C A E S A

COMPUTER AIDED ESTIMATION OF SYNTHETIC ACCESSIBILITY

CAESA version 2.4, July 2003.

Contents:

1. Introduction

CAESA is a program designed for the Computer-Aided Estimation of Synthetic Accessibility. CAESA aims to score a target compound on the estimated difficulty of its synthesis. CAESA performs a retrosynthetic analysis on the compound to find appropriate available starting materials. CAESA then analyses sections of the compound not covered by the starting materials and the overall complexity of the molecule to obtain a numerical estimate of synthetic accessibility.

1.1 Performance Objectives

For CAESA to be useful commercially, it is necessary for the results produced to be realistic and reproducible. The retrosynthetic analysis performed should match (or exceed) that of an experienced synthetic chemist:

- the lowest possible number of synthetic steps
- consideration of reaction difficulty and/or yields
- sensible starting materials selected
- minimal number of FGIs (Functional Group Interconversions)

1.2 Performance of CAESA

CAESA considers the reaction difficulty and/or yields by allocating each reaction or transform in the databases a score, with a higher number indicating a more difficult or a lower yielding reaction. For example a simple FGI like a halide conversion to an alcohol would score 1. A more complicated, specific reaction like the *Pauson-Khand* reaction would score 4. The synthetic distance is input by the user to begin the analysis, which can limit the total difficulty or number of steps allowed. CAESA selects starting materials based on their total coverage of the target compound, so that the starting material which covers the most area will be the most favoured.

A number of validation tests were carried out with CAESA, one of them comparing the CAESA answers with the answers of a group of expert synthetic chemists at a known large pharmaceutical company (see more details in Section 4). The overall conclusion was that CAESA is matching the answers of the experts in most of the cases, thus CAESA is a useful tool for organic synthesis planning and retrosynthetic analysis in the drug discovery process.

2. How CAESA works

CAESA works using a rule based expert system. A number of knowledge bases are employed with different functions in order to identify starting materials and estimate the synthetic accessibility.

CAESA employs a two-directional retrosynthetic analysis. Synthetic transformations are applied to all of the available starting materials in the starting materials database. This creates a database of potential intermediates. Retrosynthesis is then applied to the target compound until either a starting material, or a structure from the potential intermediates database is matched. This approach speeds up the execution of CAESA as only the retrosynthetic analysis is performed at the time of the search, the potential intermediates are stored with the program files.

The synthetic knowledge base contains a list of transforms which are applied to the starting materials to create intermediates. For example alkene saturations, alcohol oxidations and ozonolysis reactions. These types of transformations can be difficult to apply retrosynthetically, for instance if alkene saturation was in the retrosynthetic knowledge bases then every time a "CH-CH" fragment appeared this would be converted to an alkene "C=C". This would mean an increase in the number of matching transforms to each target and so an increase in the time to perform the retrosynthetic analysis.

There are three retrosynthetic knowledge bases, which contain transforms which are applied to the target compound. The first retrosynthetic knowledge base contains strategic disconnections applied to the whole target structure. These would be simple, versatile disconnections such as amide or ester disconnections. These aim to cut up a large target

molecule into several manageable segments. The second retrosynthetic knowledge base contains general disconnections, which can be less versatile than the first. This is the largest of the three knowledge bases and is applied most of the time. The third retrosynthetic knowledge base contains simple FGIs, and is applied to small intermediates where no further disconnections are necessary. Some transforms, such as key FGIs or fundamental disconnections, can appear in more than one or even all three knowledge bases.

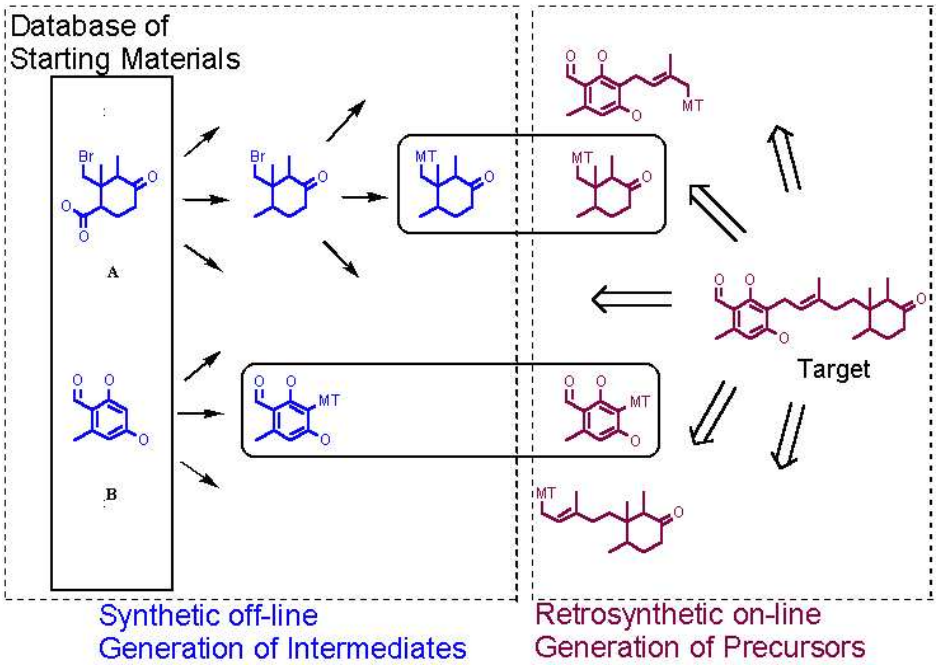

Figure 1: Bidirectional search for synthetic routes

3. Organisation of Knowledge Bases

The CAESA program uses knowledge bases of reactions and transforms. The reactions are used to generate synthetic intermediates from the starting material databases. The knowledge base of transforms generates precursors retrosynthetically from the target structure.

The reaction and transform knowledge bases are not a comprehensive analysis of reactions in synthetic chemistry. They contain generalised reactions and transforms and the structures produced by their operation may also contain generic atoms and bonds. This use of generalised reaction sequences and generic structures reduces the processing time required to search the databases. The number and difficulty of the reaction steps can be controlled, using estimates of synthetic ease associated with each rule.

The reactions in the synthetic knowledge base and transforms in the retrosynthetic knowledge bases are written in a pattern language called: PATRAN, which is an ASCII (human readable) pattern language designed specifically for reactions. It is similar to the SMILES notation, but extended to handle reactions not only individual molecules. PATRAN also allows usage of generic atom types. Examples of transforms written in PATRAN can be seen in Appendix 3.

Some important changes that were introduced recently to the knowledge bases that make the maintenance of the knowledge bases more user friendly. The EXPLANATION line of the RULE in the PATRAN string is a definition of the reaction type and appears above the reaction arrow on the display. To organise the retrosynthetic knowledge bases the following steps were carried out:

- reading and compiling list of the transforms in each existing database
- correcting the explanation lines for each one
- creating one new list with all the transforms organised into sections (for example transforms involving nitrogen groups or transforms forming heterocycles)
- allocating each transform from the list into one of three new lists, which would form the new retrosynthetic knowledge bases (based on reaction type and priority)
- creating three new retrosynthetic knowledge bases from the new transforms list, using the original transforms
- changing the chemical-label definitions (so they described the functional group in the pattern)
- compiling a list of all chemical-labels (to prevent duplication of a label with different structures and to enable a chemical-label to be reused later in a database)
- creating headings and sub-heading for different sections
- naming each transform in the knowledge base (for clarity)
- numbering all reactions in the knowledge bases and the transform lists (to enable them to be located quickly)
- adding literature references to some of the transforms in the system, i.e. about 20% of the reactions currently have associated literature references (the rest of the reactions is very common, so no single literature reference can be identified for them)

The organisation of the synthetic knowledge base was carried out in the same way. The lists of transforms appearing in each knowledge base is shown in Appendix 4. An example of chemical-labels used is in Appendix 5.

4. Validation of CAESA version 2.4

75 test structures were used to validate CAESA version 2.4. A spreadsheet comparing the scores given by chemists and generated by CAESA for the 75 test compounds is shown in Appendix 1. Speed test results are given in Appendix 2. Out of the 75 target structures, 65 are showing the correct retrosynthetic schemes and starting materials, agreeing with or exceeding the experienced organic chemists. Of the 10 unsatisfactory cases, seven contain fused heterocycles hence could be improved with addition of further transforms in this area. Two involve saturated heterocycles, an area which too is lacking. Another one example contains phosphate groups, which is also an area with hardly any transforms.

CAESA is currently a useful tool for performing retrosynthetic analysis of a target compound. It can produce a list of available starting materials which a synthetic organic chemist may or may not be aware of and suggestions for synthetic routes. The results produced will always need to be coupled with experience and chemical knowledge to produce the best solutions and there will always be limitations to CAESA's use. Symmetry, stereochemistry, regiochemistry and selectivity are often not sensed by CAESA,

thus human input is still necessary.

5. Appendices

Appendix 1: Validation Results

#	Structure	Chemists	CAESA	Comments
1	Etanid	76.3	100	ok
2	TS 19	97.5	100	ok
3	wal2014	91.3	99	ok
4	TS 73	97	99	ok
5	TS 74	99	99	ok
6	aspal	72.5	98.5	ok
7	TS 33	82.5	95.1	ok
8	itopride	91.3	93.8	ok
9	ak2123	81.3	93	ok
10	lemin	73.8	92.9	ok
11	pantoprazole	90	91.3	ok
12	TS 18	76.3	90.7	ok (phthalimide)
13	TS 17	75	89.8	ok (phthalimide)
14	TS 20	85	89.2	ok
15	remif	73.8	89.1	ok
16	TS 22	92.5	88.5	ok
17	TS 9	93.8	88.2	ok
18	sdzfox	71.3	86.6	ok
19	atevir	83.8	86.3	ok
20	letroz	86.3	84.7	ok
21	Lesop	86.3	84	ok
22	TS 16	73.8	83.4	ok
23	ae0047	68.8	83.1	ok
24	bm210955	66.3	79.9	PHOSPHO
25	eliprod	80	79.5	ok
26	cystem	57.5	79.2	ok
27	TS 21	77.5	77.7	ok
28	myco	52.5	76.5	ok
29	TS 23	86.3	74.9	FUSED HET
30	delavir	75	73.9	ok
31	TS 35	60	73	ok
32	TS 24	78.8	72.5	ok

33	Vorozole	75	72.4	ok
34	fk565	75	71	ok
35	TS 7	71.3	69.6	ok
36	TS 15	68.8	69.5	ok
37	epirop	62.5	67.7	ok
38	Zatos	55	67.4	ok
39	TS 32	81.3	67.2	ok
40	zd1033	70	65.5	ok
41	Tomud	72.5	63.7	ok
42	rs21607	58.8	63.3	ok
43	TS 6	81.3	63	ok
44	TS 10	50	60.8	ok
45	c1316	62.5	58.2	ok
46	ly246	38.8	58.1	ok
47	sm8849	66.3	57.9	ok
48	terflav	60	56.1	ok
49	TS 4	72.5	54.2	ok
50	TS 37	65	53.4	ok
51	fk037	53.8	52.3	ok
52	zipras	70	51.4	FUSED HET
53	ksg504	62.5	50.3	ok
54	ari509	58.8	49	ok
55	n3389	58.8	48.3	FUSED HET
56	TS 14	58.8	46.2	ok
57	er30346	50	45.8	ok
58	e09	43.8	43.6	ok
59	abt719	51.3	42.4	FUSED HET
60	ifetro	51.3	41.6	ok
61	snk860	56.3	39.5	FUSED HET
62	Vamic	45	39.4	ok
63	TS 26	50	38.7	ok
64	TS 29	41.3	38.4	ok
65	s1090	38.8	32.1	SATURATED HET
66	sch51048	52.5	32	ok
67	du6859	37.5	29.5	FUSED HET
68	agm	20	28.7	Ok
69	TS 11	45	27.9	FUSED HET

70	artef	15	24.4	ok
71	TS 27	21.3	22.1	ok
72	bn2727	36.3	20.9	SATURATED HET
73	er35786	36.3	19.6	ok
74	pimil	35	17.3	ok
75	TS 13	17.5	15.5	ok

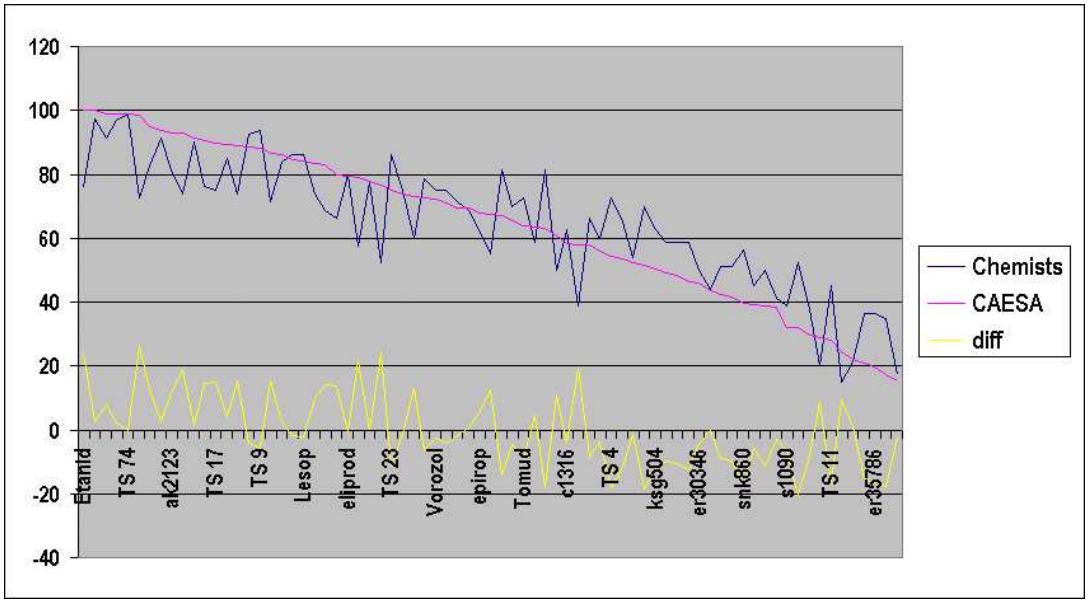

Figure 2: Comparison of the synthetic accessibility estimates generated by CAESA and a group of expert synthetic chemist at a known large pharmaceutical company (results are sorted by CAESA generated estimates in descending order)

Appendix 2: Speed test results of CAESA v2.4

HW	# of Structures	Total Processing time (s)*	Processing Time (s) / structure
Intel(R) Pentium(R) 4 CPU 1.50GHz	75 (validation set)	373.50	4.98
	300 (from maybridge DB)	1191.68	3.97
	448 (from maybridge DB)	1622.57	3.62
Intel(R) Pentium(R) 4 CPU 2.40GHz	75 (validation set)	210.75	2.81
	300 (from maybridge DB)	520.35	1.73
	448 (from maybridge DB)	870.15	1.94

Note*: Three processing parameters influence the processing times: (1) the amount of starting materials that the system has to work with, (2) the amount of transforms in the knowledge-bases of the system and (3) the number of steps considered, i.e. synthetic distance. The above speed results were generated with the standard knowledge-bases of CAESA, at a synthetic distance of three and with using the entire catalog of organic compounds from the following suppliers:

1. Acros catalog from Jan 2003
2. Aldrich catalog from Jan 2003
3. Lancaster catalog from Jan 2003

Appendix 3: Examples from Retrosynthetic Knowledge Bases

```
N) S U L F U R G R O U P S
**************************

nn) DISULFIDE FORMATION
```

```
CHEMICAL-LABEL <Disulfide>
...STARTP
...C[HETS=1]-S[CONNECTIONS=2]-S[CONNECTIONS=2]-C[HETS=1]
...ENDP
RULE
EXPLANATION Disulfide Formation.
IF Disulfide
THEN assign substitution priority 3
disconnect-bond 2
generalise-atom 2 as SG
generalise-atom 3 as SG
END-THEN

nn) THIOKETONE FORMATION
CHEMICAL-LABEL <Thioketone>
...STARTP
...C[HETS=1]=S[CONNECTIONS=1]
...ENDP
RULE
EXPLANATION Thioketone Formation from Carbonyl and Hydrogen Sulfide.
IF Thioketone
THEN assign substitution priority 2
substitute-atom 2 with O
END-THEN

nn) THIOESTER FORMATION
CHEMICAL-LABEL <Thioester>
...STARTP
...C-C[HETS=2];[ARYL=NO](=O)-S[CONNECTIONS=2]-C[HETS=1]
...ENDP
RULE
EXPLANATION Thioester Formation.
IF Thioester
THEN assign substitution priority 2
disconnect-bond 3
generalise-atom 4 as SG
add-bond to 2 of - HL
END-THEN
```

Appendix 4: Transform Lists for CAESA version 2.4 Knowledge Bases

Appendix 4.1: Synthetic Database

A. Functional group Interconversions

1. Amide Formation from Cyanide
2. Aldehyde Formation from Cyanide
3. Oxidation of Hydrazone to Diazo
4. Azo Disconnection
5. Isocyanate Formation
6. Isothiocyanate Formation
7. Keto/Enol Tautomerism
8. Carbonyl Generalisation
9. Hydrazone from Carbonyl
10. Wittig Methylenation
11. Acid Halide conversion to Acid
12. Acid Halide conversion to Amide
13. Acid Halide conversion to Aldehyde
14. Conversion of Alkyl Halide to Nitrile
15. Nitrile Formation from Amide
16. Alkyl Halide Formation from Tertiary Alcohol
17. Alkyl Halide Formation from Secondary Alcohol
18. Alkyl Halide Formation from Primary Alcohol
19. Acyl Halide from Carboxylic Acid
20. Halide formation by Decarboxylation of Acid
21. Arndt-Eistert (x2)
22. Oxidation of Aromatic-Methyl Substituents to Acids

B. Reductions/Saturations

23. Reduction of Cyanide to Amine
24. Imine Reduction
25. Hydrazone Reduction
26. Reduction of Alkyne to Alkene
27. Oxygen Group Reduction
28. Carboxylic Acid Reduction
29. Halogen Reduction
30. Sulphur Reduction
31. Olefin Saturation
32. 5M-Aromatic Ring Saturation
33. 6M-Aromatic Ring Saturation

C. New Name Transforms

34. Clemmenson Reduction
35. Curtius Rearrangement
36. Sandmeyer Reaction (x4)
37. Nef Reaction

D. General Synthetic Disconnections

38. Periodate Oxidation (x5)
39. Ester Hydrolysis
40. Hydrolysis of Acetal (x2)
41. Dithioacetal Disconnection
42. Ozonolysis (x3)

E. Organometallic Preparations

43. Grignard/Organolithium formation from halides (x3)
44. Metallation Alpha to Heteroatom (x2)
45. Metallation of Terminal Alkyne
46. Hydrometallation of Terminal Alkyne
47. Hydroboration of Alkene
48. Shapiro Reaction
49. Metallation of Aldehyde

Appendix 4.2: Retrosynthetic Database 1

A. 1,1 DISCONNECTIONS

1. ACETAL FORMATION
2. HEMI-ACETAL FORMATION

B. 1,3 DISCONNECTIONS

3. ALDOL CONDENSATION
4. CLAISEN ESTER CONDENSATION
5. MANNICH REACTION

C. 1,5 DISCONNECTIONS

6. MICHAEL REACTION

D. OXYGENGROUPS

7. ESTERIFICATION
8. ETHER FORMATION
20. ISOCYANATE FORMATION FROM PRIMARY
21. UREA FORMATION FROM ISOCYANATE
22. CARBAMATE FORMATION FROM ISOCYANATE
23. OXIME FORMATION
24. HYDRAZONE FORMATION
25. HYDRAZINE ACYLATION

F. SULPHURTRANSFORMATIONS

26. THIOETHER FORMATION
27. THIOESTER FORMATION
28. THIOAMIDE FORMATION
29. SULFONAMIDE FORMATION
30. SULFONAMIDE ALKYLATION

9. ENOL ETHER FORMATION
10. ENOL ESTER FORMATION

E. N I T R O G E N G R O U P S

11. AMINE FORMATION
12. PRIMARY AMIDE FORMATION
13. SECONDARY AMIDE FORMATION
14. TERTIARY AMIDE FORMATION
15. IMINE FORMATION
16. ENAMINE FORMATION
17. ENAMIDE FORMATION
18. N-ALKYLATION
19. REDUCTIVE AMINATION

G. G E N E R A L

31. FRIEDEL CRAFTS ACYLATION
32. GRIGNARD REACTION
33. WITTIG REACTION
34. HECK REACTION
35. KETO/ENOL TAUTOMERISM
36. IMINE/ENAMINE TAUTOMERISM
37. HALIDE FORMATION
38. NITROBENZENE REDUCTION

Appendix 4.3: Retrosynthetic Database 2

Strategic Disconnections

A. Cycloadditions

1. Diels-Alder (x 4)
2. Ene Cycloaddition (x2)
3. Photochemical 2+2 Cycloadditions (x4)
4. Thermal 2+2 Cycloadditions (x2)
5. 1,3-Dipolar Additions (x2)

B. 1, 2 Difunctionalised Compounds

6. Dihydroxylation of Alkenes
7. Acyloin Condensation
8. Formation and hydrolysis of a cyanohydrin

C. 1,3 Difunctionalised Compounds

9. Aldol Condensation
10. Aldol Dehydration
11. Aldol Condensation of Enolates (x2)
12. Claisen Condensation
13. Crossed Ketone-Ester Claisen Condensation
14. Mannich Reaction
15. 1,3-keto-ester formation
16. Alkylation of 1,3-keto-esters
17. Alkylation next to Phosphorus

D. 1,4 Difunctionalised Compounds

18. Conjugate addition of HCN (x2)
19. 1.4 Conjugate addition of amines (x2)
20. Organocopper Reactions
21. 1,4-Conjugate addition of phosphorus

E. 1, 5 Difunctionalised Compounds

22. Michael Reaction (x10)

Functional Group Interconversions

F. Aromatics

23. Friedel-Crafts Acylation
24. Friedel-Crafts Alkylation
25. Arene Diazonium Salt Formation
26. Nitrobenzene Reduction
27. Aromatic Addition/Eliminations

G. Alkenes

28. Peterson Olefination
29. McMurry Olefination
30. Wittig Reaction
31. Horner-Wadsworth-Emmons(x2)
32. Formation of cis-alkene from terminal alkyne
33. Formation of trans-alkene from terminal alkyne
34. Halogenation of terminal alkynes
35. Dehydrohalogenation
36. Dehydration of an Alcohol
37. Hydrogenation of an alkene next to Carbonyl

H. Alkynes

38. Alkylation of a Terminal Alkyne

I. Alkyl Halides

39. Halide Formation (x2)
40. Dihalide Formation from an alkene
41. Halohydrin formation form an Alkene
42. Halo-imine formation (x2)
43. 1,4-halo-imine formation
44. Enamine Halogenation

J. Alcohols

45. Alcohol formation (x5)
46. Primary alcohol formation (x2)
47. Tertiary alcohol formation (x2)
48. Alkyl Ether formation
49. Aryl Ether Formation (x6)
50. Enol Ether formation
51. Enol Ester formation

K. Carbonyls

52. Esterification
53. Carbonyl Formation
54. Ketonization of an enol
55. Formation of carboxylic acid (x5)
56. Formation of an Acid Halide
57. Anhydride Formation
58. Iodolactonisation
59. Alpha Halogenation

L. Epoxides

60. Epoxide formation (x2)
61. Nucleophilic Opening of an Epoxide (x6)
62. Opening of a Cyclic Epoxide
63. Halo-epoxide Alkylation

M. Nitrogen Groups

64. Primary Amide Formation
65. Secondary Amide Formation
66. Tertiary Amide Formation
67. Imine Formation
68. Enamine Formation
69. Enamide Formation
70. N-Alkylation (x6)
71. Aromatic N-Alkylation
72. Secondary Amine from Aryl Halide (x6)
73. Amine Formation (x3)
74. Imide formation (x3)
75. Reductive Amination forming Secondary Amine
76. Reductive Amination forming Tertiary Amine
77. Reductive Amination via N-Nitroso Intermediate
78. Reductive Amination of Aromatic Amine
79. Isocyanate Formation
80. Urea formation from Isocyanate
81. Carbamate formation from Isocyanate
82. Oxime formation
83. Hydrazone formation
84. Azo formation from primary amines
85. Diazotisation
86. Nitrile Formation
87. Nitrone Formation
88. Nitrile Oxide Formation
89. N-Oxide Formation
90. Imine/Enamine Tautomerism

N. Sulfur Groups

91. Aryl C-S Formation (x8)
92. Alkyl C-S Formation
93. Acyl Sulfonamide Formation (x2)
94. Disulphide Formation
95. Thioketone Formation
96. Thioester Formation
97. Thioamide Formation
98. Thiocyanate Formation
99. Sulfonamide Formation
100. Dithiane Formation
101. Isothiocyanate Formation
102. Thiourea Formation
103. Thiourethane Formation
104. Sulfoxide Oxidation to Sulfone
105. Sulfide Oxidation to Sulfone
106. Sulfide Oxidation to Sulfoxide
107. Sulfonamide Alkylation

Metal Catalysed Reactions

O. Magnesium
108. Grignard

P. Palladium
109. Heck Reaction
110. Palladium Cross-Coupling
111. Palladium Cross-coupling with carbonylation
112. Buchwald Chemistry
113. Palladium Catalysed Aryl and Allene Coupling
114. Palladium Catalysed Allylic Amination

Q. General
115. Organometallic 1,4-additions (x3)
116. Organometallic addition to acyl halide

Heterocycles

R. Pyridines
117. Hantzsch Pyridine Synthesis (x6)
118. Pyridine Reduction (x2)
119. Chichibabin Reaction (x3)
120. Nucleophilic Addn/Elim (x2)

S. Quinolines and Isoquinoloines
121. Quinoline Synthesis (x3)
122. Isoquinoline Synthesis (x3)
123. Pictet-Spengler Annelation

T. Quinones & Quinolones
124. Quinolone Formation
125. Quinone Formation (x3)

U. Diazines and Triazines
126. Pyridazine Formation
127. Pyrimidine Formation (x3)
128. Guanidine Formation
129. Pyrazine Formation (x3)
130. Triazine Formation

V. Pyrroles,Furans&Thiophenes
131. Pyrrole Formation (x3)
132. Furan Formation (x2)
133. Thiophene Formation
134. Mannich reaction (x3)
135. Cyanide from Amine

W. Indoles
136. Indole Synthesis (x2)
137. Coumarin Synthesis
138. Chromone Synthesis

X. Azoles
139. Imidazole Formation (x2)
140. Fused Thiazole Formation (x2)
141. Thiazole Formation (x2)
142. Oxazole Formation (x2)
143. Isooxazole Formation
144. Pyrazole Formation
145. Isothiazole Formation
146. Oxadiazine Formation
147. Amidoxime Formation (x2)
148. Triazole Formation
149. Tetrazole Formation
150. S,O,N-C=N Disconnection
151. Pthalimide Formation

Miscellaneous

Y. Other Named Reactions
152. Arndt-Eistert
153. Beckmann Rearrangement
154. Birch Reaction (x2)
155. Claisen Rearrangement (x2)
156. Clemmensen Reduction (x2)
157. Diekmann Condensation
158. Favorskii Rearrangement
159. Gatterman Formulation
160. Hoffmann-Loffler-Freytag Reaction
161. Jacobsen Asymmetric Epoxidation
162. Nazarov Cyclisation
163. Pauson Khand Reaction
164. Payne Rearangement
165. Perkin Reaction
166. Reformatsky Reaction
167. Reimer-Tiemann
168. Robinson Annelation
169. Rosenmund Reduction
170. Sandmeyer Reaction (x3)
171. Sharpless Asymmetric Epoxidation
172. Sharpless Asymmetric Aminohydroxylation
173. Simmons-Smith Reaction
174. Stephen Reduction
175. Strecker Synthesis of Amino Acids
176. Stobb Condensation
177. Swern Oxidation
178. Thorpe Reaction
179. Vilsmaier-Haack Reaction
180. Wacker Process

Z. Functional Group Activations
181. Hydrogenolysis of Benzyl Ether
182. Hydrogenation of Alkene next to Carbonyl

Appendix 4.4: Retrosynthetic Database 3

A. Aromatic Substitutions
1. Friedel Crafts Alkylation
2. Friedel Crafts Acylation
3. Nitration
4. Halogenation
5. Sulfonation
6. Nitrobenzene Reduction
7. Formation of arenediazonium salt
8. Sandmeyer Reaction (x3)

B. Alkanes
9. Alkane formation (x2)
10. Reduction of ketone/aldehyde to alkane

C. Alkenes
11. Conversion of an Olefin to a pair of Stereocentres (x3)
12. Halogenation of Terminal Alkyne
13. Dehydrohalogenation to form Alkene
14. Dehydration of Alcohol to form Alkene
15. Formation of an Alkyne from an Alkene

D. Alkyl Halides
16. Halide Formation (x2)
27. Formation of a Carboxylic Acid (x5)
28. Formation of an Acid Halide
29. Epoxide Formation (x2)
30. Alpha Halogenation
31. Enol Ether Formation
32. Enol Ester Formation
33. Carbonyl Formation (x2)
34. Rosenmund Reduction

F. Nitrogen Groups
35. Amine Formation (x3)
36. Primary Amide Formation
37. Secondary Amide Formation
38. Tertiary Amide Formation
39. Oxime formation
40. Hydrazone formation
41. Azo formation from Primary Amines
42. Nitrile Formation
43. Quaternization

G. Sulfur Group Transformation
44. Thioester Formation
45. Thioamide Formation

17. Dihalide formation from alkene
18. Halohydrin formation from alkene
19. Halo-imine formation (x2)
20. 1,4-halo-imine formation
21. Enamine Halogenation

E. Oxygen Compounds

22. Alcohol Formation (x5)
23. Primary Alcohol Formation (x2)
24. Esterification
25. Anhydride Formation
26. Phthalimide Formation (x2)
46. Thioketone Formation
47. Sulfonamide Formation
48. Dithiane Formation
49. Thiourethane Formation
50. Sulfone Oxidation to Sulfoxide
51. Sulfide Oxidation to Sulfoxide
52. Sulfide Oxidation to Sulfone

H. Phosphorus

54. Phospho Group Formation

I. General

55. Keto/Enol Tautomerism
56. Imine/Enamine Tautomerism

Appendix 5: Examples of Chemical Labels used in CAESA version 2.4

1,3-diketone	C-O-C(=O)-C-C(=O)-C
1,3-keto-ester	C-O-C(=O)-C-C(=O)-C
1,3-keto-ester 2	OG-C(=O)-C-C(=O)-C
Cyclic 1,3-keto-ester	O=C[RINGS=YES]-C[SAMERING=2]-C(=O)-OG
1,3-keto-nitrile	C(=O)-C[HS>0]-C-C#N
1,3-keto-amine	C[HETS=1](=O)-C-C-N[HS=0,1,2]
1,3-keto-secondary-amine	C(=O)-C[HS>0]-C-N[HS=0]
1,3-keto-tertiary-amine	C(=O)-C[HS>0]-C-N[HS=0]
1,1-hydroxy-acid	OG-C[HS=0,1]-C(=O)-OG
1,2-hydroxy-ketone	C[HETS=1](=O)-C[HETS=1]-(O[HS=1])
1,2-hydroxy-ketone 2	C-C(=O)-C[HS=1](-OG)-C
1,2-keto-nitrile	C[RINGS=YES](=O)-C[SAMERING=1];[HS=1]-C#N
1,3-hydroxy-ketone	C[HETS=1](=O)-C-C[HETS=1](-OG)-C
1,3-hydroxy-acid	OG-C-C-C(=O)-OG
Cyclic 1,3-hydroxy-ketone	C[RINGS=YES](-OG)-[RINGS=YES]C-[RINGS=NO]C=O[RINGS=NO]
1,4-keto-acid	C-C(=O)-C[HS=1](-C)-C(-C)(-C)-C(=O)-O
1,4-diester	C=C(-C(=O)-OG)-C[HS=2]-C(=O)-OG
1,4-dicarbonyl	C[HS=0]-C(=O)-[RINGS=NO]C[HS>0]-C(-C=,#O,N)-C=,#O,N
1,5-diketone	C(=O)-C[HS>0]-C-C-C(=O)
1,5-diketone 2	C(=O)-C[HS=2]-C[HS=2]-C[HS=1](-C=O)-C=O
1,5-diketone 3	C(=O)-C[HS=2]-C[HS=2]-C[HS=1](-C=O)-C#N
1,5-keto-nitro	C(=O)-C[HS=2]-C[HS=2]-C[HS=1](-C)-N(=O)-O
1,5-keto-nitro 2	O=N(-O)-C[HS=2]-C[HS=2]-C[HS=1](-C=O)-C=O
1,5-keto-nitrile	C(#N)-C[HS=2]-C[HS=2]-C[HS=1](-C=O)-C=O
1,5-dinitrile	C(#N)-C[HS=2]-C[HS=2]-C[HS=1](-C=O)-C#N
1,5-dinitro	O=N(-O)-C[HS=2]-C[HS=2]-C[HS=1](-C)-N(=O)-O
1,5-nitro-nitrile	C(#N)-C[HS=2]-C[HS=2]-C[HS=1](-C)-N(=O)-O
1,5-nitro-nitrile 2	O=N(-O)-C[HS=2]-C[HS=2]-C[HS=1](-C=O)-C#N
1,6-dicarbonyl	C[RINGS=NO](=O)-C[RINGS=NO]-C-C-C[RINGS=NO]-C[RINGS=NO](=O)
General Acid	OG-C[HETS=2]=O

Extended Acid
Extended Acid 2
Extended Acid 3
Unsaturated Acid
Alkylated Acid
Alcohol
Primary Alcohol
Secondary Alcohol
Tertiary Alcohol
Tertiary Alcohol 2
Alcohol 2
General Alcohol
Extended Alcohol
Extended Alcohol 2