Automatic Assignment of Chemical Connectivity to Organic Molecules in the Cambridge Structural Database

Jon C. Baber and Edward E. Hodgkin*

British Bio-technology Limited, Watlington Road, Oxford OX4 5LY, United Kingdom

A procedure for assigning chemical connectivity to organic structures in the Cambridge Structural Database is described. Coordination states and bond orders are estimated by measurement of bond angles and bond lengths. An iterative cycle, in which bond orders and coordination states are modified using a set of rules, is applied until a consistent molecular structure is attained. The procedure assigns chemical connectivity with a success rate comparable to a previously published approach for assigning atom hybridization states.

INTRODUCTION

Databases of three-dimensional molecular structures play an important role in drug design.[1-3] Sources of structural data include X-ray diffraction and 2D to 3D conversion programs such as COBRA[4] and CONCORD.[5] While the latter produce structures that may be readily transferred to modeling programs, structures from X-ray diffraction take the form of atom coordinates without chemical connectivity information. Chemical connectivity (atom types and bond orders) is required to some extent by all modeling and 3D database programs to enable structures to be interpreted as chemical entities. Determination of atomic hybridization states has been performed for small organic molecules in the Cambridge Structural Database[6] (CSD) by Meng and Lewis's IDATM program.[7] This paper describes the assignment of bond orders (in contrast to atom types) to organic molecules, as required by some 3D database programs, for example MACCS-3D.[8]

METHODS

Assignment of chemical connectivity to structures is achieved in a number of stages, collectively known as the OXBRIDGE suite of programs after the university towns of Oxford, where the work described here was carried out, and Cambridge. The C source code, written by Jon C. Baber, is available from the corresponding author by sending a Macintosh 3.5 in. disk.

Calculation of Atomic Connectivity. Table I lists the covalent radii[9] of elements found in the subset of organic molecules in the CSD. This is the same group of elements considered in the study of Allen et al.[10] of bond lengths in the CSD. Pairs of atoms, whose interatomic distance is less than the sum of the covalent radii plus a tolerance of 0.45 Å, are considered to be bonded. The tolerance of 0.40 Å used by Meng and Lewis failed to recognize a small number of bonded pairs (indeed, they reported such a failure in their study), and increasing the tolerance to 0.45 Å corrected the problem without recognizing any nonbonded pairs as bonded. Assignment of this "atomic connectivity" is performed for all molecules.

Identification of Discrete Molecules. Once atomic connectivity has been assigned, it is possible to identify discrete molecular structures in each of the crystal unit cells. This is achieved by a depthwise search through the atomic connectivity table until all atoms are accounted for and assigned to a

Table I. Covalent Radii Used in Initial Connectivity Distance Search[a]

atom	radius, Å	atom	radius, Å
As	1.21	N	0.68
B	0.83	O	0.68
Br	1.21	P	1.05
C	0.68	S	1.02
Cl	0.99	Se	1.22
F	0.64	Si	1.20
H	0.23	Te	1.47
I	1.40		

[a] Taken from ref 9. A pair of atoms is considered to be bonded if the distance between them, $D_{AB} \leq R_A + R_B + 0.45$, where D_{AB} is the interatomic distance and R_A and R_B are the covalent radii of atoms A and B.

Table II. Atom Pairs with One Observed Bond Order (X = Halogen)

atom pair	observed bond order	confidence
H–any	1	10000
X–any	1	10000
Si–any	1	$30 \times (l_{obs} - 1.4)$
B–any	1	$30 \times (l_{obs} - 1.2)$
P–S	2	$1.2 \times (0.08 - |l_{obs} - 1.93|)$
P–Te	2	$(0.07 - |l_{obs} - 2.09|)$
P–Se	2	$1.5 \times (0.10 - |l_{obs} - 2.09|)$

molecule. Metal atoms are ignored during this procedure so that "chemically independent" ligands bound to the same ion are considered separately. For the purpose of this study, metal atoms include all elements not listed in Table I. The following steps are applied to the discrete molecules rather than the unit cells.

Initial Assignment of Bond Orders and Confidence Values. Bond orders and associated confidence values are estimated for all pairs of bonded atoms by comparison of the bond length with ideal values derived from ref 10. There are three classes of bond according to the atom pair involved:

(i) Only one bond order is reported in the survey of Allen et al.[10] The relevant atom pairs are listed in Table II with associated confidence values. While other bond orders were not observed in organic molecules at the time of the survey for Si, B, P–S, P–Te, or P–Se, it is possible that future releases of the CSD will contain such molecules. Therefore, it is important to assign confidence values to these bond orders so that they can be modified later. None of the parameters for these atom pairs was required to process the test data set.

(ii) The pair of atoms may be singly or doubly bonded. In this case, three characteristic lengths of the atom pair were estimated from values in ref 10, namely, the ideal bond lengths

* Author to whom correspondence should be addressed.

Table III. Bond Length Characteristics[a]

atom pair	allowed bond orders	i_s	i_d	i_t	w_{sd}	w_{dt}
As–N	1, 2	1.86	1.835		1.845	
As–O	1, 2	1.71	1.66		1.68	
As–S	1, 2	2.28	2.08		2.15	
C–C	1, 2, 3	1.49	1.31	1.18	1.38	1.21
C–N	1, 2, 3	1.42	1.32	1.14	1.34	1.20
C–O	1, 2	1.41	1.22		1.28	
C–S	1, 2	1.78	1.68		1.70	
C–Te	1, 2	2.20	1.80		2.10	
N–N	1, 2	1.40	1.22		1.32	
N–O	1, 2	1.39	1.22		1.25	
N–P	1, 2	1.69	1.59		1.62	
N–S	1, 2	1.66	1.54		1.58	
N–Se	1, 2	1.83	1.79		1.80	
O–P	1, 2	1.60	1.48		1.52	
O–S	1, 2	1.58	1.45		1.54	
P–P	1, 2	2.23	2.04		2.06	

[a] All bond lengths are in Å.

for single and double bonds, i_s and i_d, and the watershed distance, w_{sd}, above which a bond is likely to be single and below which it is probably double. Values of these parameters were assigned "consensus" values based on the highly detailed bond length information about a variety of chemical environments given in the survey. These are listed in Table III. Ideal bond lengths are an average of the values given in Allen et al.[10] for a variety of environments while watershed values were estimated and reflect the standard deviations observed for different bond orders. Thus, for example, the watershed distances between single and double bonds are invariably closer to the ideal double bond distance because double bonds show less length variation. There may be scope for improving the algorithm by tuning these values. Each pair of bonded atoms was assigned an initial guess bond order depending on the measured bond length, l_{obs}, and a confidence value, C_{bo}, according to the formulas:

single bond, $l_{obs} > w_{sd}$:

$$C_{bo} = \frac{l_{obs} - w_{sd}}{2(i_s - w_{sd})}$$

double bond, $l_{obs} < w_{sd}$:

$$C_{bo} = \frac{w_{sd} - l_{obs}}{2(w_{sd} - i_d)}$$

(iii) The pair of atoms may be singly, doubly, or triply bonded. C–C and C–N bonds fall into this category and are treated similarly to the previous case, except that there are three ideal lengths, i_s, i_d, and i_t, and two watershed distances, w_{sd} and w_{dt}. Again, consensus values were derived from values in ref 10 and are listed in Table III. Initial guess bond orders were assigned depending on the measured bond length, l_{obs}, and confidence values, C_{bo}, were assigned according to the formulas:

single bond, $l_{obs} > w_{sd}$:

$$C_{bo} = \frac{l_{obs} - w_{sd}}{2(i_s - w_{sd})}$$

double bond, $i_d < l_{obs} < w_{sd}$:

$$C_{bo} = \frac{w_{sd} - l_{obs}}{w_{sd} - i_d}$$

double bond, $i_d > l_{obs} > w_{dt}$:

$$C_{bo} = \frac{l_{obs} - w_{dt}}{i_d - w_{dt}}$$

triple bond, $l_{obs} < w_{dt}$:

$$C_{bo} = \frac{w_{dt} - l_{obs}}{2(w_{dt} - i_t)}$$

Each bond order confidence, C_{bo}, is multiplied by 10 in order to have an appropriate value relative to coordination-state confidences.

Initial Assignment of Coordination States and Confidence Values. This is performed in a similar manner to the previous step, except that bond angles are used to estimate coordination states. Confidence values are assigned according to how close the average valence angle at an atom, a_{obs}, is to the ideal. Four categories of coordination state are considered: linear (ideal = 180°), trigonal (ideal = 120°), tetrahedral (ideal = 109.5°), and octahedral (ideal = 90°) with confidence values, C_{at}, assigned according to the formulas:

linear, $a_{obs} > 150°$:

$$C_{at} = \frac{a_{obs} - 150}{30}$$

trigonal, $120° > a_{obs} > 115°$:

$$C_{at} = \frac{a_{obs} - 115}{5}$$

trigonal, $150° > a_{obs} > 120°$:

$$C_{at} = \frac{150 - a_{obs}}{30}$$

tetrahedral, $109.5° > a_{obs} > 99°$:

$$C_{at} = \frac{a_{obs} - 99}{10.5}$$

tetrahedral, $115° > a_{obs} > 109.5°$:

$$C_{at} = \frac{115 - a_{obs}}{5.5}$$

octahedral, $a_{obs} < 99°$:

$$C_{at} = \frac{99 - a_{obs}}{9}$$

Each coordination-state confidence, C_{at}, is multiplied by 100 and the number of angles contributing to the average bond angle (i.e., ×300 for three angles) to produce appropriate values relative to the bond order confidences, C_{bo}.

Resolution of Coordination State/Bond Order Conflicts. There are three types of conflict between coordination state and bond orders. These are considered in turn and corrections

ASSIGNMENT OF CHEMICAL CONNECTIVITY

J. Chem. Inf. Comput. Sci., Vol. 32, No. 5, 1992 **403**

Table IV. Test Set

refcode	compd	R-factor	no. of atoms	no. of bond errors	hydrogens present	IDATM successful
ACTBOL	actinobolin	–	21	0	no	yes
ACTBOL	actinobolin	–	21	0	no	yes
ACTBOL	actinobolin	–	21	0	no	yes
ACTBOL	actinobolin	–	21	0	no	yes
NDMSCN	nodusmicin	0.0360	30	0	yes	yes
AAGAGG10	cyclo(Ala-Ala-Gly-Ala-Gly-Gly)	0.0373	27	0	yes	yes
PILLMA	pillaromycin A	0.0380	39	7	yes	yes
ACMBPN	2-amino-*N*-(3-dichloromethyl-3,4,4a,5,6,7-hexahydro-5,6,8-trihydroxy-4-yl)propanamide	0.0400	24	0	yes	yes
NBPENC	*p*-nitrobenzyl-5-pen-2-em-3-carboxylate	0.0414	22	0	yes	yes
ACFUCN	*N*-acetylfuranomycin	0.0430	14	0	yes	yes
OXOFMB	oxoformycin B	0.0430	20	0	yes	yes
NAHACA	hadacidin	0.0440	8	0	yes	yes
AAGGAG10	cyclo(Ala-Ala-Gly-Gly-Ala-Gly)	0.0445	27	0	yes	yes
APOMRC	apomorphine	0.0450	20	1	no	yes
APOMRC	apomorphine	0.0450	20	1	no	yes
MORPHM	morphine	0.0450	21	0	yes	yes
MORPHC	morphine	0.0460	21	0	yes	yes
NONACU	nonactin	0.0460	52	0	no	yes
TBUCBD10	tetra-*tert*-butylcyclobutadiene	0.0460	20	0	yes	yes
ANTMYC01	anthramycin methyl ester	0.0480	24	0	yes	yes
APYMPR	*β*-amino-*β*-(4-amino-6-carboxyamino-5-methylpyrimidin-2-yl)propionic acid amide	0.0480	17	2	no	no
BEVJER10	bis(dimethylammonium)octacyanotetra-methylidenecyclobutanediide	0.0480	24	0	yes	yes
TBUCBD02	tetra-*tert*-butylcyclobutadiene	0.0480	20	0	yes	yes
ANTMYC03	anthramycin methyl ester	0.0500	24	0	no	yes
NIGERI	nigericin	0.0500	51	0	no	yes
PIPBCX	6-*β*-phthalimide-6-*α*-methylpenam-3-*α*-*p*-bromocarboxanilide *β*-oxide	0.0500	33	0	no	yes
PIPCIL	piperacillin	0.0500	36	0	yes	yes
ACMPXC	4-acetyl-3-methyl-7-*β*-phenoxyacetamido-C^b-3-cephem	0.0510	24	0	yes	yes
ACANOB	*N*-acetylactinobolin	0.0520	24	0	yes	yes
BAMLIK	cyclo(Gly-His-Gly-Ala-Tyr-Gly)	0.0530	38	0	yes	yes
OXYTET01	oxytetracycline	0.0540	33	0	no	yes
PXMPEN	phenoxymethyl-C^b-2-desacetoxy-cephalosporin	0.0540	23	0	yes	yes
PURMYC10	puromycin	0.0550	34	2	yes	yes
TBUCBD01	tetra-*tert*-butylcyclobutadiene	0.0550	20	0	yes	yes
PODACE	phenoxymethyl-C^b-desacetoxy-cephalosporin	0.0560	24	0	yes	yes
AMICET10	amicetin	0.0600	44	0	yes	yes
OXYTET	oxytetracycline	0.0600	33	0	yes	yes
AMOXCT	amoxicillin	0.0610	25	0	yes	yes
CIMMUG	morphine	0.0610	22	0	yes	yes
NAHACB	hadacidin	0.0610	8	0	yes	yes
AMDMCN	amidinomycin	0.0620	14	0	no	yes
ANTETC	4-deamino-4-hydroxy-4,11a-anhydro tetracycline	0.0620	30	3	no	no
CIMNAN	morphine	0.0620	21	0	yes	yes
AGNGEC11	nigericin	0.0650	51	0	no	yes
ANTROS01	antibiotic A-130A	0.0650	60	0	yes	yes
NETRSN	netropsin	0.0670	31	5	yes	yes
NAPMYC10	naphthyridinomycin	0.0700	30	0	yes	yes
AZPCOH	*trans*-2-azabicyclo[2.1.0]pentane-3-carboxylic acid	0.0720	8	0	yes	yes
OXTETD	oxytetracycline	0.0720	33	0	yes	no
NEBULR	nebularine	0.0730	18	0	yes	yes
PRPENG	procaine	0.0730	17	0	yes	no
PRPENG	penicillin G	0.0730	23	0	yes	yes
AOTETC	5,12a-diacetyloxytetracycline	0.0760	39	0	yes	no
PRMESA	2,4-diamino-5-(3,4-dichlorophenyl)-6-methylpyrimidine	0.0760	17	0	yes	yes
OXERTH	9-deoxy-11-deoxy-9,11-(imino-(2-(2-methoxy-ethoxy)ethylidene)oxy)erythromycin	0.0770	58	0	yes	no
PENTBH10	tetrahydropentalenolactone	0.0770	18	0	yes	yes
OXTETK	oxytetracycline	0.0800	33	0	no	no
APLASM	aplasmomycin	0.0840	55	0	no	yes
ABHPTB	5'-anhydro-7-bromo-8-hydroxy-2',3'-isopropylidenetubercidin	0.0850	23	0	no	yes
ABHPTB	5'-anhydro-7-bromo-8-hydroxy-2',3'-isopropylidenetubercidin	0.0850	23	2	no	no
ANTSUL	antibiotic 593A	0.0866	22	0	yes	yes
ANFLCN	ancanthofolicin	0.0870	59	0	no	yes

Table IV (Continued)

refcode	compd	R-factor	no. of atoms	no. of bond errors	hydrogens present	IDATM successful
ANTINA	antibiotic K41 *p*-iodobenzoate	0.0930	75	0	no	yes
ACTDGU10	deoxyguanosine	0.0940	19	0	no	yes
ACTDGU10	deoxyguanosine	0.0940	19	0	no	no
NONACT	nonactin	0.1030	52	0	yes	yes
ANSMYC10	*N*-acetylbromoanisomycin	0.1050	23	3	no	yes
AMCILL	ampicillin	0.1060	24	0	yes	yes
ANTBPE	1-amino-1-(4-bromophenyl)ethanebis(dimethyl-(1-ethyl-4-(2-pyrrolocarbonyl)ethyltetra-hydroindanylbutadienyl)tetrahydropyran-2-acetic acid)	0.1070	34	0	yes	yes
ANTBRN	antibiotic K41 *p*-bromobenzoate	0.1070	75	3	no	no
NONAMT	nonactin	0.1080	52	0	yes	yes
ACIGRA	tri-*O*-acetyl-*O*-iodoacetylgranaticin	0.1090	45	0	no	yes
AERMYC10	anhydroerythromycin A cyclic carbonate methiodide	0.1090	53	0	no	no
OTETCB	oxytetracycline	0.1170	33	2	no	no
NONKCS	nonactin	0.1251	52	0	no	yes
MORPHI	morphine	0.1300	21	0	no	yes
PMEPEN	phenoxymethylpenicillin	0.1300	24	0	yes	no
PROMYC10	picric acid	0.1300	16	2	no	no
PROMYC10	toluene	0.1300	7	0	no	yes
PROMYC10	prolinomycin	0.1300	84	0	no	yes
PROMYC10	prolinomycin	0.1300	84	0	no	no
PRTYLD	protylonolide	0.1310	28	0	no	yes
PILLBA10	pillaranone monobromoacetate	0.1330	29	0	no	yes
HAZMOR	6-deoxy-6-azide-14-hydroxydihydroisomorphine	0.1340	24	0	yes	yes
AMPIAB10	*N*-iodoacetylamphotericin B	0.1370	69	2	no	yes
NIVBIO	novobiocin	0.1400	44	0	no	no
PEANNA	antibiotic A204A	0.1400	66	0	no	yes
PEANAG	antibiotic A204A	0.1500	66	0	no	yes
NONACS	nonactin	0.1720	52	0	yes	yes
NOSHEP10	nosiheptide	0.1800	82	10	no	no
PRMARI	9-propionylmaridomycin	0.1800	62	2	no	no

Table V. Initial Assignment of Chemical Connectivity of Toluene from PROMYC10[a]

(a) Coordination States

atom	av valence angle (no. of angles)	initial guess coordination state	confidence
1	120.000 (3)	trigonal	300
2	116.910 (1)	trigonal	38
3	118.837 (1)	trigonal	77
4	111.249 (1)	tetrahedral	75
5	134.345 (1)	trigonal	52
6	105.024 (1)	tetrahedral	57
7	– (0)		

(b) Bond Orders[b]

bond	length	initial guess bond order	confidence
1–2	1.111	triple	16.5
2–3	1.508	single	5.8
3–4	1.396	single	0.7
4–5	1.390	single	0.5
5–6	1.301	double	9.1
1–6	1.491	single	5.1
1–7	1.470	single	4.1

[a] Figure 1 indicates the atom numbering scheme, its elemental type, and the atomic connectivity. Angles are given in degrees. [b] Bond lengths are given in Å.
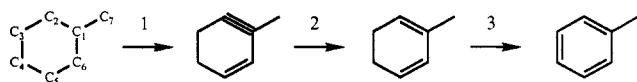

Figure 1. Assignment of bond orders to toluene from PROMYC10. The numbers indicate the steps in the assignment described in the text.

made to the structure until all conflicts are removed. If conflicts remain after application of the three rules below the process is started again.

Table VI. Typical Errors

molecule	OXBRIDGE structure	correct structure
APYMPR		
NETRSN		
ANTBRN		
PROMYC10		
AMPIAB10		
NOSHEP10		

Rule 1: nconns > valence

If the number of connections to an atom exceeds its valence there are two possible outcomes. First, it may be possible to change the valence of the atom as in the case of sulfur. If so, the valence is increased and the process starts again. Secondly, the number of connections may be greater than the highest

ASSIGNMENT OF CHEMICAL CONNECTIVITY

J. Chem. Inf. Comput. Sci., Vol. 32, No. 5, 1992 **405**

valency available to that atom, probably due to the low resolution of the structure. In this case, an error is logged, and the calculation continues with the incorrect valence, assigning chemical connectivity to the remainder of the structure in the usual way. Rule 1 was not invoked in processing the test data set.

Rule 2: nbonds > valence

If the number of bonds to an atom is greater than the valence of that atom, the order of the multiple bond with the lowest confidence is reduced by one.

Rule 3: nbonds + maxconns – nconns < valence

If the atom cannot satisfy its valence by adding hydrogens, either one of the bond orders or the coordination state is too low. The bond order of the single bond (or double if C–C or C–N) with the lowest confidence is increased by one if the confidence is less than that of the associated atom, or the confidence of the atom is greater than 95. Otherwise the atom coordination state is increased by one.

RESULTS AND DISCUSSION

The performance of the procedure was assessed by deriving chemical connectivity for the subset of the CSD used by Meng and Lewis[7] to test their IDATM program. The molecules were extracted from the CSD using the QUEST program.[9] Fractional coordinates of molecules were converted to Cartesian coordinates in the GSTAT program[11] (a coordinate file was written with the "OUTPUT COORD" command) and used as input to the OXBRIDGE program. Chemical connectivities were calculated for all the molecules in the data set using the procedure described in the Methods section.

Performance. Of the 91 molecules, 75 (82%) were correctly assigned chemical connectivity, which compares favorably with the Meng and Lewis correct atom type assignment to 73 structures (80%).[7] Of all bonds, 97% were correctly assigned, compared with 99% of atom types assigned by IDATM. This discrepancy is a result of "knock-on" effects of incorrectly assigned bond orders, so that, for example, a wrong keto/enol assignment gives two errors of bond order but only one atom type error. Table IV lists the molecules in order of R-factor, indicating those with hydrogen coordinates, the number of errors of bond order assignment by OXBRIDGE, and the performance of the IDATM program. In general, structures with hydrogen atom coordinates were more likely to be solved correctly than those without. Hydrogen atoms help to define atom geometry and coordination states more precisely.

Example of Success in Assigning Chemical Connectivity. Consider one example, the toluene structure in refcode PROMYC10, in which the chemical connectivity assignment procedure was successful. Figure 1 shows the steps taken by OXBRIDGE to assign bond orders to this molecule. Step 1 is to make an estimation of coordination states and bond orders. These are listed in Table V with their associated confidence values and the molecular geometry used in the calculation. The atom numbering scheme refers to Figure 1. The example illustrates how bad the initial assignment can be for a low resolution structure. Rules 1–3 are applied in sequence until no more changes need to be made to the structure. In step 2, rule 2 identifies atom 1 as having too many bonds and reduces the order of bond 1–2 from 3 to 2. In step 3, rule 3 identifies atom 3 as having insufficient total bonds or coordination state to attain its valency of 4. In this instance, the bond order of 3–4 is increased from 1 to 2. A further pass through the rules reveals that the structure is consistent and the procedure is complete. Note that, while none of the coordination states change in this example and there are

tetrahedral atoms in the final solution, the goal of the procedure is to obtain correct bond orders. Experience has shown that the rules are sufficient to produce a high success rate.

Examples of Errors in Chemical Connectivity Assignment. Table VI lists some typical errors taken from the test set. Some of the errors are of the keto/enol type (e.g., APYMPR and AMPIAB10) identified by Meng and Lewis as a problem with their approach. More serious errors, however, involve assigning bond orders to conjugated ring systems. The thiophen/3-hydroxypyridine construct in NOSHEP10 is a particularly demanding example in which OXBRIDGE finds a reasonable structure but not quite the correct one. ANTBRN and PROMYC10 are simpler molecules in which the current approach is unable to achieve the correct structure. There are seven structures which OXBRIDGE fails to assign correctly but which are successfully processed by IDATM. They are PILLMA, APOMRC, OXYTET01, PURMYC10, NETRSN, ANSMYC10, and AMPIAB10. Both keto/enol and ring assignment errors occur in these examples. The most common error occurs when an extended conjugated system is present and the iterative connectivity assignment procedure is unable to find the correct arrangement of alternating single and double bonds. Solutions in which the single and double bonds are switched are often the best that the algorithm can find, resulting in structures which have unsaturated atoms at the periphery of a conjugated system (see NETRSN in Table VI). Work is in progress to correct these problems.

CONCLUSION

A program to assign bond orders to crystal structures of organic molecules in the Cambridge Structural Database[6] has been developed.[12] The approach uses measured bond lengths and bond angles to make an initial guess of the chemical structure. Bond orders and atom coordination states are altered according to a set of rules until a satisfactory structure is encountered. A test set of 91 molecules was processed with 82% success, an slight improvement on the atom-type assignment method of Meng and Lewis,[7] albeit to perform a slightly different task. The approach has problems distinguishing keto and enol forms when hydrogen atom coordinates are not present in the crystal structure, and with conjugated ring systems in which correct placement of single and double bonds is challenging.

ACKNOWLEDGMENT

REFERENCES AND NOTES

(1) Van Drie, J. H.; Weininger, D.; Martin, Y. C. ALADDIN: An integrated tool for computer-aided 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.

(2) Sheridan, R. P.; Rusinko, A.; Nilakantan, R.; Venkataraghavan, R. Searching for pharmacophores in large coordinate data bases and its use in drug design. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 8165–8169.

(3) DesJarlais, R. L.; Seibel, G. L.; Kuntz, I. D.; Furth, P. S.; Alvarez, J. C.; Ortiz de Montellano, P. R.; DeCamp, D. L.; Babe, L. M.; Craik, C. S. Structure-based design of nonpeptide inhibitors specific for the human immunodeficiency virus I protease. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 6644–6648.

(4) Leach, A. R.; Prout, K. Automated conformational analysis: Directed conformational search using the A* algorithm. *J. Comput. Chem.* **1990**, *11*, 1193–1205.

(5) Rusinko, A.; Skell, J. M.; Balducci, R.; McGarity, C. M.; Pearlman, R. S. CONCORD: A program for the rapid generation of high quality approximate 3-dimensional molecular structures; Tripos Associates: St. Louis, 1988.

(6) Allen, F. H.; Bellard, S.; Brice, M. D.; Cartwright, B. A.; Doubleday, A.; Higgs, H.; Hummelink, T.; Hummelink-Peters, B. G.; Kennard, O.; Motherwell, W. D. S.; Rodgers, J. R.; Watson, D. G. The Cambridge crystal data centre: computer-based search, retrieval, analysis and display of information. *Acta Crystallogr.* **1979**, *B35*, 2331–2339.

(7) Meng, E. C.; Lewis, R. A. Determination of molecular topology and atomic hydridization states from heavy atom coordinates. *J. Comput. Chem.* **1991**, *12*, 891.

(8) MACCS is an acronym for Molecular Access System, a chemical database management system supplied by Molecular Design Limited, San Leandro, CA.

(9) QUEST, Cambridge Structural Database, Version 3.4, User's Manual, Part I, 1989.

(10) Allen, F. H.; Kennard, O.; Watson, D. G.; Brammer, L.; Orpen, A. G.; Taylor, R. Tables of bond lengths determined by X-ray and neutron diffraction. Part I. Bond lengths in organic compounds. *J. Chem. Soc. Perkin Trans. 2* **1987**, S1–S19.

(11) GSTAT, Cambridge Structural Database, Version 3.4, User's Manual, Part II, 1989.

(12) Personal communication. A referee pointed out that, while the CSD has been used as a source of data, the OXBRIDGE algorithm will not be required for routine automatic analysis of CSD structures in the near future. The advent of matching records between the crystal connectivity data and the 2D bond order data in version 5 of the database will eliminate this need.