A Stereochemically Accurate Chemical Substance Database Based on the Systematic Names of Organic Compounds. 1. Low Molecular Weight Organic Compounds

KEISUKE ARAKI* and MASANORI KAJI

Japan Information Center of Science & Technology (JICST), 5-2 Nagatacho 2 Chome, Chiyoda-ku,
Tokyo 100, Japan

A chemical substance database, containing chemical structures generated automatically from their IUPAC systematic names, is described. The structure generation program can convert the names of natural products, peptides, and stereochemically complex names into stereochemically correct connection tables. More than 300 000 commonly encountered compounds are stored in the database and are publicly accessible through the JOIS-F online system, which also offers mass spectral, thermochemical, and other factual databases.

1. INTRODUCTION

The Japan Information Center of Science and Technology (JICST) has developed a chemical substance dictionary called STARS (STereochemically Accurate Registry of Substances). The system has two main purposes. The first is to facilitate indexing of the chemical literature and the technology literature, and the second purpose is to register the chemical substances in various databases, assigning unique identification numbers to each, thus allowing cross-referencing between the dictionary and the different databases. The construction of a chemical information network system has been supported since 1981 by the Japanese Science and Technology Agency (STA) and has entailed the cooperative efforts of nine Research Institutions and Organizations involved in the development of factual databases.[1] Indexing began at the end of 1989, and the system has been publicly accessible since 1988 via the JICST Online Information System of Factual Data (JOIS-F).

The chemical substance dictionary database (STARS) is necessary for several reasons. First, Japanese chemical names are essential in the record-keeping required by Japanese law and recorded in Japan's Official Gazette. Second, the chemical structure-containing databases that presently are available do not routinely handle stereochemistry either in searching or in displays. Stereoisomerism is common among both synthetic chemicals and natural products such as sugars, peptides, antibiotics, terpenes, and alkaloids and is often pivotal for biological activity. There is thus a need for a chemical substance dictionary that will deal with stereochemistry correctly.[2,7] Third, a national information network requires control of a chemical dictionary that allows entry and retrieval of chemical structures.

Establishing chemical substance databases requires considerable technical skill and is expensive, and these factors ensure that such a project will be subject to serious scrutiny. In spite of this, a comprehensive nomenclature-to-structure translation system has been developed, building on the work of Vander Stouw.[4,5] The resulting system will produce the structure of virtually any chemical compound whose name is available.

2. NOMENCLATURE TRANSLATION SYSTEM

A detailed description of Vander Stouw's system is not available, and the method does not appear to be completely adequate. A thorough examination of the IUPAC systematic chemical nomenclature was carried out, and the following criteria were established for a nomenclature translation program:

1. Almost 100% of names must be translated.
2. The nomenclature used must be based upon the IUPAC rules.
3. Slight revision or rewriting of systematic names to fit the computer program is permissible.

After two years of system analysis, two further years were spent on detailed design of the system and programming and coding of the "analytical dictionary"—a dictionary of chemical fragment names and the corresponding structures. As a result, a highly efficient nomenclature translation system, known as the Stereochemically Accurate Registry of Substances (STARS) has been developed.[3,6,7]

2.1. General Design. The processing routine classifies chemicals into seven categories, as follows:

type 1	clearly defined organic compounds of low molecular weight containing less than 248 nonhydrogens
type 2	Markush compounds
type 3	inorganic compounds
type 4	alloys
type 5	polymers
type 6	mixtures including racemates and relative stereoisomers
type 7	undefined substances with only common names, and other organic compounds having more than 248 non-hydrogens

Depending upon the type, the chemical names that are entered are processed by different subprograms. Compounds of type 1 are identified at the level of the absolute configuration, which is represented by the connection table. Markush formulations (type 2) are used as input for compounds in either the chemical literature or patents and entered as tables or as abbreviated structural diagrams. Systematic names are written according to formulas containing variables defined by the substituents, and the corresponding individual names are generated by a second program. Inorganic compounds that are clearly defined in the stoichiometrical sense are treated as type 1 compounds but where the stoichiometry is undefined, inorganic compounds are entered into the Quantity Table (QT) as alloys. Alloys (type 4) are entered into the QT as a list of their components. Polymers are registered in the QT as repeating units or monomers. Mixtures of various sorts including salts (addends), onium salts, hydrates, racemates, and relative
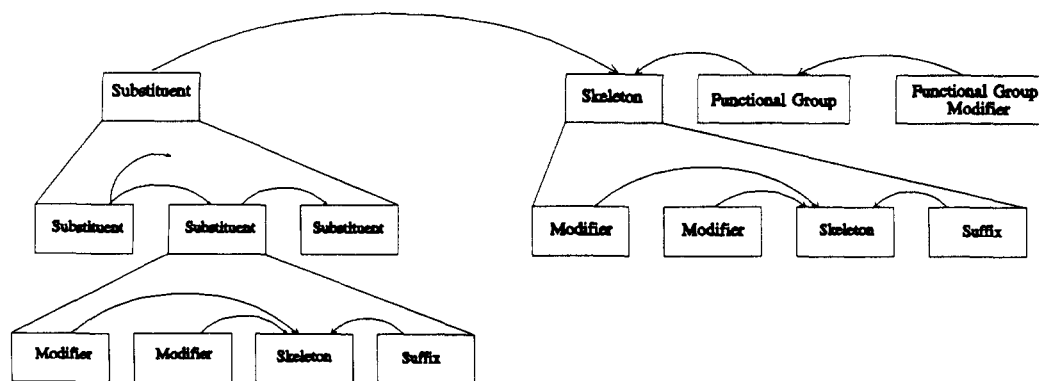
Figure 1. General pattern of systematic names.

stereoisomers are identified as such in the QT.

The record for each compound in the Quantity Table include substance names, classification codes, ID number and ID numbers of any components in the molecule, atomic ratio, weight percent, and atom percent. With racemates and relative stereoisomers, a pair of stereospecific connection tables are generated from the nomenclature. These have rac- or rel-prefixes, and an asterisk is used to tag each chiral center descriptor, as in R^*, S^*, or α^* and so on. Substances of undefined structure (type 7) are identified only by name, although the molecular formula is known in some cases.

2.2. Systematic Names of Organic Compounds. Systematic names generated according to the IUPAC rules (see Figure 1) and therefore including most CAS names, generally fall into any of three basic component classes: main components, subcomponents, and auxiliary components.

Skeletal names, substituents, functional groups, skeleton-modifying operators such as nor, homo, seco, spiro, aza, and oxa belong to the main component class. Subcomponents include locants, multipliers, and suffixes such as yl, ylidene, and ene, stereochemical symbols, element symbols in the case of unusual isotopes and + or – signs indicating optical rotatory power. A variety of auxiliary characters such as commas, hyphens, colons, semicolons, periods, brackets, parentheses, dots, and slashes are also encountered.

Skeletons are classified as follows:
(1) one node elements
(2) hydrogenated one node atoms
(3) rings and ring systems requiring locants for substituent positions
(4) acyclic hydrocarbons and hetero chains
(5) rings and ring systems requiring hydrogen, as indicated
(6) skeletons including main functional groups
(7) skeletons with unspecified multiple bonds
(8) skeletons with main functional groups at unspecified positions

Functional groups are classified as follows:
(1) Functional groups with special connections. These replace terminal carbon atoms to form functional group containing compounds, such as hexen + oic acid or hexane + nitrile.
(2) Functional groups with simple connections (e.g., carboxylic acid, ol, one).
(3) Amine (used both as a simple functional group and as a substituent-requiring skeleton, as in benzenamine or phenylamine).
(4) Functional groups of the halogenide type.
(5) Functional groups of the chalcogenide type.

Substituents are classified as follows:
(1) Not itself substituted, but only a substituent, e.g., chloro, nitro, oxo.
(2) Not itself substituted, but directly connected and

carrying a second substituent, e.g., carbonyl, oxy, thio.
(3) Substituted directly, connected directly, and also a substituent, e.g., imino, methylene.
(4) Directly substituted, e.g., amino.
(5) Substituted directly, but not further substituted by alkyl radicals, e.g., methyl, methoxy, acetyl.
(6) Amino acid residues, designated by the standard three-letter code, e.g., Phe-, Ala-, Gly-, Lys-.
(7) Amino acid residues with two valences, e.g., aspartoyl, glutamoyl.
(8) Ring substituents requiring locants to be substituted for each other, e.g., phenyl, benzoyl.
(9) Alkyl, alkenyl, alkynyl.
(10) Ring substituents with two valences requiring locants to be substituted for each other, e.g., phenylene.
(11) Alkylidene, alkylidyne, alkenylidene, alkenylidyne.

The types of locant that may be used are as follows:
(1) Numbers or alphanumeric combinations, e.g., 1, 2, 3, 3a, 3b.
(2) Superscripted numbers, e.g., 1^7, 3^1.
(3) Lower-case letters, e.g., o-, m-, p-.
(4) Lower-case Greek characters, e.g., α, β, γ.
(5) Upper-case letter as ring identifier in steroids or terpenes, e.g., A-norcholestane.
(6) Element symbols.
(7) Superscripted element symbols.
(8) Number–element symbol combination, e.g., 1N, 3N.
(9) Number–element symbol combination, e.g., 1-S, 4-O.
(10) Isotopic element symbol.
(11) Number(number) combination, e.g., 5(10), 5a(10).
(12) Number(number)lower-case letter combination, e.g., 3(8)a.
(13) Upper-case letter(number) combination, e.g., A(1), D(15).
(14) Upper-case letter followed by (number–lower-case letter) combination, e.g., A(9a), AB(10a).
(15) Number followed by isotopic element symbol.
(16) Two element symbols, e.g., SS, TeTe.

The stereodescriptors used are defined as follows:

R, S	absolute configuration of chiral atoms
R, S	absolute configuration of chiral axes of allenes
aR, aS	absolute configuration of chiral axes of non-allenes
pR, pS	absolute configuration of a chiral plane
E, Z	geometric isomerism
α, β, ξ	upper or lower position with respect to ring plane
syn, anti	direction of a substituent on a bicyclo bridge
D, L	absolute stereochemical symbols in sugars and amino acids

Table I. Pointers to Tables II–VIII[a]

	front							
rear	--V	L-V	-MV	LMV	--B	L-B	-MB	LMB
--V	VIII		IV		VIII		IV	
L-V	VII		III		VII		III	
-MV	VI							
LMV	V				V			
--B	VIII		IV		VIII		IV	
L-B	VII		III		VII		III	
-MB	VI							
LMB	V				V			

[a] The numbers in this table refer to the corresponding Tables II–VIII.

gluco, ribo	relative configurations of OH groups in sugars
rel, R*, S*	relative configuration
rac	racemic mixture

Analysis of the dictionary using these definitions shows the following breakdown:

skeletons	31 000	e.g., methane, benzene, acetic acid
functional groups	1 000	e.g., carboxylic acid, ol, chloride
substituents	1 500	e.g., methyl, phenyl, hydroxy, chloro
suffixes	20	e.g., yl, ylidene, ene, ium
bridges	230	e.g., methano, ethano, epoxy, epimino
functional operators	1 150	e.g., homo, nor, seco, aza, oxa, deoxy
multipliers	150	e.g., di, tri, tetra, bis, tris, tetrakis
total	35 050	

3. ANALYSIS OF ENTERED SYSTEMATIC NAMES AND CONSTRUCTION OF CONNECTION TABLES

At input, a systematic name is analyzed, and its elements are placed in one of three boxes in memory. The first box is for locants, the second for multipliers, and the third for significant items from the analysis dictionary. A set of these three boxes makes a unit, and each unit has a sequence number and a level number. The level number is incremented by 1 when a unit is preceded by a left parenthesis "(", or bracket "[", and is decremented by 1 if the unit is preceded by ")" or "]". Each unit is characterized according to box content as, for example, LMV, L-V, --V, -MV, L-B, and so on, where L = locant, M

= multiplier, V = significant dictionary item, B = bracket, and - = blank. These three-letter codes indicate the occurrence of name fragment patterns in systematic names. The patterns of each two units from the beginning toward the end of the name are compared with those in Table I to determine whether they can be connected to one another. If such a connection is permissible, careful examination is made of the appropriate section of Tables II–VIII (as indicated by Table I) in order to determine the order of the actual connections between skeletons, functional groups, and substituents.

For example, the pattern LMV represents 2,3-dimethyl, and the pattern - -V is simply phenyl, oxy or amino. According to Table I, these two patterns may be connected, and one is referred to Table V for details as to the type of connection. According to Table V, however, methyl (code S24) can be connected to phenyl (code S40), but it may not be connected to oxy (code S20). Methyl and oxy can be connected only in the case of patterns of the - -V type, as indicated in Table VIII.

A connection table is then constructed as follows:

(1) Division of a word into its parts, e.g., enal → ene, al; ynal → yne, al.

(2) Clarification of uncertain structures by use of locants, e.g., 2-hexene signifies insertion of a double bond between C_2 and C_3.

(3) Assembly of spiro structures, e.g., spiro[5.2.3.2]-tetradecane.

(4) Modification of skeletons by functional operators, e.g., ring assembly by bi, ter, quater; ring modification by homo, nor, seco, cyclo, aza, oxa; deletion of radicals by de or des; removal of water and formation of the anhydride by anhydro or anhydride; bridge formation using bridge element names, e.g., methano, ethano, epoxy; processing of indicated hydrogens.

(5) Processing of suffixes, e.g., ene, yne, ylidine, ylidene, ylene, ium, ylium, yde, ylide.

(6) Connection of main functional groups to a skeleton, e.g., 2-naphthalene + carboxamide.

(7) Processing of hydrogens in response to hydro and dehydro prefixes.

(8) Connection between substituents and between substituents and a skeleton, e.g., 4-(3-amino + phenyl) + -2-naphthalenecarboxamide.

Table II. Connections and Classification Codes[a]

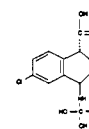|	K10	K13	K15	K17	K20	K21	S20	S21	S22	S23	S24	S30	S32	S40	S41	S42	S43	S44	A11	A13	E11	F30	F40	F41	F50
S10																									
S20			A2 K15	A2 K17	A2 K20	A2 K21	C1 S20	C1 S20						A2 S42	A2 S42	C1 S42			A2 A11	A2 A13	A2 E11				
S21			A2 K15	A2 K17	A2 K20	A2 K21	C1 S20	C1 S20						A2 S42	A2 S42	C1 S42			A2 A11	A2 A13	A2 E11				
S23																									
S22																									
S24																									
S30																									
S32																									
S40																									
S41																									
S42			A2 K15	A2 K17	A2 K20	A2 K21	C1 S20	C1 S20						A2 S42	A2 S42	C1 S42			A2 A11	A2 A13	A2 E11				
S43			A2 K15	A2 K17	A2 K20	A2 K21	C1 S20	C1 S20						A2 S42	A2 S42	C1 S42			A2 A11	A2 A13	A2 E11				
S44			A2 K15	A2 K17	A2 K20	A2 K21	C1 S20	C1 S20						A2 S42	A2 S42	C1 S42			A2 A11	A2 A13	A2 E11				

[a] The numbers in each box are codes indicating (top) the method of connection (see Table IX for an explanation) and (bottom) the new, postconnection classification codes.

366   J. Chem. Inf. Comput. Sci., Vol. 31, No. 3, 1991

ARAKI AND KAJI

Table III. Connections and Classification Codes

	K10	K13	K15	K17	K20	K21	S20	S21	S22	S23	S24	S30	S32	S40	S41	S42	S43	S44	A11	A13	E11	F30	F40	F41	F50
S10																									
S20	B1/K10	B1/K13	B1/K15	B1/K17	B1/K20	B1/K21	C1/S20	C1/S20								C1/S42					B1/E11	B1/K20	C1/K20		
S21	B1/K10	B1/K13	B1/K15	B1/K17	B1/K20	B1/K21	C1/S20	C1/S20								C1/S42					B1/E11	B1/K20	C1/K20		
S23																									
S22																									
S24																									
S30																									
S32					B1/K20								B1/S42												
S40																									
S41																									
S42	B1/K10	B1/K13	B1/K15	B1/K17	B1/K20	B1/K21	C1/S20	C1/S20								C1/S42					B1/E11	B1/K20	C1/K20		
S43	B1/K10	B1/K13	B1/K15	B1/K17	B1/K20	B1/K21	C1/S20	C1/S20								C1/S42					B1/E11	B1/K20	C1/K20		
S44	B1/K10	B1/K13	B1/K15		B1/K20	B1/K21	C1/S20	C1/S20								C1/S42					B1/E11	B1/K20	C1/K20		

Table IV. Connections and Classification Codes

	K10	K13	K15	K17	K20	K21	S20	S21	S22	S23	S24	S30	S32	S40	S41	S42	S43	S44	A11	A13	E11	F30	F40	F41	F50
S10			A2/K15	A2/K17	A2/K20	A2/K21		A2/S21	A2/S22	A2/S23	A2/S24	A2/S30	A2/S32	A2/S40	A2/S41	A2/S42	A2/S43	A2/S44	A2/A11	A2/A13	A2/E11	A2/F30	A2/F40		A2/F50
S20			A2/K15	A2/K17	A2/K20	A2/K21						A2/S30	A2/S32	A2/S40	A2/S41	A2/S42	A2/S43	A2/S44							A2/F50
S21			A2/K15	A2/K17	A2/K20	A2/K21		A2/S21	A2/S22			A2/S30	A2/S32	A2/S40	A2/S41	A2/S42	A2/S43	A2/S44	A2/A11	A2/A13	A2/E11	A2/F30	A2/F40		A2/F50
S22			A2/K15	A2/K17	A2/K20	A2/K21		A2/S21	A2/S22	A2/S23	A2/S24	A2/S30	A2/S32	A2/S40	A2/S41	A2/S42	A2/S43	A2/S44	A2/A11	A2/A13	A2/E11	A2/F30	A2/F40		A2/F50
S23			A2/K15	A2/K17	A2/K20	A2/K21		A2/S21	A2/S22	A2/S23	A2/S24	A2/S30	A2/S32	A2/S40	A2/S41	A2/S42	A2/S43	A2/S44	A2/A11	A2/A13	A2/E11	A2/F30	A2/F40		A2/F50
S24			A2/K15	A2/K17	A2/K20	A2/K21		A2/S21	A2/S22			A2/S30	A2/S32	A2/S40	A2/S41	A2/S42	A2/S43	A2/S44	A2/A11	A2/A13	A2/E11	A2/F30	A2/F40		A2/F50
S30			A2/K15	A2/K17	A2/K20	A2/K21		A2/S21	A2/S22	A2/S23	A2/S24	A2/S30	A2/S32	A2/S40	A2/S41	A2/S42	A2/S43	A2/S44	A2/A11	A2/A13	A2/E11	A2/F30	A2/F40		A2/F50
S32																									
S40			A2/K15	A2/K17	A2/K20	A2/K21		A2/S21	A2/S22	A2/S23	A2/S24	A2/S30	A2/S32	A2/S40	A2/S41	A2/S42	A2/S43	A2/S44	A2/A11	A2/A13	A2/E11	A2/F30	A2/F40		A2/F50
S41			A2/K15	A2/K17	A2/K20	A2/K21		A2/S21	A2/S22			A2/S30	A2/S32	A2/S40	A2/S41	A2/S42	A2/S43	A2/S44	A2/A11	A2/A13	A2/E11	A2/F30	A2/F40		A2/F50
S42								A2/S21	A2/S22																
S43			A2/K15	A2/K17	A2/K20	A2/K21		A2/S21	A2/S22			A2/S30	A2/S32	A2/S40	A2/S41	A2/S42	A2/S43	A2/S44	A2/A11	A2/A13	A2/E11	A2/F30	A2/F40		A2/F50
S44			A2/K15	A2/K17	A2/K20	A2/K21		A2/S21	A2/S22			A2/S30	A2/S32	A2/S40	A2/S41	A2/S42	A2/S43	A2/S44	A2/A11	A2/A13	A2/E11	A2/F30	A2/F40		A2/F50

(9) Secondary modification of the main functional group, e.g., cyclohexanone + oxime, 5-hydroxypentanoic acid lactone.

(10) Assignment of stereodescriptors to the appropriate atoms.

(11) Bond check.

(12) Aromatization and tautomerization.

(13) Rearrangement of the connection table into one that is stereospecific in terms of stereodescriptors, e.g., (R)-2-propanol → stereospecific CT.

(14) Generation of stereodescriptors and their assignment to specific atoms, e.g., 2-amino-2-deoxy-D-glucose → generation of stereodescriptors.

(15) Standardization of connection table by the extended and modified Morgan algorithm.
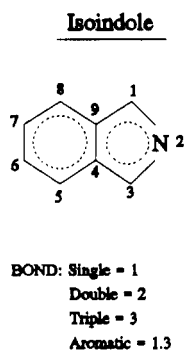
(16) Final check.

Steps 1–6 are complete first, then steps 7–11 are carried out beginning at the deepest level of parentheses. When all the components have been assembled, steps 11–16 are completed.

3.1. Connectivity Examination and Assembly of Connection Table.

The construction of a connection table is demonstrated with the following compound:

1. Morphological Analysis

(1R,3S) -(| (S)-1-HYDROXYETHYL | AMINO)-5-CHLORO-2,3-DIHYDRO-1H-ISOINDOLE-1-CARBOXYLIC ACID

class code   (|) 3-(| | |) 5-CHLORO 2,3 DI HYDRO 1H ISOINDOLE 1-CARBOXYLIC ACID

Level 0   1R,3S

Level 1   ( ) 1-HYDROXY ETHYL | AMINO

Level 2

Level 3

unit pattern | L-8 -8   L-V   -V   -8 -V   -8   L-V   LMV   -V   L-V

stereo   S10   S41   S22   S10   hydrogen hydrogen K34   F20

Components [12] and [13] are processed first. The code for isoindole (K34) is retrieved from the analysis dictionary,

Isoindole

BOND: Single = 1
Double = 2
Triple = 3
Aromatic = 1.3

Atom No	Locant 1	Locant 2	Bt	CV	HN	St	B₁	A₁	A₂	A₃	B₂
1	1		C	3	1			9 1.3	2 1.3		
2	2		N	3				1 1.3	3 1.3		
3	3		C	3	1			2 1.3	4 1.3		
4	3a		C	4				3 1.3	5 1.3	9 1.3	
5	4		C	3	1			4 1.3	6 1.3		
6	5		C	3	1			5 1.3	7 1.3		
7	6		C	3	1			6 1.3	8 1.3		
8	7		C	3	1			7 1.3	9 1.3		
9	7a		C	4	1			8 1.3	1 1.3	4 1.3	

Bt:Element
CV:Connected Valence
HN:Number of Hydrogens
St:Stereodescriptor

Figure 2. Building the connection table.
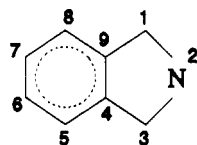
1H-isoindole

Atom No	Locant 1	Locant 2	Bt	CV	HN	St	B₁	A₁	A₂	A₃	B₂
1	1		C	3^2	1^2			9 1.3^1	2 1.3^1		
2	2		N	3				1 1.3^1	3 1.3^2		
3	3		C	3	1			2 1.3^2	4 1.3^1		
4	3a		C	4				3 1.3^1	5 1.3	9 1.3	
5	4		C	3	1			4 1.3	6 1.3		
6	5		C	3	1			5 1.3	7 1.3		
7	6		C	3	1			6 1.3	8 1.3		
8	7		C	3	1			7 1.3	9 1.3		
9	7a		C	4	1			8 1.3	1 1.3^1	4 1.3	

Connected Valence - 0.5 < ΣBond Value ≤ Connected Valence

Figure 3. Treatment of explicit hydrogens.

2,3-dihydro-1H-isoindole

Atom No	Locant 1	Locant 2	Bt	CV	HN	St	B₁	A₁	A₂	A₃	B₂
1	1		C	2	2			9 1	2 1		
2	2		N	3^2	0^1			1 1	3 2^1		
3	3		C	3^2	1^2			2 2^1	4 1		
4	3a		C	4				3 1	5 1.3	9 1.3	
5	4		C	3	1			4 1.3	6 1.3		
6	5		C	3	1			5 1.3	7 1.3		
7	6		C	3	1			6 1.3	8 1.3		
8	7		C	3	1			7 1.3	9 1.3		
9	7a		C	4				8 1.3	1 1	4 1.3	

Figure 4. Treatment of "hydro" qualifiers.

and the corresponding connection table shown in Figure 2 is used as a starting point. The hydrogen indicated by the "1H" is attached to position 2 of the isoindole nucleus, and the bond order in this nucleus is established giving the new connection tables shown in Figure 3. The classification code of the isoindole ([13] in the above morphological analysis) is then

Table V. Connections and Classification Codes

	K10	K13	K15	K17	K20	K21	S20	S21	S22	S23	S24	S30	S32	S40	S41	S42	S43	S44	A11	A13	E11	F30	F40	F41	F50
S10	B2/K10	B2/K13	B2/K15	B2/K17	B2/K20	B2/K21		B2/S21	B2/S22	B2/S23	B2/S24			B2/S40	B2/S41	B2/S42	B2/S43	B2/S44	B2/A11	B2/A13	B2/E11	B2/F30		B2/F41	B2/F50
S20	B2/K10	B2/K13																							
S21	B2/K10	B2/K13	B2/K15	B2/K17	B2/K20	B2/K21			B2/S22					B2/S40	B2/S41	B2/S42	B2/S43	B2/S44	B2/A11	B2/A13	B2/E11				
S22	B2/K10	B2/K13	B2/K15	B2/K17	B2/K20	B2/K21		B2/S21	B2/S22	B2/S23	B2/S24			B2/S40	B2/S41	B2/S42	B2/S43	B2/S44	B2/A11	B2/A13	B2/E11	B2/K20	C1/K20		B2/F50
S23	B2/K10	B2/K13	B2/K15	B2/K17	B2/K20	B2/K21		B2/S21	B2/S22	B2/S23	B2/S24			B2/S40	B2/S41	B2/S42	B2/S43	B2/S44	B2/A11	B2/A13	B2/E11	B2/K20			B2/F50
S24	B2/K10	B2/K13	B2/K15	B2/K17	B2/K20	B2/K21		B2/S21	B2/S22					B2/S40	B2/S41	B2/S42	B2/S43	B2/S44	B2/A11	B2/A13	B2/E11	B2/K20	C1/K20		B2/F50
S30	B2/K10	B2/K13	B2/K15	B2/K17	B2/K20	B2/K21		B2/S21	B2/S22	B2/S23	B2/S24			B2/S40	B2/S41	B2/S42	B2/S43	B2/S44	B2/A11	B2/A13	B2/E11	B2/K20			B2/F50
S32																									
S40	B2/K10	B2/K13	B2/K15	B2/K17	B2/K20	B2/K21		B2/S21	B2/S22	B2/S23	B2/S24			B2/S40	B2/S41	B2/S42	B2/S43	B2/S44	B2/A11	B2/A13	B2/E11	B2/K20	B2/K20		B2/F50
S41	B2/K10	B2/K13	B2/K15	B2/K17	B2/K20	B2/K21		B2/S21	B2/S22					B2/S40	B2/S41	B2/S42	B2/S43	B2/S44	B2/A11	B2/A13	B2/E11	B2/K20	B2/K20		B2/F50
S42																									
S43									B2/S22														B2/K20		
S44									B2/S22														B2/K20		

Table VI. Connections and Classification Codes

	K10	K13	K15	K17	K20	K21	S20	S21	S22	S23	S24	S30	S32	S40	S41	S42	S43	S44	A11	A13	E11	F30	F40	F41	F50
S10			A2/K15	A2/K17	A2/K20	A2/K21	C1/S10	A2/S21	A2/S22	A2/S23	A2/S24	A2/S30	A2/S32	A2/S40	A2/S41	A2/S42	A2/S43	A2/S44	A2/A11	A2/A13	A2/E11	A2/K20	A2/F40		A2/F50
S20			A2/K15	A2/K17	A2/K20			A2/S21	A2/S22	A2/S23	A2/S24	A2/S30	A2/S32	A2/S40	A2/S41	A2/S42	A2/S43	A2/S44	A2/A11		A2/E11				
S21			A2/K15	A2/K17	A2/K20	A2/K21		A2/S21	A2/S22	A2/S23	A2/S24	A2/S30	A2/S32	A2/S40	A2/S41	A2/S42	A2/S43	A2/S44	A2/A11	A2/A13	A2/E11	A2/K20	A2/F40		A2/F50
S22			A2/K15	A2/K17	A2/K20	A2/K21	C1/S10	A2/S21	A2/S22	A2/S23	A2/S24	A2/S30	A2/S32	A2/S40	A2/S41	A2/S42	A2/S43	A2/S44	A2/A11	A2/A13	A2/E11	A2/K20	A2/F40		A2/F50
S23			A2/K15	A2/K17	A2/K20	A2/K21	C1/S10	A2/S21	A2/S22	A2/S23	A2/S24	A2/S30	A2/S32	A2/S40	A2/S41	A2/S42	A2/S43	A2/S44	A2/A11	A2/A13	A2/E11	A2/K20	A2/F40		A2/F50
S24			A2/K15	A2/K17	A2/K20	A2/K21	C1/S10	A2/S21	A2/S22			A2/S30	A2/S32	A2/S40	A2/S41	A2/S42	A2/S43	A2/S44	A2/A11	A2/A13	A2/E11	A2/K20	A2/F40		A2/F50
S30			A2/K15	A2/K17	A2/K20	A2/K21	C1/S10	A2/S21	A2/S22	A2/S23	A2/S24	A2/S30	A2/S32	A2/S40	A2/S41	A2/S42	A2/S43	A2/S44	A2/A11	A2/A13	A2/E11	A2/K20	A2/F40		A2/F50
S32																									
S40			A2/K15	A2/K17	A2/K20	A2/K21	C1/S10	A2/S21	A2/S22	A2/S23	A2/S24	A2/S30	A2/S32	A2/S40	A2/S41	A2/S42	A2/S43	A2/S44	A2/A11	A2/A13	A2/E11	A2/K20	A2/F40		A2/F50
S41			A2/K15	A2/K17	A2/K20	A2/K21	C1/S10	A2/S21	A2/S22			A2/S30	A2/S32	A2/S40	A2/S41	A2/S42	A2/S43	A2/S44	A2/A11	A2/A13	A2/E11	A2/K20	A2/F40		A2/F50
S42			A2/K15	A2/K17	A2/K20									A2/S40	A2/S41	A2/S42	A2/S43	A2/S44	A2/A11		A2/E11				A2/F50
S43			A2/K15	A2/K17	A2/K20	A2/K21		A2/S21	A2/S22	A2/S23	A2/S24	A2/S30	A2/S32	A2/S40	A2/S41	A2/S42	A2/S43	A2/S44	A2/A11	A2/A13	A2/E11	A2/K20	A2/F40		A2/F50
S44			A2/K15	A2/K17	A2/K20	A2/K21		A2/S21	A2/S22			A2/S30	A2/S32	A2/S40	A2/S41	A2/S42	A2/S43	A2/S44	A2/A11	A2/A13	A2/E11	A2/K20	A2/F40		A2/F50

changed from K34 (a code indicating a skeleton requiring hydrogen) to K15 (code of a fully defined skeleton) and the element [12] is then erased.
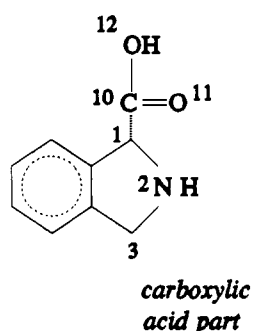
The "dihydro" qualifier is dealt with as shown in Figure 4, and the connections between the skeleton and the main functional groups are established as shown in Figures 5 and 6. Then, starting from the deepest level of parentheses or brackets, the possibility of connecting two elements from front to back is examined, using Table I. When such a possibility is found, the actual connection and the new classification codes are determined by using Tables II–VIII. At positions where the hydro qualifier is assigned, the node attributes are changed; CV (the connected valence) is decreased by 1 and HN (the number of hydrogens) is increased by 1.

In the case of the deeply nested units [5] and [6], which have the patterns L-V and --V, respectively, Table I indicates that they may be connected and reference is made to Table VII, which indicates that S10 and S41 may be connected by the method A2, i.e., carboxy (S10) is attached to position 1

of ethyl (S41). The S41 radical always requires a locant (a position number) to connect radicals. One hydrogen is removed from the corresponding position. In this case, the new classification code is also S41. A stereodescriptor is assumed to be assigned within the atoms at this level. Next, components [3]–[7] are consolidated, and the new pattern --B is assigned. Then the possibilities for a connection between --B and --V are examined by using Table I. This table indicates that in this case, Table VIII must be consulted, and there it is learned that (S41 can be connected to S22 by the method of B3, i.e., the radical node of S41 is attached to one of the radicals of S22 with no net change in the number of hydrogens. Thus the ethyl simply connects to one of the radical sites in the amino radical, which retains one radical site. The new classification code of [3]–[8] is S22, and as these fragments are surrounded by the brackets of [2] and [9], they are assigned the L-B pattern code. Table I shows that L-B cannot be connected to either L-V [10] or LMV [11], but can be connected to --V (13) and reference to Table VII reveals that an

Table VII. Connections and Classification Codes

	K10	K13	K15	K17	K20	K21	S20	S21	S22	S23	S24	S30	S32	S40	S41	S42	S43	S44	A11	A13	E11	F30	F40	F41	F50
S10	B3/K10	B3/K13	B3/K15	B3/K17	B3/K20	B3/K21	C1/S20	B3/S21	B3/S22	B3/S23	B3/S24	B3/S30	B3/S32						B3/A11	B3/A13	B3/E11	B3/F30	C1/K20		
S20	B3/K10	B3/K13	B3/K15	B3/K17	B3/K20	B3/K21	C1/S20	B3/S21	B3/S22	B3/S23	B3/S24					C2/S20			B2/A11		B2/E11				
S21	B3/K10	B3/K13	B3/K15	B3/K17	B3/K20	B3/K21	C1/S20	B3/S21	B3/S22	B3/S23	B3/S24					C2/S20			B2/A11	B3/A13	B2/E11				B2/F50
S22	B3/K10	B3/K13	B3/K15	B3/K17	B3/K20	B3/K21	C1/S20	B3/S21	B3/S22	B3/S23	B3/S24								B3/A11	B3/A13	B3/E11	B3/F30	C1/K20	C1/F40	B3/F50
S23	B3/K10	B3/K13	B3/K15	B3/K17	B3/K20	B3/K21	C1/S20	B3/S21	B3/S22	B3/S23	B3/S24								B3/A11	B3/A13	B3/E11	B3/F30	C1/K20		
S24	B3/K10	B3/K13	B3/K15	B3/K17	B3/K20	B3/K21	C1/S20	B3/S21	B3/S22			B3/S30	B3/S32						B3	B3/A13	B3/E11	B3/F30	C1/K20	C1/F40	B3/F50
S30	B3/K10	B3/K13	B3/K15	B3/K17	B3/K20	B3/K21	C1/S20	B3/S21	B3/S22	B3/S23	B3/S24	B3/S30	B3/S32						B3/A11	B3/A13	B3/E11	B3/F30	C1/K20	C1/F40	
S32																									
S40	B3/K10	B3/K13	B3/K15	B4/K17	B3/K20	B3/K21	C1/S20	B3/S21	B3/S22	B3/S23	B3/S24	B3/S30	B3/S32						B3/A11	B3/A13	B3/E11	B3/F30	C1/K20	C1/F40	B3/F50
S41	B3/K10	B3/K13	B3/K15	B3/K17	B3/K20	B3/K21	C1/S20	B3/S21	B3/S22			B3/S30	B3/S32						B3/A11	B3/A13	B3/E11	B3/F30	C1/K20	C1/F40	B3/F50
S42							C1/S20	B3/S21	B3/S22							C2/S20									B2/F50
S43	B3/K10	B3/K13	B3/K15	B3/K17	B3/K20	B3/K21	C1/S20	B3/S21	B3/S22	B3/S23	B3/S24					C2/S20			B3/A11	B3/A13	B3/E11				B2/F50
S44	B3/K10	B3/K13	B3/K15	B3/K17	B3/K20	B3/K21	C1/S20	B3/S21	B3/S22							C2/S20			B3/A11	B3/A13	B3/E11				B2/F50

2,3-dihydro-1H-isoindole - 1α-carboxylic acid

Atom No	Locant 1	Locant 2	Et	CV	HN	St	B_1		A_1		A_2	1	A_3		B_2
1	1		C	2^3	2^1		10	1	9	1	2	1			
2	2		N	2	1				1	1	3	1			
3	3		C	2	2				2	1	4	1			
4	3a		C	4					3	1	5	1.3	9	1.3	
5	4		C	3	1				4	1.3	6	1.3			
6	5		C	3	1				5	1.3	7	1.3			
7	6		C	3	1				6	1.3	8	1.3			
8	7		C	3	1				7	1.3	9	1.3			
9	7a		C	4					8	1.3	1	1	4	1.3	
10			C	4					1	1	11	2	12	1	
11			O	2					10	2					
12			O	1	1				10	1					

Figure 5. Addition of main functional groups.

A2 type of connection must be involved. The exact nature of this and all other connections is given by Table IX. Thus the radical of (2) is connected to the 3 position of the isoindole. Similarly, the L-V unit [10] is examined by using Table I and is shown also to be connected to atom 6 of the isoindole.

3.2. Algorithm for Bond Fixing. Each atom has connected valences and hydrogen numbers. In each connection table, bond values of single, double, triple, and aromatic are indicated by 1, 2, 3, and 1.3, respectively. For every atom, the following condition must be satisfied:

$$C - 0.5 < \Sigma B < C$$

where C is the connected valence and B is the bond value of the atom. For the hydrogen designated as "1H", the connected valence of atom 1 decreases by 1 and the hydrogen number increases by 1. The values of bonds between atom 1 and other atoms change from 1.3 to 1. At an atom connected to atom 1, the bond value of a cited bond automatically becomes 1 and in this case, too, the equation above must be satisfied. The value of the bond from atom 2 to a third atom must change from 1.3 to 2, and likewise, bonds from atom 3 will change.

The process of bond definition is propagated through the structure as shown in Figures 2-4.

At the points of concentration between the substituents, the skeleton, and the main functional groups, the valence of the atom concerned increases as a result of the presence of radicals or functional groups and the hydrogen count associated with that atom must decrease by the same amount. In this way, the partial connection tables corresponding to the various units are assembled to provide a complete connection table for the molecule as shown in Figures 5 and 6.

Stereodescriptors are compared with a special table to determine their meaning, which is then merged into the connection table. All unsymmetrically substituted atoms are examined for weight, and the precedence according to the Cahn-Ingold-Prelog rules[11] is determined with the help of Figure 8a and 8b. The locations of these atoms in the connection table are then rearranged according to the stereo pattern tables, R, S, E, Z, syn, and anti to give the stereospecific connection table shown in Figure 7. For example, atom number 14 has an R configuration and thus atom 2 must be entered in the B1 column (lower direction). The lower

(1R,3S)-3-[[(S)-1-
hydroxyethyl]amino]
5-chloro-
2,3-dihydro-1H-
isoindole -
1α-carboxylic acid

Atom No	Locant 1	Locant 2	Bt	CV	HN	St	B₁	A₁	A₂	1	A₃	B₂
1	1		C	3	1		10 1	9 1	2 1			
2	2		N	2	1			1 1	3 1			
3	3		C	2³	2¹			2 1	4 1			16 1
4	3a		C	4				3 1	5 1.3	9 1.3		
5	4		C	3	1			4 1.3	6 1.3			
6	5		C	2⁴	1⁰			5 1.3	7 1.3			17 1
7	6		C	3	1			6 1.3	8 1.3			
8	7		C	3	1			7 1.3	9 1.3			
9	7a		C	4				8 1.3	1 1		4 1.3	
10			C	4				1 1	11 2		12 1	
11			O	2				10 2				
12			O	1	1			10 1				
13			O	1	1			14 1				
14		1	C	2³	2¹	S		15	16 1			13 1
15		2	C	1	3			14 1				
16			N	2	1			14	3			
17			Cl	1				6				

OH
|
C=O
|
(ring structure)
17 Cl
16 NH

carboxylic acid part (rows 11, 12)
hydroxy (row 13)
ethyl (rows 14, 15)
amino (row 16)
chloro (row 17)

NH
|
H—C—OH
|
CH₃

Figure 6. Connection of substituents.

OH
|
C=O
 ⟍H
  NH
Cl
   ⟋H
  NH
|
H—C⟋OH
|
CH₃

① - ④ in the
connectivity represents
the CIP precedences

Substituents of each
asymmetric atom are
adjusted using the stereo-
chemical adjustment tables.

Atom No	Locant 1	Locant 2	Bt	CV	HN	St	B₁	A₁	A₂	A₃	B₂	C
1	1		C	3	1		10① 1	9③ 1	2① 1			H④
2	2		N	2	1			1 1	3 1			
3	3		C	3	1	S	H④	2 1	4 1		16 1	
4	3a		C	4				3 1	5 1.3	9 1.3		
5	4		C	3	1			4 1.3	6 1.3			
6	5		C	4				5 1.3	7 1.3		17 1	
7	6		C	3	1			6 1.3	8 1.3			
8	7		C	3	1			7 1.3	9 1.3			
9	7a		C	4				8 1.3	1 1	4 1.3		
10			C	4				1 1	11 2	12 1		
11			O	2				10 2				
12			O	1	1			10 1				
13			O	1	1			14 1				
14		1	C	3	1	S	13①	15③	16② 1			H④
15		2	C	1	3			14 1				
16			N	2	1			14	3			
17			Cl	1				6				

Figure 7. Addition of stereochemistry to connection table.

direction of the ring is defined by the A2 column (clockwise direction) and the A1 column (counter-clockwise direction) of the atoms in the ring. Hydrogen is entered in the B2 column (upper direction) and in this way, the *R* configuration of atom 14 is registered in the connection table.

Finally, all arrangements of asymmetrical atoms are compared with the stereo pattern tables in order to obtain the correct stereodescriptors. It is essential that this procedure be followed in the case of compounds assembled from stereoparent names.

4. STANDARDIZATION OF THE CONNECTION TABLE

4.1. Extended Morgan Algorithm. The assembled connection table is renumbered according to the extended Morgan algorithm, as described by Moreau.[8] The numbering prece-

dence is as follows: (1) degree of connectivity, (2) bond value, 3 > 2 > 1.3 > 1, (3) atomic number of the atom, (4) atomic numbers of neighboring atoms, and (5) existence of chirality.

4.2. Determination of the Basic Ring System. The connection table for the whole molecule is examined for chains, rings, or ring systems, and these are then decomposed into component rings using the algorithms developed by Choplin and Wipke.[9] The interrelationships between the component rings and any spiro, fused, or bridged rings are examined, and the positions of the latter are determined. The ring group with the greatest number of fused rings is considered to be the basic ring system, and residual portions of rings are treated as bridges. Atoms belonging to the basic ring system are registered in the connection table. The outer atoms of rings taken up in a clockwise direction are placed in the A2 column, and those encountered in the opposite direction are placed in the

a

R-O Table

R-Configuration of an
ordinary asymmetric atom

	B₁	A₁	A₂	A₃	B₂
1	1	2	3		4
2	1	3	4		2
3	1	4	2		3
4	3	1	2		4
5	2	3	1		4
6	4	2	1		3
7	3	2	4		1
8	2	4	3		1
9	4	1	3		2
10	4	3	2		1
11	2	1	4		3
12	3	4	1		2

R-J Table

R-configuration of a
ring-fused atom

	B₁	A₁	A₂	A₃	B₂
1	4	2	3	1	
2	4	1	2	3	
3	4	3	1	2	
4	3	4	2	1	
5	3	1	4	2	
6	3	2	1	4	
7	2	4	1	3	
8	2	3	4	1	
9	2	1	3	4	
10	1	4	3	2	
11	1	2	4	3	
12	1	3	2	4	

	B₁	A₁	A₂	A₃	B₂
13	1	3	2	4	
14	2	1	3	4	
15	3	2	1	4	
16	1	2	4	3	
17	4	1	2	3	
18	2	4	1	3	
19	1	4	3	2	
20	3	1	4	2	
21	4	3	1	2	
22	2	3	4	1	
23	4	2	3	1	
24	3	4	2	1	

The numbers in the Tables reflect the CIP order of precedence of the atoms

Pattern Number 1　　Pattern Number 11　　Pattern Number 13
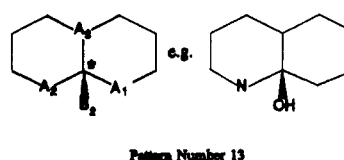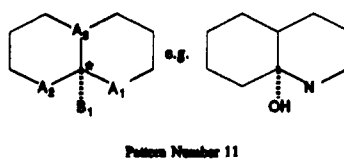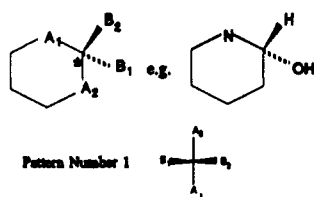
b

S-O Table

S-Configuration of an
ordinary asymmetric atom

	B₁	A₁	A₂	A₃	B₂
1	1	3	2		4
2	1	2	4		3
3	1	4	3		2
4	2	1	3		4
5	3	2	1		4
6	4	3	1		2
7	2	3	4		1
8	3	4	2		1
9	4	1	2		3
10	4	2	3		1
11	3	1	4		2
12	2	4	1		3

S-J Table

S-Configuration of an
ring-fused atom

	B₁	A₁	A₂	A₃	B₂
1	4	3	2	1	
2	4	1	3	2	
3	4	2	1	3	
4	3	2	4	1	
5	3	4	1	2	
6	3	1	2	4	
7	2	4	3	1	
8	2	1	4	3	
9	2	3	1	4	
10	1	3	4	2	
11	1	4	2	3	
12	1	2	3	4	

	B₁	A₁	A₂	A₃	B₂
13	2	3	1	4	
14	3	1	2	4	
15	1	2	3	4	
16	4	2	1	3	
17	1	4	2	3	
18	2	1	4	3	
19	3	4	1	2	
20	4	1	3	2	
21	1	3	4	2	
22	4	3	2	1	
23	2	4	3	1	
24	3	2	4	1	

The numbers in the Tables reflect the CIP order of precedence of the atoms

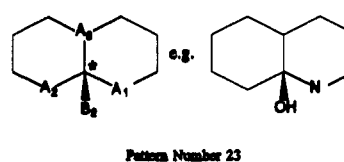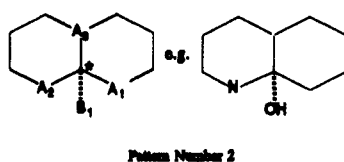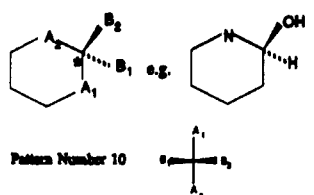Pattern Number 10　　Pattern Number 2　　Pattern Number 23

Figure 8. Stereochemistry adjustment table.

A1 column. Bridges are then placed in B1 (lower) or B2 (upper) proceeding toward the basic ring system. Chains are standardized according to essentially the same rules for the main and side chains in the IUPAC nomenclature rules and placed in the B1 or B2 columns depending upon the configuration of the atom in question.

**4.3. Tautomerism and Aromatization.** Tautomerism and aromatization are handled by the method of Chemical Abstracts Service (CAS)[10] except that aromatization is applied only to 6-membered rings. Aromatic and tautomeric bonds both have a bond value of 1.3. The stepwise processing of tautomerism and aromatization, and then tautomerism again, is conducted in such a way that the connection tables derived from superficially different systematic names will be as similar as possible. Keto–enol tautomerism in carbon skeletons devoid of heteroatoms is not examined.

Table VIII. Connections and Classification Codes

	K10	K13	K15	K17	K20	K21	S20	S21	S22	S23	S24	S30	S32	S40	S41	S42	S43	S44	A11	A13	E11	F30	F40	F41	F50
S10																									
S20							C2 S20	C2 S20								C2 S20									
S21							C2 S20	C2 S20								C2 S20									
S22																									
S23																									
S24																									
S30					B3 K20							B3 S30		B3 S40	B3 S41										
S32					B3 S30									B3 S40	B3 S41									B3 S30	
S40																									
S41																									
S42							C2 S20	C2 S20								C2 S20									
S43							C2 S20	C2 S20								C2 S20									
S44							C2 S20	C2 S20								C2 S20									

Table IX. Codes for Methods of Connection

code	method of connection
A2	connects a radical to the position indicated by the locant of a target and decrements the number of hydrogens at the position by a number equal to the valence of the radical.
	example: 1-carboxy + ethyl—the carboxy is connected to the 1 position of the ethyl
B1	connects a radical to a position with a default mark
	example: methyl + pyrazine—the pyrazine has a default 1 mark at position 2, where only one radical can be connected without a locant
B2	connects radicals to a position with a default mark of 2 until all hydrogens at this positions have been replaced
	example: trichloro + acetic acid—chlorines are added to the 2 position until no hydrogens remain there
B3	connects a radical to another special radical by replacing hydrogens not using the radical node
	example: 2,2'-methylene bis (methyl imino) bis ethanol—the methyl connects to the imino, replacing a hydrogen, but preserving the two radicals present in the imino
C1	connects a radical to a second radical with 2 or more radical values, reducing the radical value by 1
	example: acetyl→ + ←oxy→ = acetyloxy→
C2	connects a radical with more than two prioritized valencies (e.g., *1, 2, 3,* etc.) to each priority 1 radical atoms of another radical with more than two radical valencies
	example: 2,2',2''-[nitrilo tris (3-chloro-4,1-phenylene) tris ethanol—the nitrilo has three valencies with priority numbers *1, 2,* and *3.* The 4,1-phenylene has a priority *1* at position 4 and priority *2* at position 1. First, the chloro is attached to the 3 position of the 4,1-phenylene, making the radical asymmetric. Connections are then made between the *1, 2,* and *3* nitrilo radicals and the 1 position of three 4,1-phenylene radicals, producing the correct structure:
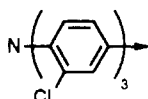
5. RESULTS

5.1. Program Steps and Processing Speed. The nomenclature translation program is written in PL/1 with some assembler and has about 140K steps. The processing speed per compound is between 0.5 and 8.0 s on a HITAC M680 computer under a VOS-3 operating system.

5.2. Capability. Nearly all types of systematic names for organic compounds of low molecular weight, including sugars, peptides, antibiotics, and other complex compounds can be successfully translated into a connection table. Some examples are shown in Figure 9.

6. JICST NOMENCLATURE RULES

These rules are based upon the IUPAC rules which have been supplemented and modified. The new rules form an input rule book for the JICST Chemical Substance Dictionary Database. Some of the modifications that were made to the IUPAC rules are described below:

1. Number concordance. Example: diacetic acid calcium salt is used instead of acetic acid calcium salt.

2. Restriction of usable locant types. Example: 2-*O*,3-*O*,4-*O*-trimethyl-α-D-glucopyranose is used instead of 2,3,4-tri-*O*-methyl-α-D-glucopyranose.

3. Description of spiro compounds using numerical figures. Example: trispiro[3.0.3.4.3.2]octadecane is only for
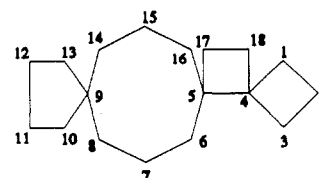
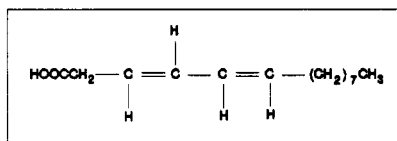trispiro[3.0.3:4.3.2]octadecane is for
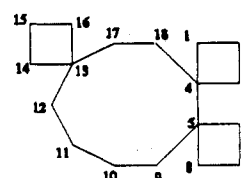
Chemical Compound Search System   DATE:87.02.04 PAGE: 28
                                          JICST COPYRIGHT

SN = J4.085a (C)
RN = 23400-52-4
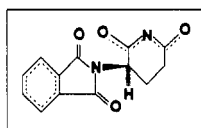Molecular Formula = C14H24O2
Molecular Weight = 224.344

Systematic Name = (3E,5Z) -3, 5- Tetradecadienoic acid
Common Name =ガトモケン
Common Name = Megatomoic acid

Chemical Compound Search System   DATE:90.03.14 PAGE: 5
                                          JICST COPYRIGHT

#0002
SN = J15.228E (C)
RN = 2614-06-4
Molecular Formula = C13H10N2O4
Molecular Weight = 258.233

Optical Rotation = +
Systematic Name = (+)-2-[(R)-2,6-Dioxo-3-piperidinyl]
                       -1H-isoindole-1,3(2H)-dione
Systematic Name = (+)-N-[(R)-2,6-Dioxo-3-piperidinyl]
                       -1H-isoindole-1,3(2H)-dione
Common Name = (+)-リンドマイド
Common Name = (+)-Thalidomide
Common Name = (+)-リンドミド
Common Name = (+)-リンドマイド
Common Name = (+)-リンドミド
Common Name = R-(+)-Thalidomide
Common Name = TI4925000

Chemical Compound Search System   DATE:90.03.28 PAGE: 1
                                          JICST COPYRIGHT

#0002
SN = J53.847G (C)
RN = 841-67-8
Molecular Formula = C13H10N2O4
Molecular Weight = 258.233
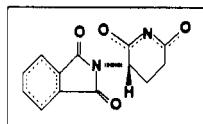
Optical Rotation = -
Systematic Name = 2-[(3S)-2,6-Dioxo-3-piperidinyl]
                       -1H-isoindole-1,3(2H)-dione
Systematic Name = (-)-N-[(S)-2,6-Dioxo-3-piperidinyl]
                       -1H-isoindole-1,3(2H)-dione
Common Name = S-(-)-Thalidomide
Common Name = TI4925050
Common Name = S-(-)-Thalidomide
Common Name = S-(-)-リンドマイド
Common Name = (+)-リンドミド

Figure 9. STARS output.

4. Peptides are described by the standard three-letter abbreviation in linear formulas such as L-Ala-L-Cys(1)-L-Met-L-Lys-Gly-L-Cys(1)-L-Phe-L-Val-L-Leu-NH$_2$. In this case, the Cys residues with the same parenthetic number "(1)" are connected together by disulfide bonds, and this connection is reflected in the connection table: N^6-(L-Ala-L-Phe-)cyclo[L-Lys-L-Met-D-Phe-L-Val-Gly-]. In this case, the peptide bond between the Gly and the N of the Lys forms a ring, and the chain of the Ala-Phe residue is connected to the N^6 atom of the Lys.

5. The stereodescriptor description does not follow the CAS method. Each stereodescriptor is inserted into the corresponding bracket level where an unsymmetrical target atom is found.

7. ONLINE SERVICES PROVIDED BY JOIS-F

7.1. Data Sources. The JICST Chemical Substance Dictionary Database contains registered chemical compounds from the following sources.

(1) Chemical literature.
(2) *Handbook of Existing and New Chemical Substances*, edited by the Ministry of Trade and Industry in accordance with the law concerning examination and regulation, manufacture, etc. of chemical substances.
(3) List of chemical substances, edited by the Ministry of Labor in accordance with the Industrial Safety and Health Law.
(4) Registry of Toxic Effects of Chemical Substances (RTECS) prepared by the U.S. Department of Labor, National Institute of Occupational Safety and Health.
(5) Chemical substances examined and entered into factual databases and registered in the chemical dictionary database.
(6) Major, important directories and handbooks.

At the end of September 1990, the number of registered substances in the database was more than 300 000.[7]

7.2. Online Service. Since January 11, 1988, the JICST Chemical Substance Dictionary Database, in conjunction with the factual database of TH (Thermochemistry, compiled by JICST) and MS (Mass Spectral, issued by both the U.S. NIST and the Japan Society for Mass Spectrometry) have been available via the JOIS-F Online Service.[3] The following search capabilities are available:

(1) SN (JICST Substance Number).
(2) Molecular Formula (total and partial).

(3) Systematic name (full name and fragments, left or right truncated).
(4) Trivial, common, and commercial names and various code numbers, experimental substance numbers, and so on.
(5) Chemical structures (total, partial, or Markush type by drawing structural diagrams or entering structural codes. Stereospecific searching is possible).
(6) Other features (peptides can be sought with three-letter amino acid codes and with structures. Specific sequences of amino acids can also be found by means of a secondary iterative search.)

7.3. Output. Alphanumeric strings of the dictionary and structural diagrams are output to TTY-type terminals using hyphens, equal signs, and dots. High-quality structure diagrams may be obtained with graphics terminals. Some examples of search and output are shown in the figures.

REFERENCES AND NOTES

(1) Fujiwara, Y.; Araki, K.; Kaji, M.; Maeda, C.; Kajiwara, Y.; Nanba, M.; Nagai, Y.; Takahashi, K.; Karasawa, Y.; Obara, Y. *Proceedings of the 21st Research Meeting of Information Sciences & Technology of Japan*; 1984; pp 75–83.
(2) Petrarca, A. E.; Lynch, M. F.; Rush, J. E. *J. Chem. Doc.* **1967**, *7*, 154–165.
(3) Fujiwara, Y.; Komuro, S.; Souma, T.; Suzuki, K.; Hasegawa, N.; Hinatsu, K.; Nishiyama, H. *Proceedings of the 1984 International Chemical Congress on Pacific Basin Society*; 1984; pp 61–74.
(4) Vander Stouw, G. G.; Naznitsky, I.; Rush, J. E. *J. Chem. Doc.* **1974**, *14*, 185–193.
(5) Dittmar, P. G.; Stobaugh, R. E.; Watson, C. E. *J. Chem. Inf. Comput. Sci.* **1976**, *16*, 111–121.
(6) Uchino, H.; Araki, K. *Proceedings of the 14th Research Meeting of Information Sciences & Technology of Japan*; 1977; pp 101–121.
(7) Araki, K. European Patent 090895, Jan 18, 1989.
(8) Moreau, G. *Nouv. J. Chim.* **1980**, *4*, 17–22.
(9) Choplin, F.; Wipke, W. T. *J. Chem. Inf. Comput. Sci.* **1978**, *18*, 110–119.
(10) Mockus, J.; Stobaugh, R. E. *J. Chem. Inf. Comput. Sci.* **1980**, *20*, 18–22.
(11) Cahn, R. S.; Ingold, C.; Prelog, V. *Angew. Chem., Int. Ed. Engl.* **1966**, *5*, 385–415.