Chemical Abstracts Service Chemical Registry System. 11. Substance-Related Statistics: Update and Additions

ROBERT E. STOBAUGH

Chemical Abstracts Service, P.O. Box 3012, Columbus, Ohio 43210

Received May 18, 1988

Statistics are updated for types of substances, ring systems, and elemental composition that have been determined for the Chemical Abstracts Service Registry Structure File at different points in time. This paper reports the updated figures and in addition some new statistics and offers some comparisons to show various shifts in file characteristics.

INTRODUCTION

The Chemical Abstracts Service (CAS) Chemical Registry System is a computer-based system that uniquely identifies chemical substances on the basis of their molecular structure. The design, content, functions, statistics, special features, and input structure conventions have been described in detail in previous papers.[1-10]

The computer-readable structure records that make up the basis of the CAS Chemical Registry System are basically records of the atoms and bonds present in the molecular structure of the substances. They represent the ring systems that are present, the substituents attached to the rings, and any substituents that link two or more rings. From these structure records, statistics can be obtained for analyses of elemental composition and ring characteristics. The statistics were first published in 1980[6] and in this paper are updated as of 1987. Some new statistics are also included concerning the occurrence of various elements in chemical substances.

The tables in this paper present a comparison of cumulative occurrence data concerning ring graphs for the years 1974, 1978, and 1987 and ring systems for 1974, 1976, 1978, and 1987. Also compared are the cumulative occurrence data for elemental composition for the years 1967, 1974, 1979, and 1987. Tables report the percentage increase from 1979 to 1987 for the occurrence of elements and for substances containing the given elements. Some statistics are provided for the percentage increase for ring graphs and ring systems. New statistics are provided for the frequency of occurrence of various elements in the presence of the halogens fluorine, chlorine, bromine, and iodine.

TYPES OF SUBSTANCES

The numbers of substances for several classes of chemical substances as of June 1987 are provided in Table I. These are for machine-registered substances and do not include those registered by manual techniques.[8] The classes listed are mutually exclusive and are those for which statistics are routinely obtained during operation of the CAS Chemical Registry System. The percentage distribution of these classes is very similar to that found for the same classes as of December 1978.[6] The numbers themselves are larger for all classes, considerably so in most cases.

STRUCTURAL CHARACTERISTICS

Chemical substances are recorded in the CAS Chemical Registry System in terms of components,[6] a component being a set of contiguous atoms. For some substances more than one component or one component plus additional information [for example, a homopolymer like the homopolymer of 2-propenoic acid $(C_3H_4O_2)_x$] is necessary for adequate representation, for example, salts of acids and of bases, polymers, mixtures, and others. The combination or augmentation of

Table I. CAS Chemical Registry Coverage (June 1987)

machine registered types of substances	no.	%
fully defined substances	6 794 065	84.1
incompletely defined substances	116 087	1.4
polymers	289 898	3.6
coordination compounds	587 775	7.3
alloys	190 800	2.4
mixtures	17 306	0.2
minerals	1 712	0.02
radical ions	14 351	0.18
ring parents	68 560	0.8
total	8 080 554	100.00

Table II. Expression Statistics (June 1987)

expressions	no.	%
1 component	63 810	6.5
2 components	660 627	67.0
3 components	107 628	10.9
4 or more components	153 973	15.6
total	986 038	100.0

components/expression av: 2.5
components/expression high: 19.0

Table III. Component Statistics (June 1987)

components	7 094 516	
acyclic	777 867	(11% of total)
atoms/acyclic component av:	15.9	
atoms/acyclic component high:	253	
cyclic	6 316 649	
with one ring systems	2 821 733	(44.6%)
with two ring systems	2 045 446	(32.4%)
with three ring systems	871 762	
with four or more ring systems	577 708	
ring systems/cyclic fragment av:	1.9	
ring systems/cyclic fragment high:	36	

Table IV. Ring Statistics (June 1987)

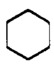
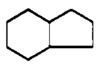
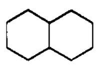
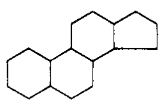
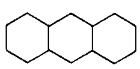
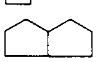
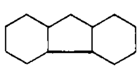
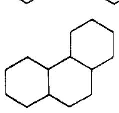
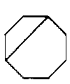
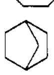
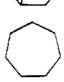
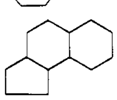
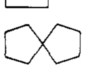
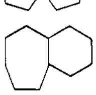
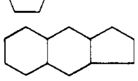
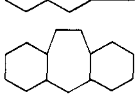
	total	av	high
ring graphs	38 842		
ring systems/ring graph		6.9	6908
ring-node sets	149 690		
sets/ring graph		3.8	2089
ring-node-bond sets (ring systems)	270 782		
sets/ring node variant		1.8	386

0095-2338/88/1628-0180$01.50/0

Table V. Ring Graph Categorization of 200 Most Frequent Ring Graphs

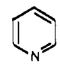
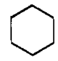
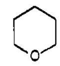
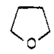
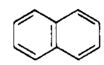
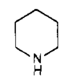
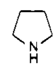
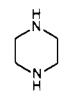
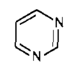
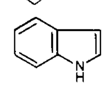
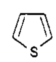
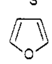
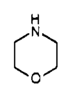
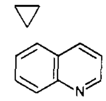
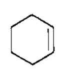
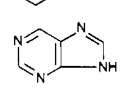
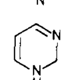
type	occurrence 1978	1987
ortho-fused	98	86
spiro-fused	26	31
fused-bridged	25	25
bridged	17	21
single	15	14
spiro	13	11
macro	3	4
π-complex	2	6
polyhedral	1	2

Table VI. Ring Graphs

		1974	1978	1987
(1)		2 362 044	4 143 118	8 263 432
(2)		368 894	689 176	1 466 857
(3)		202 188	358 567	697 101
(4)		210 128	342 957	657 161
(5)		47 116	85 826	170 569
(6)		67 395	93 378	147 342
(7)		45 175	70 063	131 155
(8)		21 325	39 303	89 952
(9)		15 150	32 784	79 200
(10)		22 954	39 161	74 060
(11)		21 282	35 109	61 215
(12)		7 194	23 189	56 538
(13)		16 299	26 560	51 823
(14)		8 339	19 218	47 678
(15)		16 239	26 321	47 456
(16)		13 572	23 124	42 715
(17)		13 500	23 447	40 798
(18)		11 372	19 298	34 832
(19)		8 729	15 739	30 035
(20)		11 018	16 421	25 595

components is called an "expression". About 12.2% of the recorded substances are made up of expressions, broken down

Table VII. Ring Systems

		1974	1976	1978	1987
(1)		1 181 106	2 577 641	3 269 266	6 436 928
(2)		95 086	130 190	161 097	310 502
(3)		80 004	112 132	142 050	285 470
(4)		44 935	69 716	99 482	268 217
(5)		37 767	56 355	76 062	191 739
(6)		61 509	78 793	95 295	188 160
(7)		57 048	74 685	93 236	175 495
(8)		40 285	56 110	75 046	160 831
(9)		27 794	39 152	48 643	96 315
(10)		25 531	33 743	40 829	83 829
(11)		16 366	26 527	40 224	83 233
(12)		23 139	32 739	41 068	82 337
(13)		22 549	32 128	41 141	81 768
(14)		27 116	36 751	44 873	80 868
(15)		27 959	37 565	46 152	79 262
(16)		17 130	24 570	33 304	73 589
(17)		25 087	32 564	39 564	71 175
(18)		17 919	25 801	33 147	69 973
(19)		12 505	19 484	25 899	66 203
(20)		10 811	16 921	23 748	64 822

into various types, as listed in Table II. The overall percent is the same as that for the 1978 analysis. The variation as to number of components is similar, with a larger percentage found for expressions with three and four or more components. As before, from 1 to 19 components are found in expressions, the average being 2.5. Two-component expressions are the most common; those with five or more are almost always alloys.

The majority of chemical substances are represented by a single component, and these may either be acyclic (containing no ring structure) or cyclic. Table III lists statistics for these single-component substances. A small portion (11%, down from 12.5% in 1978) contain no rings. These acyclic substances contain up to 253 non-hydrogen atoms, the average being about 16. Of the substances containing rings, 44.6% contain one ring system and 32.4% two ring systems. The

Table VIII. Elemental Composition Statistics by Substance

element symbol	5/31/67 no. of substances	5/31/67 %	2/28/74 no. of substances	2/28/74 %	2/18/79 no. of substances	2/18/79 %	7/30/87 no. of substances	7/30/87 %
Ac			62	0.002 275	106	0.002 328	224	0.002 658
Ag	88	0.014 756	6 336	0.232 474	12 162	0.267 151	22 733	0.269 754
Al	909	0.152 422	19 436	0.713 127	37 902	0.832 558	68 590	0.813 902
Am	2	0.000 335	332	0.012 181	451	0.009 906	638	0.007 571
Ar	2	0.000 335	127	0.004 660	200	0.004 393	318	0.003 773
As	2 165	0.363 030	15 162	0.556 309	25 162	0.552 710	38 905	0.461 654
At	2	0.000 335	86	0.003 155	160	0.003 514	444	0.005 269
Au	144	0.024 146	3 117	0.114 366	6 641	0.145 876	14 212	0.168 642
B	5 417	0.908 333	36 057	1.322 970	58 428	1.283 433	109 481	1.299 123
Ba	39	0.006 539	4 282	0.157 111	7 035	0.154 531	10 961	0.130 065
Be	76	0.012 744	2 669	0.097 928	4 174	0.091 686	5 790	0.068 705
Bi	186	0.031 189	3 002	0.110 146	5 367	0.117 891	9 091	0.107 876
Bk			53	0.001 945	107	0.002 350	164	0.001 946
Br	24 378	4.087 750	147 014	5.394 095	243 199	5.342 127	438 177	5.199 495
C	594 706	99.676 353	2 630 958	96.532 551	4 293 917	94.320 495	7 961 367	94.471 147
Ca	72	0.012 073	6 839	0.250 930	11 770	0.258 540	17 568	0.208 465
Cd	120	0.020 113	6 266	0.229 906	10 445	0.229 435	16 991	0.201 619
Ce	8	0.001 341	2 412	0.088 499	4 363	0.095 837	8 123	0.096 389
Cf			76	0.002 789	134	0.002 943	185	0.002 195
Cl	83 432	13.990 043	544 891	19.992 610	894 992	19.659 459	1 632 270	19.368 837
Cm			130	0.004 770	214	0.004 700	370	0.003 643
Co	179	0.030 015	38 157	1.400 014	67 243	1.477 064	119 284	1.415 447
Cr	136	0.022 805	22 740	0.834 354	49 691	1.091 516	99 886	1.185 267
Cs	25	0.004 192	3 789	0.139 022	6 275	0.137 837	8 888	0.105 467
Cu	251	0.042 088	32 635	1.197 412	61 081	1.341 709	116 034	1.376 882
D	2 806	0.470 516	15 977	0.586 213	30 028	0.659 597	59 658	0.707 914
Dy	7	0.001 173	1 334	0.048 946	2 378	0.052 235	4 397	0.052 176
Er	10	0.001 676	1 575	0.057 788	2 655	0.058 319	4 356	0.051 689
Es			33	0.001 211	88	0.001 933	123	0.001 460
Eu	8	0.001 341	2 043	0.074 960	3 308	0.072 663	5 497	0.065 228
F	58 687	9.840 752	166 386	6.104 873	265 714	5.836 693	519 113	6.159 897
Fe	146	0.024 481	40 322	1.479 456	90 688	1.992 059	190 117	2.255 966
Fm			42	0.001 541	77	0.001 691	108	0.001 282
Fr			45	0.001 651	78	0.001 713	446	0.005 292
Ga	104	0.017 439	3 252	0.119 319	5 594	0.122 878	9 906	0.117 547
Gd	9	0.001 509	1 888	0.069 273	3 447	0.075 717	6 206	0.073 642
Ge	1 306	0.218 992	8 677	0.318 368	14 030	0.308 184	23 591	0.279 935
H	596 347	99.996 646	2 636 306	96.728 775	4 289 454	94.222 461	7 923 788	94.025 228
He			107	0.003 926	156	0.003 426	219	0.002 599
Hf	8	0.001 341	1 252	0.045 937	2 455	0.053 926	5 046	0.059 877
Hg	1 769	0.296 629	11 601	0.425 653	18 273	0.401 386	30 216	0.358 549
Ho	3	0.000 503	1 214	0.044 543	2 017	0.044 305	3 720	0.044 142
I	6 957	1.166 563	64 849	2.379 376	99 516	2.185 975	172 552	2.047 536
In	45	0.007 546	3 279	0.120 310	5 817	0.127 776	10 228	0.121 367
Ir	16	0.002 683	4 717	0.173 072	8 472	0.186 096	15 794	0.187 415
K	161	0.026 997	17 433	0.639 635	28 174	0.618 872	47 248	0.560 654
Kr			184	0.006 751	290	0.006 370	349	0.004 141
La	11	0.001 844	2 843	0.104 313	4 847	0.106 469	8 593	0.101 966
Li	298	0.049 696	7 786	0.285 676	14 496	0.318 420	28 095	0.333 381
Lr			14	0.000 514	53	0.001 164	86	0.001 020
Lu	3	0.000 503	931	0.034 159	1 521	0.033 410	2 678	0.031 778
Md			25	0.000 917	58	0.001 274	86	0.001 020
Mg	320	0.053 658	8 726	0.320 166	16 277	0.357 541	27 352	0.324 564
Mn	66	0.011 067	19 931	0.731 289	48 062	1.055 733	96 636	1.146 702
Mo	29	0.004 863	14 476	0.531 139	34 336	0.754 227	71 840	0.852 468
N	383 050	64.230 583	1 751 974	64.281 725	2 872 142	63.089 681	5 400 264	64.080 595
Na	187	0.031 356	41 217	1.512 295	71 326	1.566 752	126 625	1.502 557
Nb	50	0.008 384	5 328	0.195 490	11 441	0.251 313	22 111	0.262 373
Nd	10	0.001 677	2 435	0.089 343	4 016	0.088 215	8 848	0.100 720
Ne			100	0.003 669	135	0.002 965	160	0.001 899
Ni	117	0.019 619	34 294	1.258 282	69 436	1.525 236	131 312	1.558 174
No			23	0.000 844	64	0.001 405	93	0.001 104
Np			561	0.020 584	886	0.019 461	1 254	0.014 880
O	492 746	82.624 625	2 215 027	81.271 616	3 630 656	79.751 256	6 808 457	80.790 490
Os	47	0.007 881	2 000	0.073 382	4 062	0.089 226	10 226	0.121 344
P	28 044	4.702 473	169 677	6.225 623	292 133	6.417 014	552 793	6.559 551
Pa			356	0.013 062	441	0.009 687	589	0.006 989
Pb	404	0.067 743	6 707	0.246078	11 351	0.249 336	18 370	0.217 982
Pd	36	0.006 036	8 564	0.314 222	17 305	0.380 122	35 873	0.425 676
Pm			143	0.005 247	205	0.004 503	366	0.004 343
Po	62	0.010 396	162	0.005 944	241	0.005 293	399	0.004 735
Pr	9	0.001 509	1 894	0.069 493	3 181	0.069 874	6 159	0.073 084
Pt	172	0.028 841	12 378	0.454 162	24 745	0.543 550	48 771	0.578 726
Pu	3	0.000 503	656	0.024 069	980	0.021 526	1 318	0.015 640

Table VIII (Continued)

	5/31/67		2/28/74		2/18/79		7/30/87	
element symbol	no. of substances	%	no. of substances	%	no. of substances	%	no. of substances	%
Ra	4	0.000 671	72	0.002 642	119	0.002 613	285	0.003 382
Rb	14	0.002 347	2486	0.091 214	4 140	0.090 939	6 022	0.071 458
Re	13	0.002 180	3 606	0.132 308	6 526	0.143 380	13 440	0.159 482
Rh	3	0.000 503	6 906	0.253 388	14 108	0.309 897	30 548	0.362 489
Rn			54	0.001 981	81	0.001 779	195	0.002 314
Ru	7	0.001 174	4 746	0.174 136	9 724	0.213 598	26 884	0.319 011
S	118 536	19.876 351	569 805	20.906 731	967 434	21.250 726	1 865 753	22.139 392
Sb	881	0.147 728	8 442	0.309 746	14 741	0.323 801	23 717	0.281 431
Sb	16	0.002 683	1 513	0.055 514	2 163	0.047 512	3 155	0.037 438
Se	2 116	0.354 815	14 765	0.541 743	25 251	0.554 665	46 125	0.547 328
Si	8 628	1.446 760	66 152	2.427 185	125 165	2.749 383	264 263	3.135 797
Sm	46	0.007 713	2068	0.075 877	3 483	0.076 507	6 771	0.080 346
Sn	2 857	0.479 067	21 605	0.792 710	36 652	0.805 100	64 731	0.768 111
Sr	25	0.004 192	2 325	0.085 307	3 863	0.084 854	5 585	0.066 273
T	579	0.097 088	2 313	0.084 866	4 069	0.089 379	8 428	0.100 008
Ta	38	0.006 372	2 971	0.109 009	6 024	0.132 323	11 094	0.131 644
Tb	5	0.000 838	1 190	0.043 662	2 002	0.043 976	4 472	0.053 066
Tb	4	0.000 671	345	0.012 658	650	0.014 277	1 923	0.022 819
Te	225	0.037 728	3 506	0.128 639	6 153	0.135 157	13 632	0.161 760
Th	48	0.008 049	1 966	0.072 135	3 053	0.067 062	4 867	0.057 753
Ti	407	0.068 247	12 402	0.455 042	24 710	0.542 781	46 239	0.548 681
Tl	129	0.021 631	3 467	0.127 208	5 846	0.128 413	8 622	0.102 310
Tm	3	0.000 503	829	0.030 417	1 361	0.029 895	2 468	0.029 286
U	137	0.022 972	5 799	0.212 771	9 675	0.212 521	14 764	0.175 193
V	145	0.024 314	8 705	0.319 395	17 437	0.383 022	32 809	0.389 318
W	57	0.009 558	9 255	0.297 564	18 122	0.398 069	38 842	0.460 907
Xe			422	0.015 484	574	0.012 608	632	0.007 523
Y	9	0.001 509	2 192	0.080 427	3 985	0.087 534	7 391	0.087 703
Yb	5	0.000 838	1 320	0.048 432	2 233	0.049 050	3 939	0.046 741
Zn	595	0.099 771	14 703	0.539 468	25 030	0.549 810	43 157	0.512 109
Zr	57	0.009 558	4 914	0.180 300	9 703	0.213 136	21 013	0.249 344

figures from 1978 were 48.0% and 31.8%.

A basic design feature of the CAS Chemical Registry System is that the ring systems present in a structure are recognized during the registration process.[1] The systems are stored in the structural record of the substance as an identifying number, linking the structure record to a file of ring systems. In this file, ring systems are recorded as composites of the ring graphs or basic patterns, as graph-node (atom) variations, and as graph-node-bond variations for fully specified systems. Statistics for the ring system constituents, listed in Table IV, show that in the total file of over 8 million substances there are 38 842 basic graphs and 270 782 different ring systems. The number of ring systems per ring graph ranges from 1 to 6908, the average being almost 7. For each graph, there is an average of 3.8 different sets of specific atoms, ranging from 1 to 2089 sets. For each graph-node variant, there are on the average almost 2 specific bond variations, ranging from 1 to 386.

Ring systems are very common in chemical substances, occurring 13 522 531 times in 6 316 649 different substances. Of the total number of occurrences, 6 436 928, 47.6%, are phenyl rings, down from 55% in the 1978 statistics. The next most common ring is the pyridyl ring, occurring 310 502 times, 2.30%, down slightly from the 2.72% reported previously. The 200 most common ring graphs, which account for over 96% of the total ring occurrences, are categorized as listed in Table V, with no substantial variation from that reported previously. The number of ortho-fused graphs decreased somewhat, and the number of spiro-fused graphs increased, with small differences in other types.

The 20 most frequently occurring ring graphs, with the number of occurrences given for 1974, 1978, and 1987, are pictured in Table VI. Again, the six-membered ring occurs most frequently, to the extent of over 61% of total graph occurrences. The same graphs are in the top 20 as previously, with some slight shifting, in most cases by one or two places in the list. The total number of different graphs showed a 73% increase from 1978 to 1987. The percentage increase in number of occurrences of individual graphs from 1978 to 1987 shows that the bicyclo[3.2.0] (no. 14 on the list) increased the most (148.1%), followed by the bicyclo[4.2.0] graph (no. 12) (141.6%) and the bicyclo[3.3.0] graph (no. 9) (141.6%). The same three graphs, in slightly different order, showed the largest percentage increase of occurrence previously, from 1974 to 1978. The typical tetracyclic steroid graph, which showed the smallest increase from 1974 to 1978, similarly showed a comparatively small increase from 1978 to 1987 (57.8%). This is the same trend found previously, indicating a higher publishing activity in such fields as cephalosporins, penicillins, prostaglandins, and other classes having the above bicyclo systems as compared to the activity in the steroid field. The rather large increase from 1978 to 1987 for the four-membered ring graph (no. 8) (128.9%) and the three-membered ring (no. 5) (98.7%) indicate increasing publishing activity involving small (three- and four-membered) ring compounds.

Occurrence statistics for the top 20 individual ring systems are shown in Table VII, figures being given for 1974, 1976, 1978, and 1987. The first four remain in the same order as in 1978. These are all six-membered rings, as are the majoriy of the top 20 and also the top 200 (supplementary material). Six-membered rings, single and fused, account for about 70% of all ring occurrences. Statistics for ring system occurrence support the trends noted above in the discussion of ring graph occurrence. There is no steroid system in the top 20 systems, the one previously occurring as no. 19 in 1978 having dropped to no. 25 in 1987 and showing a relatively small increase (52.9%). The percentage increase in occurrence of the three-membered cyclopropane ring (no. 16) is large (121.0%), as is the case for ethylene oxide (oxirane) (no. 21, not shown) at 97.6%. The dihydropyrimidine system (no. 20) shows the largest increase in frequency of occurrence at 173%, followed by the tetrahydropyran ring (no. 4) at 169.6% and the purine

Table IX. Elemental Composition Statistics by Occurrence

element	5/31/67 no. of occurrences	5/31/67 %	2/28/74 no. of occurrences	2/28/74 %	2/18/79 no. of occurrences	2/18/79 %	7/30/87 no. of occurrences	7/30/87 %
Ac			64	0.000 054	108	0.000 053	229	0.000 058
Ag	91	0.0003	9 129	0.007 725	15 448	0.007 703	28 212	0.007 147
Al	947	0.0039	29 753	0.025 177	50 972	0.025 417	86 616	0.021 943
Am	2		368	0.000 311	504	0.000 251	723	0.000 183
Ar	2		173	0.000 146	243	0.000 121	401	0.000 102
As	2 496	0.0102	25 610	0.021 672	42 531	0.021 208	64 479	0.016 335
At	2		93	0.000 079	167	0.000 083	471	0.000 119
Au	144	0.0006	4 530	0.003 833	8 092	0.004 035	17 633	0.004 467
B	7 279	0.0300	115 237	0.097 516	182 493	0.091 002	285 272	0.072 270
Ba	39	0.0002	5 717	0.004 838	7 618	0.003 798	11 753	0.002 977
Be	77	0.0003	4 664	0.003 947	5 743	0.002 863	7 459	0.001 890
Bi	193	0.0008	5 101	0.004 317	9 314	0.004 644	16 600	0.004 205
Bk			56	0.000 047	115	0.000 057	176	0.000 045
Br	34 255	0.1413	228 223	0.193 126	377 743	0.188 365	651 845	0.165 137
C	9 400 518	38.7785	44 959 807	38.045 697	76 131 591	37.963 829	149 728 571	37.931 878
Ca	73	0.0003	9 915	0.008 390	14 051	0.007 006	19 936	0.005 051
Cd	120	0.0004	7 653	0.006 476	12 190	0.006 078	19 351	0.004 902
Ce	8		3 384	0.002 864	5 548	0.002 721	9 726	0.002 464
Cf			79	0.000 067	147	0.000 073	208	0.000 053
Cl	149 248	0.6157	974 531	0.824 663	1 587 485	0.791 616	2 765 829	0.700 689
Cm			163	0.000 138	255	0.000 127	365	0.000 092
Co	198	0.0008	46 754	0.039 564	79 123	0.039 455	137 635	0.034 868
Cr	159	0.0006	26 216	0.022 184	55 462	0.027 656	108 556	0.027 501
Cs	26	0.0001	4 423	0.003 743	6 602	0.003 292	9 396	0.002 380
Cu	262	0.0010	39 908	0.033 771	72 696	0.036 250	137 501	0.034 834
D	7 833	0.0323	52 560	0.044 477	99 610	0.049 671	194 838	0.049 360
Dy	7		2 043	0.001 729	3 263	0.001 627	5 617	0.001 423
Er	10		2 764	0.002 339	3 455	0.001 722	5 507	0.001 395
Es			35	0.000 030	90	0.000 044	127	0.000 032
Eu	8		2 473	0.002 093	4 195	0.002 091	6 723	0.001 703
F	211 854	0.8730	723 577	0.612 302	1 181 628	0.589 231	2 212 825	0.560 592
Fe	149	0.0006	51 396	0.043 492	108 053	0.053 881	222 572	0.056 386
Fm			44	0.000 037	79	0.000 039	111	0.000 028
Fr			49	0.000 041	80	0.000 039	453	0.000 115
Ga	108	0.0004	5 290	0.004 476	7 872	0.003 295	13 251	0.003 357
Gd	9		3 014	0.002 550	4 820	0.002 403	8 013	0.002 030
Ge	1 563	0.0064	13 219	0.011 186	20 331	0.010 138	31 870	0.008 074
H	11 547 701	47.6360	55 600 781	47.050 257	94 297 072	47.022 240	186 360 933	47.212 234
He			137	0.00 116	203	0.000 101	289	0.000 073
Hf	8		1 835	0.001 553	3 032	0.001 511	5 838	0.001 479
Hg	1 974	0.0081	14 071	0.011 907	22 044	0.010 992	36 042	0.009 131
Ho	3		1 844	0.001 560	2 770	0.001 381	4 833	0.001 244
I	10 429	0.0430	94 153	0.079 674	146 121	0.072 864	246 240	0.062 382
In	47		4 406	0.003 728	7 187	0.003 583	12 273	0.003 109
Ir	16		5 322	0.004 504	9 682	0.004 828	18 879	0.004 783
K	180	0.0007	19 439	0.016 450	29 139	0.014 530	48 563	0.012 303
Kr			191	0.000 162	308	0.000 153	379	0.000 096
La	11		4 590	0.003 884	7 212	0.003 596	12 107	0.003 067
Li	347	0.0014	9 971	0.008 438	16 536	0.008 245	31 650	0.008 018
Lr			16	0.000 014	55	0.000 027	89	0.000 023
Lu	3		1 248	0.001 056	1 921	0.000 957	3 501	0.000 887
Md			27	0.000 023	60	0.000 029	89	0.000 023
Mg	328	0.0013	12 155	0.010 268	18 916	0.009 432	30 556	0.007 741
Mn	69	0.0002	22 979	0.019 445	52 376	0.026 117	103 454	0.026 232
Mo	29	0.0001	24 867	0.021 043	53 268	0.026 562	106 165	0.026 896
N	917 258	3.7838	4 674 936	3.956 004	8 040 072	4.009 267	16 032 745	4.061 697
Na	215	0.0009	44 699	0.037 825	73 581	0.036 691	129 397	0.032 781
Nb	50	0.0002	11 393	0.009 641	20 772	0.010 358	34 472	0.008 733
Nd	10		3 652	0.003 090	5 556	0.002 770	10 816	0.002 740
Ne			121	0.000 102	161	0.000 080	174	0.000 044
Ni	119	0.0004	39 706	0.033 600	77 127	0.038 460	145 070	0.036 752
No			27	0.000 023	66	0.000 032	113	0.000 029
Np			623	0.000 527	988	0.000 492	1 426	0.000 361
O	1 715 545	7.0769	8 702 367	7.364 080	14 932 265	7.446 133	29 456 422	7.462 420
Os	47	0.0002	2 669	0.002 259	5 864	0.002 924	17 040	0.004 317
P	35 706	0.1437	260 947	0.220 817	466 904	0.232 826	974 028	0.246 758
Pa			416	0.000 352	509	0.000 253	661	0.000 167
Pb	470	0.0019	9 529	0.008 064	15 205	0.007 582	24 470	0.006 199
Pd	37	0.0001	10 740	0.009 088	20 936	0.010 439	43 174	0.010 938
Pm			156	0.000 132	229	0.000 114	397	0.000 101
Po	62	0.0002	165	0.000 140	245	0.000 122	407	0.000 103
Pr	9		2 873	0.002 431	4 276	0.002 132	7 861	0.001 991
Pt	174	0.0007	14 426	0.012 208	28 255	0.014 089	55 628	0.014 093
Pu	3		891	0.000 754	1 088	0.000 542	1 509	0.000 382

Table IX (Continued)

	5/31/67		2/28/74		2/18/79		7/30/87	
element	no. of occurrences	%	no. of occurrences	%	no. of occurrences	%	no. of occurrences	%
Ra	4		72	0.000 061	119	0.000 059	289	0.000 073
Rb	14		3 061	0.002 590	4 445	0.002 216	6 324	0.001 602
Re	13		4 980	0.004 214	8 874	0.004 425	18 472	0.004 680
Rh	3		8 810	0.007 455	17 881	0.008 916	41 119	0.010 417
Rn			54	0.000 046	81	0.000 040	197	0.000 050
Ru	7		5 987	0.005 066	12 227	0.006 097	35 688	0.009 041
S	166 664	0.6875	881 655	0.746 070	1 500 099	0.748 040	2 885 595	0.731 030
Sb	967	0.0040	10 787	0.009 128	19 600	0.009 773	30 047	0.007 612
Sc	16		1 986	0.001 681	2 965	0.001 478	4 264	0.001 080
Se	2 534	0.0104	25 105	0.021 244	41 282	0.020 585	66 087	0.016 742
Si	16 362	0.0675	120 257	0.101 763	208 384	0.103 912	408 113	0.103 390
Sm	46	0.0002	2 903	0.002 457	4 699	0.002 343	8 474	0.002 147
Sn	3 396	0.0144	26 774	0.022 657	43 476	0.021 679	76 224	0.019 310
Sr	25	0.0001	3 684	0.003 117	4 518	0.002 252	6 253	0.001 584
T	857	0.0035	3 486	0.002 950	6 458	0.003 220	13 811	0.003 499
Ta	38	0.0001	5 653	0.004 784	10 251	0.005 111	16 845	0.004 267
Tb	5		1 740	0.001 472	2 689	0.001 340	5 457	0.001 382
Tc	4		424	0.000 360	749	0.000 373	2 303	0.000 583
Te	231	0.0009	6 889	0.005 830	10 881	0.005 425	20 806	0.005 271
Th	48	0.0002	2 349	0.001 988	3 587	0.001 788	5 518	0.001 398
Ti	478	0.0019	17 361	0.014 691	29 962	0.014 940	54 547	0.013 819
Tl	140	0.0006	4 193	0.003 548	6 888	0.003 434	10 413	0.002 638
Tm	3		1 170	0.000 990	1 760	0.000 877	3 172	0.000 804
U	141	0.0006	7 512	0.006 357	12 681	0.006 323	18 563	0.004 703
V	148	0.0006	14 492	0.012 263	26 794	0.013 361	46 203	0.011 705
W	59	0.0002	134 149	0.020 113	44 549	0.022 214	88 739	0.022 481
Xe			462	0.000 391	643	0.000 320	729	0.000 185
Y	9		3 442	0.002 913	5 700	0.002 842	9 842	0.002 493
Yb	5		1 854	0.001 569	2 932	0.001 462	4 981	0.001 262
Zn	613	0.0025	19 132	0.016 190	28 311	0.014 117	47 918	0.012 139
Zr	57	0.0002	6 747	0.005 709	12 063	0.006 015	24 553	0.006 220

Table X. Twenty Most Frequently Occurring Elements According to Number of Substances

1967				1974				1979				1987			
element	%	element	%	element	%	element	%	element	%	element	%	element	%	element	%
H	99.99	I	1.17	H	96.73	I	2.34	C	94.32	I	2.19	C	94.47	Fe	2.26
C	99.68	B	0.91	C	96.53	Na	1.51	H	94.22	Fe	1.99	H	94.03	I	2.05
O	82.62	Sn	0.48	O	81.27	Fe	1.48	O	79.75	Na	1.57	O	80.79	Ni	1.56
N	64.23	D	0.47	N	64.28	Co	1.40	N	63.09	Ni	1.53	N	64.08	Na	1.50
S	19.88	As	0.36	S	20.91	B	1.32	S	21.25	Co	1.48	S	22.14	Co	1.42
Cl	13.99	Se	0.35	Cl	19.99	Ni	1.26	Cl	19.66	Cu	1.34	Cl	19.37	Cu	1.38
F	9.84	Hg	0.30	P	6.23	Cu	1.20	P	6.42	B	1.28	P	6.56	B	1.30
P	4.70	Ge	0.22	F	6.10	Cr	0.83	F	5.84	Cr	1.09	F	6.16	Cr	1.19
Br	4.09	Al	0.152	Br	5.39	Sn	0.79	Br	5.34	Mn	1.06	Br	5.20	Mn	1.15
Si	1.45	Sb	0.148	Si	2.43	Mn	0.73	Si	2.75	Al	0.83	Si	3.14	Mo	0.85

Table XI. Twenty Most Frequently Occurring Elements According to Number of Atoms

1967		1974		1979		1987	
element	%	element	%	element	%	element	%
H	47.64	H	47.05	H	47.02	H	47.21
C	38.79	C	38.05	C	37.96	C	37.93
O	7.08	O	7.36	O	7.45	O	7.46
N	3.78	N	3.96	N	4.01	N	4.06
F	0.87	Cl	0.82	Cl	0.79	S	0.73
S	0.69	S	0.75	S	0.75	Cl	0.70
Cl	0.62	F	0.61	F	0.59	F	0.56
P	0.15	P	0.22	P	0.23	P	0.25
Cr	0.14	Br	0.19	Br	0.19	Br	0.17
Si	0.07	Si	0.10	Si	0.10	Si	0.10
I	0.04	B	0.098	B	0.093	B	0.072
D	0.032	I	0.080	I	0.074	I	0.062
B	0.030	D	0.044	Fe	0.052	Fe	0.056
Sn	0.014	Fe	0.044	D	0.050	D	0.049
Se	0.010	Co	0.040	Co	0.039	Ni	0.037
As	0.010	Cu	0.034	Ni	0.038	Co	0.035
Hg	0.008	Ni	0.034	Cu	0.036	Cu	0.035
Ge	0.006	Al	0.025	Cr	0.028	Na	0.033
Al	0.004	Sn	0.023	Al	0.025	Cr	0.028
Sb	0.004	Cr	0.022	Sn	0.022	W	0.022

Table XII. Registry Structure File Statistics

	substances		occurrences			substances		occurrences	
element	no.	%	no.	%	element	no.	%	no.	%
Ac	118	111.0	121	112.0	Mn	48 574	101.1	51 169	97.7
Ag	10 571	86.9	12 764	82.6	Mo	37 504	109.2	52 897	99.3
Al	30 688	81.0	37 644	73.9	N	2 528 122	88.0	7 992 673	99.4
Am	187	41.5	218	43.5	Na	55 299	77.5	55 816	75.9
Ar	118	59.0	158	65.0	Nb	10 670	93.3	13 700	66.0
As	13 743	54.6	21 948	51.6	Nd	4 472	111.4	5 260	94.7
At	284	177.5	304	182.0	Ne	25	18.5	13	8.1
Au	7 571	114.0	9 541	117.9	Ni	61 876	89.1	67 943	88.1
B	51 053	87.4	102 779	56.3	No	29	45.3	47	71.2
Ba	3 926	55.8	4 135	54.3	Np	368	41.5	438	44.3
Be	1 616	38.7	1 716	29.9	O	3 177 801	87.5	14 524 157	97.3
Bi	3 724	69.4	7 286	78.2	Os	6 164	151.7	11 176	190.6
Bk	57	53.3	61	53.0	P	260 660	89.2	507 124	108.6
Br	194 978	80.2	274 102	72.6	Pa	148	33.6	152	30.0
C	3 567 450	83.1	73 596 980	96.7	Pb	7 019	61.8	9 265	60.9
Ca	5 798	49.3	5 885	41.9	Pd	18 568	107.3	22 238	106.2
Cd	6 546	62.7	7 161	58.7	Pm	161	78.5	168	73.4
Ce	3 760	86.2	4 268	78.2	Po	158	65.6	162	66.1
Cf	51	38.1	61	41.5	Pr	2 978	93.6	3 585	83.8
Cl	737 278	82.4	1 178 344	74.2	Pt	24 026	97.1	27 373	96.9
Cm	93	43.5	110	43.1	Pu	338	34.5	421	38.7
Co	52 041	77.4	58 512	74.0	Ra	166	139.5	170	142.9
Cr	50 195	101.0	53 094	95.7	Rb	1 882	45.5	1 879	42.3
Cs	2 613	41.6	2 794	42.3	Re	6 914	105.9	9 958	108.2
Cu	54 953	90.0	64 805	89.1	Rh	16 440	116.5	23 238	130.0
D	29 630	98.7	95 228	95.6	Rn	114	140.7	116	143.2
Dy	2 019	84.9	2 354	72.1	Ru	17 160	176.5	23 461	191.9
Dr	1 701	64.1	2 052	59.4	S	898 319	92.9	1 385 496	92.4
Es	35	39.8	37	41.1	Sb	8 976	60.9	10 447	53.3
Eu	2 189	66.2	2 528	60.3	Sc	992	45.9	1 299	43.8
F	253 399	95.4	1 031 197	87.3	Se	20 874	82.7	24 805	60.1
Fe	99 429	109.6	114 519	106.0	Si	139 098	111.1	199 729	95.8
Fm	31	40.3	32	40.5	Sm	3 288	94.4	3 775	80.3
Fr	368	471.8	373	466.3	Sn	28 079	76.6	32 748	75.3
Ga	4 312	77.1	5 379	68.3	Sr	1 722	44.6	1 735	38.4
Gd	2 759	80.0	3 193	66.2	T	4 359	107.1	7 353	113.9
Ge	9 561	68.1	11 539	56.8	Ta	5 070	84.2	6 594	64.3
H	3 634 334	84.8	92 063 861	97.6	Tb	2 470	123.4	2 768	102.9
He	63	40.4	86	42.4	Tc	1 273	195.8	1 554	207.5
Hf	2 591	105.5	2 806	92.5	Te	7 479	121.6	9 925	91.2
Hg	11 943	65.4	13 998	63.5	Th	1 814	59.4	1 931	53.8
Ho	1 703	84.4	2 063	74.5	Ti	21 529	87.1	24 585	82.1
I	73 036	73.4	100 119	68.5	Tl	2 776	47.5	3 525	51.2
In	4 411	75.8	5 086	70.8	Tm	1 107	81.3	1 412	80.2
Ir	7 322	86.4	9 197	95.0	U	5 089	52.6	5 882	46.4
K	19 074	67.7	19 424	66.7	V	15 372	88.2	19 409	72.4
Kr	59	20.3	71	23.1	W	20 720	114.3	44 190	99.2
La	3 746	77.3	4 895	67.9	Xe	60	10.5	86	13.4
Li	13 599	93.8	15 114	91.4	Y	3 406	85.5	4 142	72.7
Lr	33	62.3	34	61.8	Yb	1 706	76.4	2 049	69.9
Lu	1 157	76.1	1 580	82.2	Zn	18 127	72.4	19 607	69.3
Md	28	48.3	29	48.3	Zr	11 310	116.6	12 490	103.5
Mg	11 075	68.0	11 640	61.5					

Table XIII. Elemental Composition Increases from 1979 to 1987

	1967	1974	1979	1987
substances	596 367	2 725 462	4 552 475	8 427 300
atoms	24 241 534	118 173 172	200 537 175	394 730 177

ring system (no. 19) at 155.6% and then the tetrahydrofuran ring (no. 5) at 152.0%.

ELEMENT STATISTICS

Statistics have been obtained at varying time intervals on the occurrence of the chemical elements in the CAS Chemical Registry Structure File. Table VIII shows the number of substances containing a given element (atomic numbers 1–103) and the percent of the total as of 1967,[11] 1974, 1979, and 1987. Table IX lists the analogous data for occurrences (number of atoms) of the elements. The 1967 statistics are included for historical purposes only, since the Registry File at that early stage was not at all representative of the entire chemical field. Examination of the statistics of Tables VIII and IX shows that the element hydrogen is the most frequently occurring element according to the number of atoms, followed by carbon, oxygen, nitrogen, sulfur, and chlorine. This is also the case according to the number of substances, although the listings in Table VIII do not seem to bear that out, showing the element carbon first. However, if data for hydrogen (H) and its isotopes deuterium and tritium are combined, then it is the most frequently occurring for substances as well as atoms. Comparisons according to time show only minor variations, as indicated in Table X, where the 20 most frequently occurring elements according to the number of substances are listed, and in Table XI, where the criterion is the number of atoms.

Table XIII lists the number of substances (both machine and manual registration) and the number of total atoms in the four times chosen for generation of statistics. The per-

Table XIV. Frequency of Occurrence of Various Elements in Halogen-Containing Compounds

	I	Br	Cl	F
Al	1.062	3.181	9.399	1.946
As	13.835	17.200	48.888	23.170
B	1.451	3.554	8.186	21.345
Br	0.270	1.000	2.868	1.018
Cl	0.570	0.206	1.000	0.372
D	2.122	6.235	12.848	5.654
F	0.287	0.725	3.680	1.000
Ge	8.742	14.246	46.950	22.125
Hg	18.608	28.925	84.641	21.926
I	1.000	1.751	5.116	2.611
Mg	4.982	27.816	28.654	8.105
N	0.034	0.080	0.366	0.103
O	0.020	0.060	0.243	0.074
Os	55.716	71.823	231.084	73.713
P	0.411	0.831	3.391	1.426
S	0.077	0.206	0.935	0.289
Sb	13.907	24.308	118.385	48.830
Se	6.178	9.895	22.398	7.604
Si	0.325	0.728	3.317	1.509
Sn	3.706	8.285	30.375	5.983
Te	35.846	57.614	106.475	68.544
Ti	1.114	3.635	27.423	5.917
Tl	36.572	67.492	150.098	120.447
Zn	4.639	8.158	36.066	4.828

centage increase in number of substance from 1979 to 1987 is 85.1%, for the number of atoms, 96.8%. For individual elements, the percentage increase (Table XII) varied from 10.5% (xenon) to 471.8% (francium) for substances and from 8.1% (neon) to 466.3% (francium) for occurrences. While two of the rarer elements, francium and technetium, show the very highest percentage increases, several more common elements showed high increases as well. For number of substances, ruthenium (176.5%), osmium (151.7%), tellurium (121.6%), and zirconium (116.6%) show high increases; for occurrences, rutheniuim (191.9%), osmium (190.6%), rhodium (130.0%), and gold (117.9%) show high increases. Ruthenium, osmium, and gold also showed high increases in the previous paper[6] on statistics.

Besides the statistics on elemental composition, others can be obtained by appropriate manipulation of data present in the CAS ONLINE Registry File. As an illustration, occurrences of elements can be determined, as shown in Table XIV. It gives the frequency of occurrence of various elements in compounds containing the halogens iodine, bromine, chlorine, and fluorine. For comparisons, figures giving actual numbers of compounds have been normalized according to the formula $B/A = C$, where A is the proportion of arsenic (for example) compounds in the total file, B the proportion of arsenic–iodine compounds to all arsenic compounds, and C the factor found in the table. In the case of arsenic–iodine compounds, the figure 13.835 appears at the intersection of As and I; that is, arsenic appears 13.835 times as often in iodine-containing compounds as it does in the total file. Sulfur, on the other hand, appears only 0.077 times as often in iodine-containing compounds as it does in the total file. In some cases a chemical connection is clear, as for boron trifluoride complexes, chloroosmates, chlorozincates, chloroantimonates, mercuric chloride complexes, and the like.

The statistics presented here show large increases in absolute values, since the Registry File increases rapidly over time. In many cases the percentage values will show only little change over a period of time, with an occasional substantial variation that indicates some shift in activity in the field as shown in the published literature.

CONCLUSION

These statistics have been collected and presented simply as such, statistics, with very little commentary on the author's part. Derek de Solla Price hypothesized[12] that the growth of knowledge is like the gradual solving of a very large or perhaps infinite jigsaw puzzle that was started at the center ca. 1660 with the first scientific papers and journals and is now working its way outward on all fronts. It is hoped that these statistics will contribute to the solution of the puzzle in stimulating ideas and research in chemistry, chemometrics, and information science.

ACKNOWLEDGMENT

The development of the CAS Chemical Registry System was substantially supported by the National Science Foundation. Chemical Abstracts Service, a division of the American Chemical Society, gratefully acknowledges this support.

Supplementary Material Available: Cumulative occurrence statistics for the 200 most frequently occurring ring graphs in the CAS Chemical Registry System for the years 1974, 1978, and 1987 and for the 200 most frequently occurring ring systems in the CAS Chemical Registry System for the years 1974, 1976, 1978, and 1987 (39 pages). Ordering information is given on any current masthead page.

REFERENCES

(1) Dittmar, P. G.; Stobaugh, R. E.; Watson, C. E. "The Chemical Abstracts Service Registry System. 1. General Design". *J. Chem. Inf. Comput. Sci.* **1976**, *16*, 111–121.
(2) Freeland, R. G.; Funk, S. A.; O'Korn, L. J.; Wilson, G. A. "The Chemical Abstracts Service Chemical Registry System. 2. Augmented Connectivity Molecular Formula". *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 94–98.
(3) Blackwood, J. E.; Elliott, P. M.; Stobaugh, R. E.; Watson, C. E. "The Chemical Abstracts Service Chemical Registry System. 3. Stereochemistry". *J. Chem. Inf. Comput. Sci.* **1977**, *17*, 3–8.
(4) Vander Stouw, G. G.; Gustafson, C. R.; Rule, J. D.; Watson, C. E. "The Chemical Abstracts Service Chemical Registry System. 4. Use of the Registry System To Support the Preparation of Index Nomenclature". *J. Chem. Inf. Comput. Sci.* **1976**, *16*, 213–218.
(5) Zamora, A.; Dayton, D. L. "The Chemical Abstracts Service Chemical Registry System. 5. Structure Input and Editing". *J. Chem. Inf. Comput. Sci.* **1976**, *16*, 219–222.
(6) Stobaugh, R. E. "The Chemical Abstracts Service Chemical Registry System. 6. Substance-Related Statistics". *J. Chem. Inf. Comput. Sci.* **1980**, *20*, 76–82.
(7) Mockus, J.; Stobaugh, R. E. "The Chemical Abstracts Service Chemical Registry System. 7. Tautomerism and Alternating Bonds". *J. Chem. Inf. Comput. Sci.* **1980**, *20*, 18–22.
(8) Moosemiller, J. P.; Ryan, A. W.; Stobaugh, R. E. "The Chemical Abstracts Service Chemical Registry System. 8. Manual Registration". *J. Chem. Inf. Comput. Sci.* **1980**, *20*, 83–88.
(9) Ryan, A. W.; Stobaugh, R. E. "The Chemical Abstracts Service Chemical Registry System. 9. Input Structure Conventions". *J. Chem. Inf. Comput. Sci.* **1982**, *22*, 22–28.
(10) Hamill, K. A.; Nelson, R. D.; Vander Stouw, G. G.; Stobaugh, R. E. "The Chemical Abstracts Service Chemical Registry System. 10. Registration of Substances from Pre-1965 Indexes of Chemical Abstracts". *J. Chem. Inf. Comput. Sci.* (preceding paper in this issue).
(11) Leighner, H. L.; Leiter, D. P. "A Statistical Analysis of the Structure Registry at Chemical Abstracts Service". Presented at the 154th National Meeting of the American Chemical Society, Chicago, IL, Sept 11, 1967.
(12) de Solla Price, D. "The Revolution of Mapping of Science". Presented at the 42nd Annual Meeting of the American Society for Information Science, Minneapolis, MN, Oct 14–18, 1979.