the sulfur atom rather than the dummy metal atom (the enzyme Zn atom) used earlier.[28] All distances are in Å.

advantage of avoiding these problems and also has the ability to deal with stereo information. However, it is computationally expensive if it has to deal with many molecules with multiple conformations.

Therefore, in the combined approach, the fast distance based method is used for initial matching, with the transformation algorithm applied to potentially identical cliques as a chirality check. The distance based algorithm also has the advantage of suggesting a good starting triplet for matching using the transformation approach, thus increasing its efficiency.

6. RESULTS

To test and evaluate the pharmacophore mapping program properly, it was necessary to prepare some test data. Several groups of candidates were chosen from the literature. The set presented here is from a study by Dammkoehler et al.[28,29] on ACE inhibitors using constrained systematic search. All our experiments were performed on a SGI Indy R5000.

6.1. ACE Inhibitors. Four ACE inhibitors were used in our test set (Figure 6). A library of low energy conformers was found by performing conformational searches using MacroModel v5.5.[18] A 10kJ/mol energy window and a user-defined tolerance of 0.6 Å were used in this experiment, but both of these parameters can be varied by the user at run time. The running time of conformational searches depends on the flexibility of the molecules and method used (Table 6). The perception of pharmacophore groups is fast, e.g., 1.1 s for the ACE inhibitors (37 conformations). The molecules with pharmacophoric groups identified by the perception program (Figure 6) were then submitted for the pharmacophore mapping program. The running time of the pharmacophore mapping program on the ACE inhibitors test set is given in Table 7. Each of the ACE inhibitor molecules was in turn used as the template molecule in the clique detection process.

The pharmacophores proposed by Dammkoehler et al.[28] are shown in Figure 7. The predefined pharmacophoric groups used in their study are the terminal carboxyl group, an amido carbonyl group, and a zinc binding group (e.g., thiol group).

Our system identifies several pharmacophores which reproduce the literature superposition and some distances of the pharmacophoric groups. Figure 8 shows two of the pharmacophores generated by the program in this study which match the pharmacophoric groups found in the Dammkoehler study (Figure 7). The other pharmacophores found are summarized in Figure 9, and their pharmacophoric distances are shown in Table 8. In our approach, the positions of the pharmacophoric groups considered are the positions of the heavy atoms and not the positions of their hydrogens or lone pairs, which can have large degrees of spatial freedom. Therefore, the pharmacophores generated are presented in terms of interfeature distances. For example, the pharmacophoric distances (Figure 8) are measured with respect to the position of the sulfur atom in the thiol group (zinc binding group). In contrast to this, Dammkoehler et al. measured distances to an estimated location of the enzyme Zn atom (Figure 7).

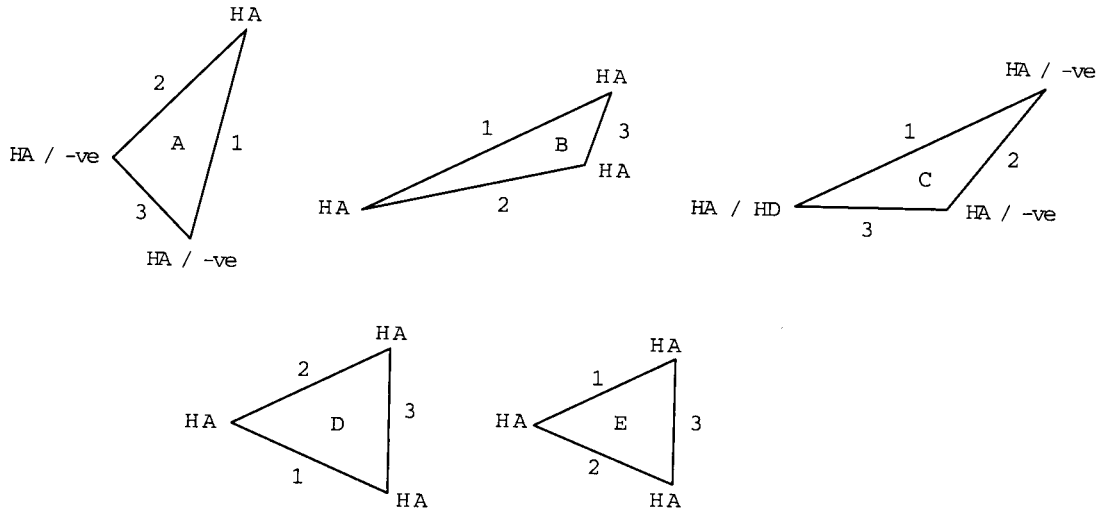

Figure 9. Summary of all pharmacophores generated with pharmacophoric groups labeled 1, 2, and 3 represent distances 1, 2, and 3, respectively. Refer to Table 8 for the pharmacophoric distances.

Table 8. The Pharmacophoric Distances of All Pharmacophores Generated Using the ACE Inhibitors in This Study[a]

template	phar	dist 1	dist 2	dist 3
ACE1	A	4.65−4.67	3.63−3.66	2.22−2.23
ACE1	B	7.50−7.96	5.53,6.67	2.22−2.23
ACE1	C	7.5−8.13	4.65−4.67	3.63−3.98
ACE1	D	5.98	5.53	3.66
ACE2	A	4.06−4.10	3.15−3.22	2.22−2.23
ACE2	B	7.49−7.99	5.50−6.12	2.22
ACE2	C	8.46	4.60	4.06
ACE2	D	5.73	4.98, 3.73	3.66, 3.09
ACE2	E	4.27−4.53	3.42−3.66	3.15−3.42
ACE3	A	4.06−4.07	3.14−3.17	2.23
ACE3	B	7.99	6.10	2.23
ACE3	C	6.04, 7.99	4.06−4.07	3.59, 4.59
ACE4	A	4.30−4.43	3.43−3.68	2.22−2.23
ACE4	B	7.50−7.56	6.38−6.47	2.23
ACE4	C	8.09−8.52	4.55−4.57	4.05−4.43
ACE4	D	6.21−6.47	4.42−4.82	3.41−3.70
ACE4	E	4.13, 5.60	3.48−3.66	3.20−3.46

[a] Each of the ACE inhibitor molecules was in turn used as the template molecule in the experiments. Pharmacophores produced were clustered into pharmacophore classes A−E with distances shown in columns dist 1−3 (all distances are in Å. Ranges are indicated by "−" and "," for discontinuous values). Refer to Figure 9 for pharmacophore properties.

7. CONCLUSION

Our system is capable of generating pharmacophores for structurally diverse and flexible molecules. The approach used consists of three main stages. The first stage is the perception of chemically important features and identification of potential pharmacophoric groups. Accurate identification of such features is essential to the pharmacophore mapping procedure. Hence, this process is carried out using an expert system which allows intelligent, accurate, and flexible identification of pharmacophoric groups, including hydrogen donors, hydrogen acceptors, and positive and negative groups, and takes protonation behavior into account. Tautomers are also considered by the expert system prior to the identification of pharmacophoric groups.

This perception phase is followed by the preparation of a library of low energy conformers of ligands by conformation search using MacroModel. The pharmacophores are then generated using a clique detection algorithm, which has the advantage of requiring no prior knowledge and no prealignment of the pharmacophore groups. The generated pharmacophores, based on hydrogen bonding properties and charges, may then be used to align ligands in order to identify further important features, such as common hydrophobic regions, and provide useful information for pharmacophore validation.

ACKNOWLEDGMENT

We are grateful to our colleagues Jon. C. Baber and Miklós Vargyas for useful discussions. We would also like to thank Jasmit Kaur from Organon Labs Ltd. for her general advice. The research was sponsored by Organon Teknika.

Supporting Information Available: Details of the tautomerism, protonation, and hydrogen bonding knowledge bases. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Martin, Y. C. Pharmacophore mapping. In *Designing Bioactive Molecules: Techniques; Applications*; Martin, Y. C., Willett, P., Eds.; American Chemical Society: Washington, DC, 1997.

(2) VanDrie, J. H.; Weininger, D.; Martin, Y. C. ALADDIN: An Integrated Tool for Computer-Assisted Molecular Design and Pharmacophore Recognition From Geometric, Steric, and Substructure Searching of Three-dimensional Molecular Structures. *J. Comput.-Aided Mol. Des.* **1989**, *3*, 225−251.

(3) Dolle, R. E. Comprehensive Survey of Chemical Libraries Yielding Enzyme Inhibitors, Receptor Agonists; Antagonists, and Other Biologically Active Agents: 1992 through 1997. *Mol. Diversity* **1998**, *3*, 199−233.

(4) Myatt, G. J. Computer Aided Estimation of Synthetic Accessibility. Ph.D. Thesis, Leeds, 1994.

(5) Baber, J. C. CAESA−Improved Algorithms for the Identification of Starting Materials. Ph.D. Thesis, Leeds, 1998.

(6) Hopkinson, G. A. Computer-Assisted Organic Synthesis Design. Ph.D. Thesis, Leeds, 1985.

(7) Weininger, D. Smiles. 3. Depict Graphical Depiction of Chemical Structures. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 237−243.

(8) Jeffrey, G. A.; Saenger, W. *Hydrogen Bonding in Biological Structures*; Springer-Verlag: New York, 1994; Part II.

(9) Morgan, H. L. The Generation of a Unique Machine Description for Chemical Structures−A Technique Developed at Chemical Abstracts Service. *J. Chem. Doc.* **1965**, *5*, 107−113.

(10) Joule, J. A.; Mills, K.; Smith, G. F. *Heterocyclic Chemistry*; Chapman and Hall: London, 1994.

(11) *Physical Methods in Heterocyclic Chemistry*; Katritzky, A. R., Ed.; Academic Press: 1963; Vol. I.

(12) *Physical Methods in Heterocyclic Chemistry*; Katritzky, A. R., Ed.; Academic Press: New York, 1971; Vol. III.

(13) Davies, D. T. *Aromatic Heterocyclic Chemistry*; Oxford Science Publications: New York, 1992.

(14) *Comprehensive Heterocyclic Chemistry−Part 2B: Six-membered Rings with Oxygen, Sulfur or Two or More Nitrogen Atoms*; Katritzky, A. R., Ed.; Pergamon: Elmsford, NY, 1984.

(15) *Comprehensive Heterocyclic Chemistry−Part 4B: Five-membered Rings with Two or More Oxygen, Sulfur or Nitrogen Atoms*; Katritzky, A. R., Ed.; Pergamon: Elmsford, NY, 1984.

(16) *Handbook of Heterocyclic Chemistry*; Katritzky, A. R., Ed.; Pergamon: Elmsford, NY, 1985.

(17) *Merck Index*; Budavari, S., Ed.; Merck & Co. Inc.: Rahway, NJ, 1996.

(18) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. MacroModel−An Integrated Software System for Modelling Organic and Bioorganic Molecules using Molecular Mechanics. *J. Comput. Chem.* **1990**, *11*, 440−467.

(19) Halgren, T. A. Merck Molecular Force Field. I. Basis, Form, Scope, Parametrization, and Performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490−519.

(20) Chang, G.; Guida, W. C.; Still, W. C. An Internal Coordinate Monte Carlo Method for Searching Conformational Space. *J. Am. Chem. Soc.* **1989**, *111* (12), 4379−4386.

(21) Martin, Y. C.; Bures, M. G.; Danaher, E. A.; DeLazzer, J.; Lico, I.; Pavlik, P. A. A Fast New Approach to Pharmacophore Mapping and its Application to Dopaminergic and Benzodiazepine Agonists. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 83−102.

(22) Bures, M. G.; Danaher, E.; DeLazzer, J.; Martin, Y. C. New Molecular Modelling Tools Using 3-D Chemical Substructures. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 218−223.

(23) Gardiner, E. J.; Holliday, J. D.; Willett, P.; Wilton, D. J.; Artymiuk, P. J. Selection of Reagents for Combinatorial Synthesis using Clique Detection. *Quant. Struct.-Act. Relat.* **1998**, *17* (3), 232−236.

(24) Gardiner, E. J.; Artymiuk, P. J.; Willett, P. Clique-Detection Algorithms for Matching Three-dimensional Molecular Structures. *J. Mol. Graphics Modell.* **1997**, *15*, 4, 245−253.

(25) Holliday, J. D.; Willett, P. Using a Genetic Algorithm to Identify Common Structural Features in Sets of Ligands. *J. Mol. Graphics Modell.* **1997**, *15* (4), 221−232.

(26) Brint, A. T.; Willett, P.; Algorithms for the Identification of Three-dimensional Maximal Common Substructures. *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 152−158.

(27) Bron, C.; Kerbosch, J. Algorithm 457: Finding all Cliques of an Undirected Graph [H]. *Commun. ACM* **1973**, *16*, 575−577.

(28) Dammkoehler, R. A.; Karasek, S. F.; Shands, E. F. B.; Marshall, G. R. Constrained Search of Conformational Hyperspace. *J. Comput. Aided-Mol. Des.* **1989**, *3*, 3−21.

(29) Leach, A. R. *Molecular Modelling Principles and Applications;* Longman: Essex, U.K., 1996; Chapter 10, p 547.

CI990106H